Project no. 033104

MultiMatch

Technology-enhanced Learning and Access to Cultural Heritage
Instrument: Specific Targeted Research Project
FP6-2005-IST-5

D1.1.3 – State of the Art

Start Date of Project: 01 May 2006
Duration: 30 Months

Organization Name of Lead Contractor for this Deliverable: ISTI-CNR

Final Version

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Document Information

Deliverable number:	D1.1.3
Deliverable title:	State of the Art Report
Due date of deliverable:	July 2008
Actual date of deliverable:	December 2008
Main Author(s):	Editor & Introduction: Carol Peters, ISTI-CNR Section 2: Johan Oomen, BandG Section 3: Carl Ibbotson OCLC PICA; Contributions from WIND, UniGE, Alinari, ISTI-CNR Section 4: Neil Ireson, USFD Section 5: Martha Larson and Jaap Kamps, UvA Section 6 Stephane Marchand-Maillet and Eric Bruno, UniGE Section 7: Gareth Jones, DCU; Contributions from UniGE, UvA Section 8: Paul Clough, USFD
Participant(s):	All Partners
Workpackage:	1
Workpackage title:	User Requirements & Functional Specification
Workpackage leader:	UNED
Dissemination Level:	PU (Public)
Version:	Final
Keywords:	cultural heritage, metadata, digital asset management, search engines, multilingual indexing and retrieval, multimedia indexing and retrieval, information classification, information extraction, user interaction, interface design

History of Versions

Version	Date	Author (Partner)	Description/Approval Level
V1	15.11.08	C.Peters, ISTI-CNR	Version1 with contributions for most sections
Final version	39.11.08	All	Completed and checked

Abstract

MultiMatch aims at complex, heterogeneous digital object retrieval and presentation. The development of the system implies addressing a number of significant research challenges in a multidisciplinary context. This report describes the state of the art in the relevant areas of research, thus specifying the scientific and technology baseline from which the consortium partners start. It has been released in three instalments (D1.1.1; D1.1.2, and the current document D1.1.3). We originally identified six main areas: existing technology for cultural heritage; search engines; information extraction and classification; multilingual/multimedia indexing; multilingual/multimedia retrieval; user interaction and interface design. Each area was first reviewed in a separate chapter in D1.1.1, released in December 2006. A substantial update describing Image Collections and Browsing was added as a separate report in December 2007 (D1.2). In this final version, we provide significant updates to the original documents and in addition each chapter terminates with a section which relates the general sate-of-the-art in that area to what has been done in MultiMatch. Our aim has been to provide a complete panorama of the actual state-of-the-art in the areas of interest to MultiMatch, covering as far as possible all relevant aspects.

Table of Contents

Executive Summary

The objective of MultiMatch is to develop a multilingual search engine specifically designed for access, organization and personalised presentation of cultural heritage (CH) information. The development of the system thus implies addressing a number of significant research challenges in a multidisciplinary context. R&D expertise is required in a diverse set of system- and user-oriented research areas. On the system side, these relate to focused Internet crawling, information extraction and analysis, multilingual information access and retrieval, multimedia complex object management, and interface design. On the user side, relevant areas include user profiling, metadata and ontology studies, user/system interaction, and user-centred interface design. The technology in these areas tends to develop rapidly. For this reason, it was decided to prepare a detailed State of the Art (SotA) report in the initial phases of the project, to be updated during and at the end of the project. This document has been released in three instalments (D1.1.1; D1.1.2, and the current document D1.1.3). We originally identified six main areas: existing technology for cultural heritage; search engines; information extraction and classification; multilingual/multimedia indexing; multilingual/multimedia retrieval; user interaction and interface design. Each area was first reviewed in a separate chapter in D1.1.1, released in December 2006. A substantial update describing Image Collections and Browsing was added as a separate report in December 2007 (D1.1.2). In this final version, we provide a series of updates to the original documents. Interestingly, one of the new areas which has been considered of importance to a number of authors is the advent of applications such as social tagging or other types of applications which leverage collective intelligence. From different perspectives, this emerging phenomenon is commented in Chapters 2, 4 and 8. Our aim has been to provide a complete panorama of the actual state-of-the-art in the areas of interest to MultiMatch, covering as far as possible all relevant aspects. In addition each chapter now terminates with a section which relates the general state-of-the-art in that area to what has been done in MultiMatch. Here below we briefly outline the main points of each chapter.

Technology for Cultural Heritage

Chapter 2 attempts to list the major new developments in the CH technology area. A wide range of technologies are used in the different domains that can be classified under the general heading of cultural heritage. In D1.1.1, we focused on metadata and encoding standards, and digital asset management systems as being of prime interest for the activity being undertaken in MultiMatch. However, in the past two years there have been major technological developments and Chapter 2 of the current document focuses on four main areas where development has been most evident: digital library software; metadata interoperability; digitization standards; the impact of Web 2.0 on Cultural Heritage activities. The first section of Chapter 2 describes latest trends in digital libraries for cultural heritage. The most important recent development here for us is Europeana, the European Digital Library. Europeana has very similar objectives to MultiMatch, aiming at providing access to multimedia collections in many languages, and intending to activate cross-language search. Chapter 2 thus briefly describes the current state of development of Europeana. It is commonly agreed that metadata interoperability is an important key to ensuring access to heritage collections. However, interoperability is hindered by the diversity of metadata formats and standards that exist in the cultural heritage domain. The second section in this chapter thus discusses some of the recent advancements in this area. The third section regards digitization standards, where probably that most relevant to MultiMatch is the development of JPEG2000. The final section concentrates on the impact of the Semantic Web and Web 2.0 on cultural heritage. There have been profound changes in user requirements since the delivery of D1.1.1. As users have found new sources of information, they have been introduced to tools that actively encourage or require user interaction. This has lead in general to a demand for improved searching functionality: better discovery through post-search filters (faceted searching), tag clouds and other visual search tools, improved displays, etc. The services offered by MultiMatch fit well into these new market trends

Search Engines

Chapter 3 discusses the state of the art for multilingual and multimedia search engines. In the revised version of this deliverable, a number of new initiatives in this field are listed, e.g. the Quaero and Theseus projects and the latest developments and intentions of Google Translate. The chapter concludes that MultiMatch's achievements relate well to the current state of the art. Rival multimedia search engines such as Theseus and the high-profile Quaero are still a long way from completion. Most multimedia searches rely on manually

generated meta-data, and those which don't have demonstrated a level of ineffectiveness. In fact, the current state of both multimedia and multilingual search still seems immature. The very few multilingual services available are limited in effectiveness and not particularly user friendly. Additionally, MultiMatch has introduced new features such as intelligent key-frame generation, and transcript searches that take the user to the appropriate place in the media file. These features are still far from common-place within other search engines.

Classification and Information Extraction

Classification (also known as categorisation) and information extraction are part of the knowledge discovery process, which attempts to find "interesting" patterns in data, i.e. those which reveal some underlying meaning (semantics). This chapter presents an extensive review of the state of the art in these two areas for text, images and videos. Much of this work has focussed on developing the pattern detection algorithms which detect the relevant features in the media type (i.e. words and phrases, textures and areas of interest, slots, etc.). Along with the computer science domain, and the world in general, possibly the most interesting challenge and opportunity facing researchers in this domain is the advent of the Internet and World-Wide-Web and in particular the increasing prevalence of Web 2.0 applications which encourage collaborative work with applications such as social tagging. The chapter concludes a number of possible directions for future web-mining including the use of multimedia and multilingual data, in addition the use of the "hidden web", i.e. the databases which are used to generate web pages from user queries, is seen as key. Within the MultiMatch project, the use of multimedia and multilingual data is obviously important and the use of structured data provided by the hidden web plays an important role in the use of information extraction to augment the metadata.

Multilingual/Multimedia Indexing

This chapter describes the state-of-the-art in the indexing of cultural heritage documents in various languages and of various media types. The special characteristics of cultural heritage documents are first described. General approaches to indexing currently being developed are then discussed and the specific approaches available for each different media type are presented. The chapter concludes by describing those areas in which MultiMatch has contributed to advancing the state of the art in multimedia indexing: structuring and indexing features for spoken audio, handling noise and processing audio from the internet, video classification, complex objects representation.

Image Collection Overviews and Browsing

Chapter 6 describes the development of image collection browsing and overviewing. This is motivated by the fact that such activities are complementary to search operations and may provide efficient solutions where search tools are deficient due to the lack of representative semantics within the documents. Initial evaluations of the work in MultiMatch pinpointed the need for complements or alternatives to the Query-by-Example paradigm. Deliverable D1.1.1 included an in-depth review of the latter. Del 1.1.2 thus proposed a review of browsing technique in a context close to or departing from retrieval. This overview was made with the view of evaluating browsing principles and technologies as useful in the context of MultiMatch. It has now been inserted into this final revised version of the SotA and a section has been added describing the advances within MultiMatch in order to ensure that the user is provided with a clear and efficient browsing strategy.

Multilingual/Multimedia Information Retrieval

The need to expand the scope of research in information retrieval (IR) beyond English text has been recognised in the last 15 years. Increasing amounts of work have been conducted and reported which explore non-English IR, cross-language information retrieval, multilingual information retrieval, and multimedia information retrieval. This work has greatly increased understanding of the issues of multilingual and multimedia information retrieval and access. A range of techniques have been proposed, explored, evaluated and refined. However, the techniques are imperfect and many challenges remain to improve effectiveness and to extend the scope of retrieval tasks. For example, significant issues arise with respect to translation between search topics and documents for cross-language and multilingual information retrieval. For multimedia IR, there are still problems related to the definition of retrieval units, i.e. what should we look for in an image or video, and the accuracy with which features can be detected automatically once they have been defined.

This chapter first provides a brief review of the relevant details and indexing assumptions of monolingual, cross-language and multilingual text IR. It then introduces multimedia IR and highlights some relevant experimental work. The final section looks toward future applications and challenges.

User-centred Interaction and Interface Design

The interface acts as the intermediary between users of information retrieval (IR) systems and the search system. The final chapter reports on studies of users' information seeking behaviour in order to provide informative insight into user interface design. The focus is on understanding the user needs in a dynamic multilingual search context, and identifying system functionalities that support those needs. Areas of relevance to the MultiMatch interface design include enabling the retrieval of multimedia objects (text, images, video, and audio) and then determining the best way of allowing the user to access this information (i.e. results visualisation). The interface should be interactive and adapt to meet a user's changing information needs. In considering interface design, an important first step is to examine functionalities currently provided by existing systems. Therefore, a brief summary of related systems and their features is provided. These include online museum collections, cultural heritage websites, multimedia search engines, and other systems designed by academic research projects. Innovative experimental approaches to aspects of interface design and results visualisation are also mentioned. Conducting such a survey provides an overview of current practice and provides a basis upon which MultiMatch can expand. By examining and testing a variety of designs with potential user groups, MultiMatch can endeavour to build an interactive, innovative interface that is first and foremost successful at meeting its users' needs. In this revised version, a new section has been added on Semantic web interfaces, reporting the new developments in this area.

Introduction

The objective of MultiMatch is to develop a multilingual search engine specifically designed for access, organization and personalised presentation of cultural heritage information. The development of the system thus implies addressing a number of significant research challenges in a multidisciplinary context. R&D expertise is required in a diverse set of system- and user-oriented research areas including, on the system side, focused Internet crawling, information extraction and analysis, multilingual information access and retrieval, multimedia complex object management, interface design, and, on the user side, user profiling, metadata and ontology studies, user/system interaction, interface design from the user perspective. The technology in these areas tends to develop rapidly. For this reason, it was decided to prepare a detailed State of the Art (SotA) report in the initial phases of the project, to be updated during and at the end of the project. This document has thus been released in three instalments (D1.1.1; D1.1.2, and the current document D1.1.3).

1.1 Structure and Contents

We originally identified six main areas: existing technology for cultural heritage; search engines; information extraction and classification; multilingual/multimedia indexing; multilingual/multimedia retrieval; user interaction and interface design. Each area was first reviewed in a separate chapter in D1.1.1, released in December 2006. A substantial update describing Image Collections and Browsing was added as a separate report in December 2007 (D1.1.2). In this final version, we provide updates to the original documents where appropriate[1]. Interestingly, one of the new areas which has been considered of importance to a number of authors is the advent of applications such as social tagging or other types of applications which leverage collective intelligence. From different perspectives, this emerging phenomenon is commented in Chapters 2, 4 and 8. In addition each chapter terminates with a section which relates the general state-of-the-art in that area to what has been done in MultiMatch. Our aim has been to provide a complete panorama of the actual state-of-the-art in the areas of interest to MultiMatch, covering as far as possible all relevant aspects.

In this Introduction, we summarise briefly the importance of these areas for MultiMatch. In the rest of the deliverable, each of these topics is discussed in detail. As is to be expected, there is some overlapping between the arguments treated in the different chapters. For example, the question of metadata is addressed in Chapters 2 and 4; but in each case from a different perspective. Similarly, indexing of multi-media data is discussed in both Chapters 4 and 5, with the focus of Chapter 4 on indexing for the purposes of information extraction whereas Chapter 5 is interested in indexing for the purpose of information access. Chapters 3 and 8 both talk about search engines, but while Chapter 3 describes the different types of existing search engines, Chapter 8 discusses the users' expectations and how they can interact with the functionality provided by the engines.

1.2 Technology for Cultural Heritage

A wide range of technologies are used in the different domains that can be classified under the general heading of cultural heritage. In the State of the Art deliverables we have focused on those of most direct interest for MultiMatch. In D1.1.1, we focused on metadata and encoding standards, and digital asset management systems. In Chapter 2 of the current document, attention is given to four main areas: digital library software; metadata interoperability; digitization standards; the impact of Web 2.0 on Cultural Heritage activities.

Of particular importance for efficient search and retrieval are decisions regarding the most suitable metadata schema(s) and conceptual reference framework(s) and consequent problems of interoperability over collections. The project recognised that content providers typically do not apply the same data model and conceptual schemas. However, it was felt that the schemas adopted for MultiMatch should contain all the

[1] The extent of revision with respect to Dels 1.1.1 and 1.1.2 varies considerably: Chapter 2 is completely modified, new sections have been added to Chapters 3, 4, 5 and 8, Chapter 6, released originally just nine months ago, contains just a few updates. The inclusion of Chapter 6 in this deliverable means that the chapters on Multilingual / Multimedia Information Retrieval and User Interaction and Interfaces have now become Chapters 7 and 8, respectively, instead of Chapter 6 and 7 as in D1.1.1

elements needed to describe the cultural heritage objects within the domain of the project. This chapter in D1.1.1 thus focused in particular on providing an overview of the technology and standards used in this area; a more in-depth description can be found in Deliverable 2.1 which provides a detailed analysis of metadata and ontologies in the cultural heritage domain. The final metadata schema decided on for MultiMatch is described in D2.2. Unfortunately, it has not been possible to exploit to the full the potential of this very powerful schema in Prototype 2 due to the variety and complexity of the information that should be derived from the CH documents acquired by MultiMatch in order to populate it completely. Despite this, we consider the MultiMatch metadata schema to be one for the important results of the Project activity.

There have been profound changes in user requirements since the delivery of D1.1.1 in December 2006. As users have found new sources of information (through services such as Google, Amazon and many others) they have been introduced to tools that actively encourage or require user interaction. This has lead in general to a demand for improved searching functionality: better discovery through post-search filters (faceted searching), tag clouds and other visual search tools, improved displays, etc. The services offered by MultiMatch fit well into these new market trends. The growing Web 2.0 movement comprises a suite of technologies for richer user experience, enabling users to easily provide their own Web content and using social-networking facilities. Any future development of MultiMatch should include functionality to handle and process user-generated content.

In the DL area probably the development that is most relevant to MultiMatch is that of Europeana, the European Digital Library, which will have it first public presentation in November 2008. Europeana has very similar objectives to MultiMatch, aiming at providing access to multimedia collections in many languages. Enjoying full collaboration with a very large number of national libraries in the European Union and beyond, Europeana will include books, films, photographs, manuscripts, and other cultural works. Initially access will mainly be to bibliographic documents, later it is hoped to include much full text. Europeana also intends to activate a cross-language search for a limited number of languages.

Metadata interoperability is key to ensuring access to heritage collections from various cultural heritage institutions. However, interoperability is hindered by the diversity of metadata formats and standards that exist in the cultural heritage domain. Chapter 2 discusses some of the recent advancements in metadata interoperability. D2.1 First Analysis of Metadata in the Cultural Heritage Domain documented an exhaustive list of metadata schema used in the CH domain. Since the finalization of this deliverable in October 2006, some new practices emerged, the main ones are listed. In particular, this chapter mentions the international OAI-ORE for archive interoperability: the Protocol for Metadata Harvesting (OAI-PMH) and Object Reuse and Exchange (OIA-ORE) and the issues involved with interoperability when using the Semantic Web technologies. In the MultiMatch project, OWL is used as a representation of the internal metadata model, which can also serve as a gateway to the Semantic Web.

With respect to recent trends regarding digitization standards probably that most relevant to MultiMatch is the development of JPEG2000. A MultiMatch partner, Alinari, has been involved in JPEF2000 discussions.

1.3 Focussed Search Engines

Chapter 3 discusses the state of the art for multilingual and multimedia search engines. In the revised version of this deliverable, a number of new initiatives in this field are listed, e.g. the Quaero and Theseus projects and the latest developments and intentions of Google Translate. The MultiMatch project has developed an advanced, domain-specific search engine which offers complex object retrieval through a combination of focused crawling, and semantic enrichment that exploits the vast amounts of metadata available in the cultural heritage domain. The major contribution of MultiMatch has been to provide a platform which integrates components for text, image, speech and video indexing and retrieval and provides an interface that permits the user to access and search simultaneously the collections provided over both media and language boundaries. Thus a single query can retrieve relevant documents that have been made available in text, image, video or spoken format and have been acquired either directly via agreements with CH institutions or indirectly through the MultiMatch CH targeted web crawler.

1.4 Information Extraction and Classification

Classification (also known as categorisation) and information extraction are part of the knowledge discovery process, which attempts to find “interesting” patterns in data, i.e. those which reveal some underlying

meaning (semantics). This chapter presents an extensive review of the state of the art in these two areas for text, images and videos. MultiMatch has used large scale information extraction from documents to identify entities and their relations in large Web corpora. This has made it possible to enable classification and clustering of documents according to their content. Documents have been classified on the basis of diverse dimensions, such as topical, geographical, and temporal. Much of this work has focussed on developing the pattern detection algorithms which detect the relevant features in the media type (i.e. words and phrases, textures and areas of interest, slots, etc.). Section 4.5 is new with respect to Del 1.1.1. This section comments on the challenges and opportunities facing researchers in this domain with the advent of the Internet and World-Wide-Web and in particular the increasing prevalence of Web 2.0 applications which encourage collaborative work with applications such as social tagging. A number of possible directions for future web-mining are suggested including the use of multimedia and multilingual data, in addition, the use of the "hidden web", i.e. the databases which are used to generate web pages from user queries, is seen as key. Within the MultiMatch project, the use of multimedia and multilingual data is obviously important and the use of structured data provided by the hidden web plays an important role in the use of information extraction to augment the metadata. The results of this approach can be most clearly seen in the Faceted Browsing in the MultiMatch User Interface.

1.5 Multilingual/Multimedia Indexing

This chapter describes the state-of-the-art in the indexing of cultural heritage documents in various languages and of various media types. The special characteristics of cultural heritage documents are first described. General approaches to indexing currently being developed are then discussed and the specific approaches available for each different media type are presented. The chapter concludes with a new section that describes those areas in which MultiMatch has contributed to advancing the state of the art in multimedia indexing: structuring and indexing features for spoken audio, handling noise and processing audio from the internet, video classification, complex objects representation.

1.6 Image Collections Overview and Browsing

Chapter 6 describes the development of image collection browsing and overviewing. This is motivated by the fact that such activities are complementary to search operations and may provide efficient solutions where search tools are deficient due to the lack of representative semantics within the documents. Initial evaluations of the work in MultiMatch pinpointed the need for complements or alternatives to the Query-by-Example paradigm. Deliverable D1.1.1 included an in-depth review of the latter. Del 1.1.2 thus proposed a review of browsing technique in a context close to or departing from retrieval. This overview was made with the view of evaluating browsing principles and technologies as useful in the context of MultiMatch. It has now been inserted into this final revised version of the SotA and a section has been added describing the advances within MultiMatch in order to ensure that the user is provided with a clear and efficient browsing strategy.

1.7 Multilingual/Multimedia Information Retrieval

For many years information retrieval research concentrated primarily on English language text documents. However, recent years have seen a significant increase in research activity extension to information retrieval techniques for multimedia and multilingual document collections. Unfortunately, so far, there has been little transfer of research advances to real world applications. MultiMatch aims at bridging this gap.

Multimedia data can be classified according to its constituent media streams: audio, visual and textual. Research in audio retrieval has largely been concentrated in speech retrieval (SR), where the key challenge is accurate automatic content recognition. Research in visual information retrieval (VIR) for images and video data streams has similarly been underway for over 10 years. Problems of VIR relate to both recognition of visual content and the definition of visual content for IR. Images and video key frames are most often indexed using low-level features such as colour and texture, or recognising named individuals or objects based on specific trained models. Research is now underway exploring the automatic recognition of shapes and their use in retrieval. The long-term challenges of visual retrieval cannot be overstated. Many multimedia data sources comprise a combination of audio and visual data with textual metadata labels. Thus multimedia IR often combines retrieval using these separate data sources.

Multilingual information retrieval (MLIR) has also become an established area of research in recent years. MLIR focuses on the problem of using a request in one language to retrieve documents from a collection in multiple different languages. MLIR also introduces the problem of how to select documents in languages for presentation to the user. A range of approaches have been introduced and explored in recent years.

The development of MultiMatch will require limitations of existing work in both areas to be addressed. A major challenge will be to merge results from queries on language-dependent (text, speech) and language-independent material (video, image). Retrieving documents from collections of mixed media also introduces problems of consistent ranking across the different media.

The CH material to be used in MultiMatch will have a high degree of heterogeneity covering many different topics, from a variety of different resources of differing linguistic forms as well as different media, and potentially published over a long period of time. Again, this introduces significant problems for high quality IR. For example, it has been demonstrated that using general translation resources for documents in a specific domain is much less effective than using ones specialised for this domain. A second key research problem for MultiMatch will be to identify the domain of requests and documents, and to build, and then to identify and exploit suitable translation resources for the domains within the CH collection. Documents will also be sourced in different media. MultiMatch will thus need to address significant issues of document selection arising from document language, media and topic.

1.8 User-centred Interaction and Interface Design

Although there has been huge progress, content-based information retrieval (e.g. video and image retrieval by visual content) still faces significant barriers when attempting to create truly effective and comprehensive retrieval with respect to the user's needs. Low-level features can be automatically extracted by analysing the audio and video stream, but human intervention is still needed to add high-level features (i.e. metadata). However, recent advances in the areas of information retrieval and information extraction make it possible to automatically associate concepts to objects when text is available. The need for human intervention to annotate material is thus reduced. The MultiMatch user interface integrates automatic techniques for low level feature extraction and automatic concept classification.

A key research problem for MultiMatch has been enabling the user to adequately formulate their query using the language of their choice and specify both low-level and high-level multimedia features.

According to a recent survey of search engine offerings, "despite the rapid growth of multimedia data that are available from the World Wide Web, current search engines have yet to provide an exciting, intuitive and user-centred set of the functionalities that support and sustain this phenomenon". MultiMatch has been working towards the aim of doing just this, following a user-centered approach to design access to multimedia material (with the unique addition of cross-language search as well.) Another main focus of MultiMatch has been on content aggregation and providing a global view to heterogeneous, distributed contents enhanced by semantic links. These features match those which Hyvönen [2007] mentions as ways in which the cultural heritage domain is well suited to the construction of semantic portals.

1.9 Summing Up

While MultiMatch has made headway into the exploration of a variety of topics relating to multimedia and multilingual information access and retrieval, unfortunately the scope and timescale of the project has meant that it was not possible to extensively investigate all areas. Future work inspired by the project could include, but is not limited to, developing tools for automatic language identification, annotation, translation, and correction of ASR output for multilingual videos; continued work with exploring the ways in which both experts and naïve users search for cultural heritage material (and ways of facilitating this); furthering knowledge of use cases of cross-language search in the cultural heritage domain; and further work with developing innovative image search and result interfaces (including multimodal search).

Technology for Cultural Heritage

by Johan Oomen

This chapter is intended as an extension and update to Chapter 2 of the original MultiMatch State-of-the-Art (SotA) document (D1.1.1). The aim is to define the scope of the technology now being used in the Cultural Heritage domain. This is not easy as it can include a very broad range of products and applications. We have focused on three main areas: digital library software; metadata interoperability; digitization standards; the impact of Web 2.0 on Cultural Heritage activities.

2.1 Trends in Digital Library Software

2.1.1 Commercial vendors update

In the original MultiMatch State of the Art document (D1.1.1), the following commercial Digital Library Software products were listed: IBM Content Manager, EMC Documentum, Autonomy, Corbis Media Management, Artesia, Oracle: Oracle Content DB. There have been major changes in the area of digital libraries since the delivery of D1.1.1 in December 2006. Notably, the rise of Autonomy, OCLC and Open Text (with the acquisition of Artesia and Corbis Media ManagemenT). The Norwegian based enterprise search vendor Fast was acquired by Microsoft Corporation early 2008. FAST technology will be integrated in Microsoft Search Server 2008 Express, and Search for Microsoft Office SharePoint Server 2007.

Together with these mergers and market shifts, there have been profound changes in the requirements for DL systems, many of them rooted in changing user expectations. As users have found new sources of information (through services such as Google, Amazon and many others) they have been introduced to tools that actively encourage or require user interaction. These tools encourage community building, plugging users into groups that share their interests or learning styles. Likewise, these new services have served to shine a spotlight on the library community and its information systems. Both the library community and its patrons have been able to see clearly how woefully unprepared our current integrated library systems are at present to participate in this very new user environment.

With this in mind, the Orbis Alliance Council[2] (consortium of libraries in the US) installed a working group to define **key dimensions of future DL systems**.

1. Improved searching: better discovery through post-search filters (faceted searching), tag clouds and other visual search tools, improved displays, etc. Likewise, results that provide more relevant results.
2. Better user experience: a more modern user experience ("Amazon-like"), with book jackets displayed, reviews, tagging, etc.
3. Same requesting functionality: equivalent to current abilities to request and borrow materials. In other words, a new catalog cannot be a step back in this area.
4. Syndication: a platform that supports pushing data to Internet search engines, desktop software, course management software, and other end-user applications, in order to integrate this data into the applications where users naturally work.
5. Developer friendly: a platform that supports and encourages interaction with the system. This can take many shapes, including OAI harvesting, SRU (Search and Retrieve by URL), OpenSearch or a simple web-services-based API to allow to take a more proactive role in developing services.

Our aim is not to be exhaustive here, but rather to point to the major ´movers and shakers´ in the area of enterprise search and asset management that became prolific after the release of D1.1.1.

[2] http://www.orbiscascade.org/index/about-the-alliance

OCLC and WorldCat

WorldCat is the flagship product of OCLC, the world's largest library service and research organization. OCLC has offices in the Netherlands, Australia, France, Germany, Switzerland, the United Kingdom and the United States and is a partner in the MultiMatch consortium.

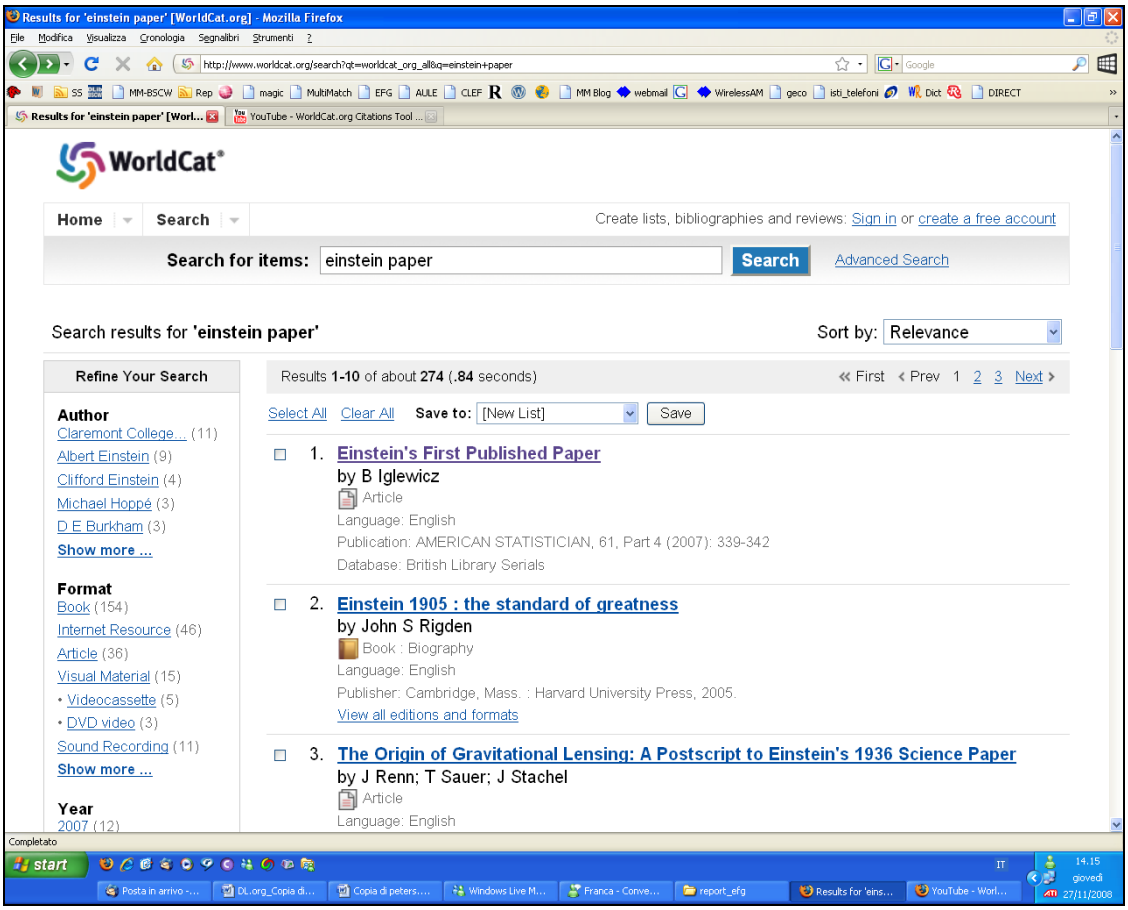


Figure 2.1: WorldCat Results Display

WorldCat is a union catalog which itemizes the collections of more than 10,000 libraries which participate in the OCLC global cooperative. It is built and maintained collectively by the participating libraries from more than ninety countries. Created in 1971, it contains more than 90 million different records pointing to over 1.2 billion physical and digital assets in more than 360 languages, as of November 2007. It is the world's largest bibliographic database. WorldCat itself is not directly purchased by libraries, but serves as the foundation for many other fee-based OCLC services (such as resource sharing and collection management).

The entire database is made available for search-engine harvesting. In August 2006, it became possible to search WorldCat directly through a central Web page at worldcat.org or through a downloadable search box.

ExLibris: DigiTool

The Ex Libris Group (headquartered in Israel) is a major provider of library automation solutions, offering a suite of products for acquiring, managing, and providing access to print, electronic, and digital materials for libraries of every type and size—from single-branch institutions to large consortia.

The open architecture and supporting interoperability standards make the ExLibris systems relatively easy to maintain and manage, and Unicode-compliant, with full multilingual capabilities. Three Ex Libris products are of particular relevance to MultiMatch: the DigiTool digital asset management system; Preservation, a large-scale digital preservation system being developed with the National Library of New Zealand for the preservation of cultural heritage; and the Primo end-user discovery and delivery tool that provides for materials of all types, regardless of the system of storage.

Autonomy IDOL

Autonomy is an enterprise software company with joint head quarters in Cambridge, United Kingdom, and San Francisco, USA. Autonomy's position as the industry leader is widely proclaimed and supported by analysts including Gartner Group, Forrester Research, and Delphi, which calls Autonomy the fastest growing public company in the space. The main technology is called Intelligent Data Operating Layer (IDOL), and is to unstructured information what an RDBMS is to structured information. IDOL allows search and processing of text, audio, video, and structured information. The processing of such information by IDOL is referred to by industry analysts (such as IDC) as the Meaning-Based Computing sector.

In May 2007 after exercising an option to buy a stake of technology start up, Blinkx Inc, and combining it with its consumer division, Autonomy spun out Blinkx Plc which was IPOed in London at a value over $250M.

Adlib Museum.

Adlib Information Systems3 (Maarssen, Netherlands) is a specialist software company with a history of more than 20 years of service to the library, museum and archive sector. The technology is widely used in the Netherlands, Belgium and the UK. Adlib Museum is a software application for managing collections and information in museums. Adlib Museum has been designed and developed by Adlib Information Systems and is based on many years' experience in museum and library automation.

Open Text Corporation a major provider of Enterprise Content Management software, its flagship project the Livelink ECM Solutions suite. Open Text has a strong customer base in the publishing, media and entertainment industry, with customers like HBO, 20th Century Fox, DreamWorks, Pearson Education. In 2005, it formed a special division targeted at the media an entertainment industry, the Artesia Digital Media Group.

In July 2008, Open Text acquired the eMotion Media Management Division of Corbis. This acquisition gives Open Text's Artesia Digital Media Group a broader portfolio of offerings for marketing departments and advertising agencies, adding capabilities that complement its industry-leading enterprise marketing asset management solution, Artesia DAM.

Interwoven[4], is an enterprise software company headquartered in San Jose, California, USA and founded in 1995. The company is mainly known for its content management system TeamSite, used to create complex intranet and Internet websites for enterprises. Interwoven claims to have over 4,200 organizations as customers. One of their Customers in the UK is the Natural History Museum. TeamSite provides an intuitive interface for content authoring, workflow, and archiving, allowing content authors, editors and reviewers to easily add, modify, and approve content..."

Vignette[5] is a suite of Content management, portal, collaboration, document management, and records management products developed by the Vignette Corporation, headquartered in Austin, Texas. Vignette V7 is the latest version of the Enterprise content management product. It consists of several suites of products allowing non-technical business users to rapidly create, edit and track content through workflows, and then publish this content through Web or portal sites. The appearance of delivery applications can be controlled via templates. Many large content-rich sites on the World Wide Web run Vignette. This includes Lexmark, Nokia, Wachovia, Time-Warner, Fox News Digital and the OECD.

Orange Logic provides a services for importing, indexing, distributing and selling digital pictures. Orange Logic claims to offer "the most versatile tool in the industry" Customers include Reuters (www.pictures.reuters.com) and smaller ones such as Art and Commerce (www.artandcommerce.com). http://www.orangelogic.com/

AquaBrowser[6] (Amsterdam, The Netherlands) is an indexing engine that resides outside the catalog and has a configurable user interface. Features include tag maps and a number of facets for filtering search results. AquaBrowser provides users with a graphical representation of search results, which are displayed contextually by topic or location. AquaBrowser only provides a search and retrieval system. Functions such

[3] http://www.adlibsoft.com/
[4] http://www.interwoven.com/
[5] http://www.vignette.com/
[6] http://www.medialab.nl/

as record display, shelf status, and requesting are handled by the underlying ILS but presented in the AquaBrowser interface. However, because AquaBrowser does perform its own indexing, the entire catalog data must be exported into the AquaBrowser software, which likely means that bibliographic data used by AquaBrowser will be up to 24 hours out of date, assuming export is done nightly. Customers include: King County Public Library (http://aquabrowser.toledolibrary.org/ aquabrowser/).

2.1.2 Open Source Software Suites

Next to the commercially available DL platforms listed above and in D1.1.1., several open source alternatives have risen to the surface in the past few yeas: Economic imperatives give designers of commercial systems a strong incentive to lock users in by preventing them from exporting the result of all their work, to increase the cost of migration to another vendor's product. This is a serious practical disadvantage. At least in principle, open source systems are immune because their code is accessible in source form and can be examined, understood, and modified by any competent programmer. In practice, however, substantial human investment is required to figure out just how to get the documents and metadata out of a digital library in a usable format. [Witten et al., 2005].

DSpace[7] and **Fedora**[8] are (still) the leading ones and were already described in D1.1.1. A third platform, Greenstone, is older and more established internationally.

Greenstone[9] is a suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License.

Key functionalities of Greenstone include:

- Design and construction of collections
- Distribution on the web and/or removable media
- Customized structure depending on available metadata
- End-user collection-building interface for librarians
- Reader and librarian interfaces in many languages
- Multiplatform operation.

Standard interoperability frameworks ~supported by Greenstone include OAI-PMH, which focuses on interoperability of metadata alone, and METS, which is a general framework that focuses on interoperability of document and metadata containers.

It needs to be noted here that DSpace and Fedora have a more impressive institutional pedigree in comparison to Greenstone.

OpenDLib[10] is a software toolkit that can be used to create a digital library easily, according to the requirements of a given user community, by instantiating the software appropriately and then either loading or harvesting the content to be managed. OpenDLib consists of an interoperable and communicating federation of services that implement the digital library functionality making few assumptions about the nature of the documents to be stored and disseminated. If necessary, the system can be easily extended with other services to meet particular needs.

The present version of OpenDLib provides a number of interoperating services that implement the basic functionality of a digital library, such as acquisition, description, storage, search, browse, selection and dissemination of documents, authorization and authentication of the users. This set of services is not fixed but can be extended to provide additional functionality. The OpenDLib services can be centralized or distributed and/or replicated on different servers. This means that an OpenDLib federation may comprise multiple instances of the same service type. Each service may require the functionality of other services in order to carry out its processing. In this case, a service instance communicates with the other service

[7] http://www.dspace.org/
[8] http://www.fedora.info/
[9] http:// www.greenstone.org/
[10] http://opendlib.research-infrastructures.eu/

instances via the OpenDLib Protocol (OLP). OLP protocol requests are expressed as URLs embedded in HTTP requests. All structured requests or responses are XML-based. OpenDLib has been developed by the DLib Group of the ISTI Institute of the Italian National Research Council – Italy, and will be shortly made open source accessible.

Nepomuk [11], or Networked Environment for Personalized, Ontology-based Management of Unified Knowledge, is an open-source software specification that is concerned with the development of a Social Semantic desktop. It is funded by the European Union. The project aimed to bring semantic information closer to the user, focusing on how it can help people find and structure information on their personal computers, and share that information with other users, instead of on the traditional area of how semantic information can be used on the Web. Nepomuk's desktop solution allows users to give meaning to documents, contact details, pictures, videos, and a variety of other data stored on a user's computer, regardless of file format, application, or language, making it easier and quicker to find information and identify connections between different items. When information is added, the Nepomuk software asks users to annotate the information so it can be correctly situated, and it also crawls the user's computer to search for information and establishes connections between different items.

2.1.3 Europeana: The European Digital Library

In an official press release, on 3 March 2006, the EC announced that the European digital library:

- Will build upon the infrastructure of The European Library[12].
- Will first encompass full collaboration among the national libraries in the European Union and gradually expand to archives and museums.
- By 2008 will include 2 million books, films, photographs, manuscripts, and other cultural works in a prototype.
- By 2010 will provide access to more than 6 million resources from every library, archive and museum in Europe.

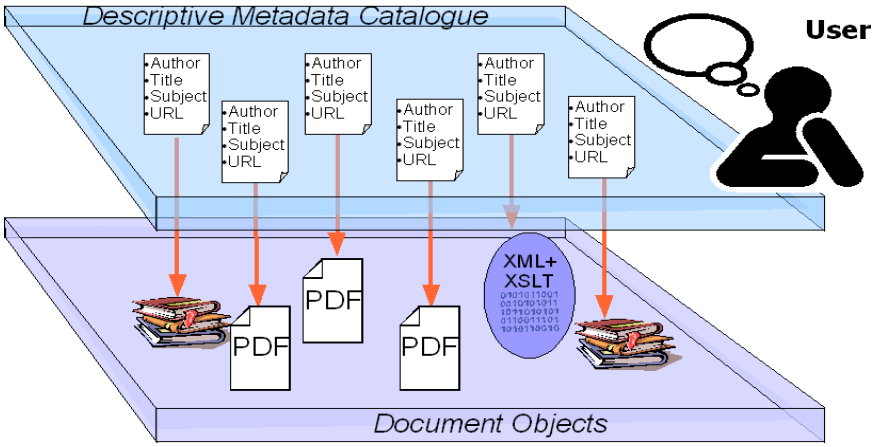
The EDLnet Thematic Network will release the prototype of Europeana and its final specifications, i.e. initial semantic and technical interoperability requirements for Europeana by the end of 2008.

After the launch of the prototype in November 2008, the EDLnet project's final task is to recommend a business model that will ensure the sustainability of the website. It will also report on the further research and implementation needed to make Europe's cultural heritage fully interoperable and accessible through a truly multilingual service. The intention is that by 2010 the Europeana portal will give everybody direct access to well over 6 million digital sounds, pictures, books, archival records and films**.**

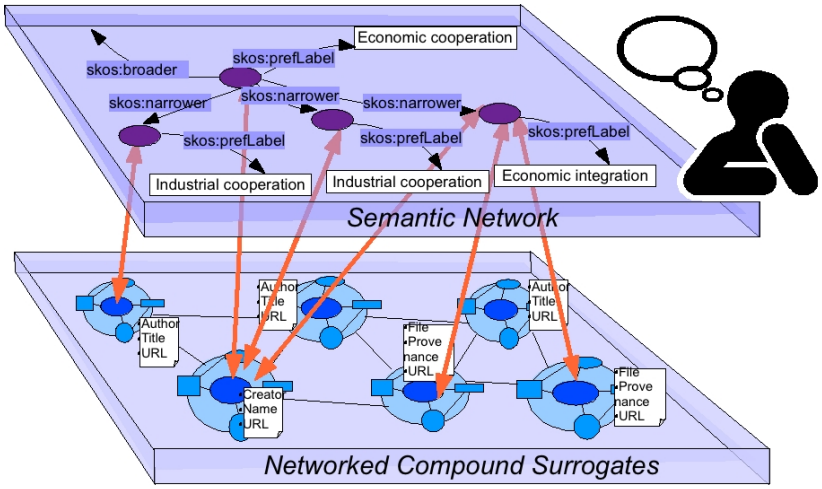
The foundation of the operational service is the EDLnet deliverable D2.5: Europeana Outline Functional Specification. A central principle for building Europeana is that a network of semantic resources will be used as the primary level of user interaction. [Dekkers et al., 2008].

The boxed text on the following page is extracted from this deliverable and outlines the Logical data model of Europeana. Figures 2.2, 2.3, 2.4 and 2.5 show the Europeana homepage, the Results display plus detailed view, and the proposed Timeline view.

[11] http://nepomuk.semanticdesktop.org/
[12] http://www.theeuropeanlibrary.org



Unlike in such librarian functional models users are expected to explore the Europeana data space using semantic nodes as primary elements for searching and browsing along paradigms indicated by the questions as to "Who?", "Where?", "When?" and "What?" The intended relation between the semantic and the object representation layers with respect to the Europeana user interface is illustrated in the figure below.



The user now primarily interacts with the semantic network to explore the Europeana surrogate space which now has the metadata as parts of the surrogates and surrogate aggregations.

In the perspective of this approach, Europeana can be thought of as a network of inter-operating object surrogates enabling semantics based object discovery and use. This network in turn is an integral part of the overall information architecture of the WWW.

Furthermore, the Europeana object model is based on the assumption that the central Europeana data store will only contain object surrogates and index files, whereas original objects are located at the content provider sites. Europeana thus will create a parallel data space inside the system that is a representation of the real world object space. As a consequence, we distinguish 'object entities' (to indicate an external object plus any associated metadata about that object) and 'surrogate entities' (to indicate the internal object with associated metadata and other composite elements). Likewise, two separate data spaces need to be distinguished: an external space of objects entities and an internal space of surrogate entities.

Figure 2.2: Europeana homepage (November 2008)

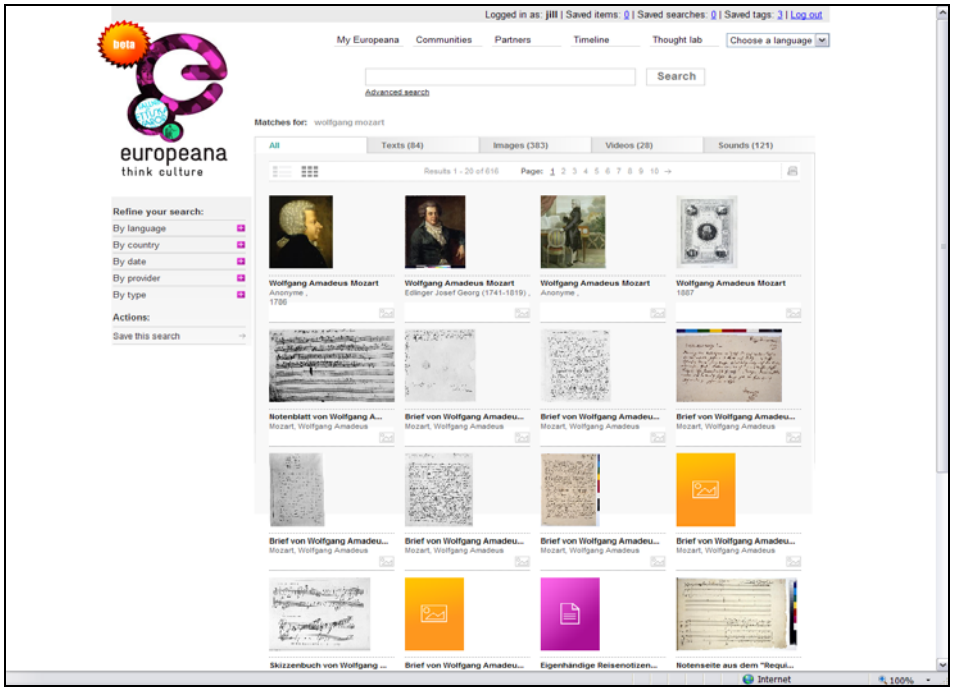


Figure 2.3: Europeana Result Display (November 2008)

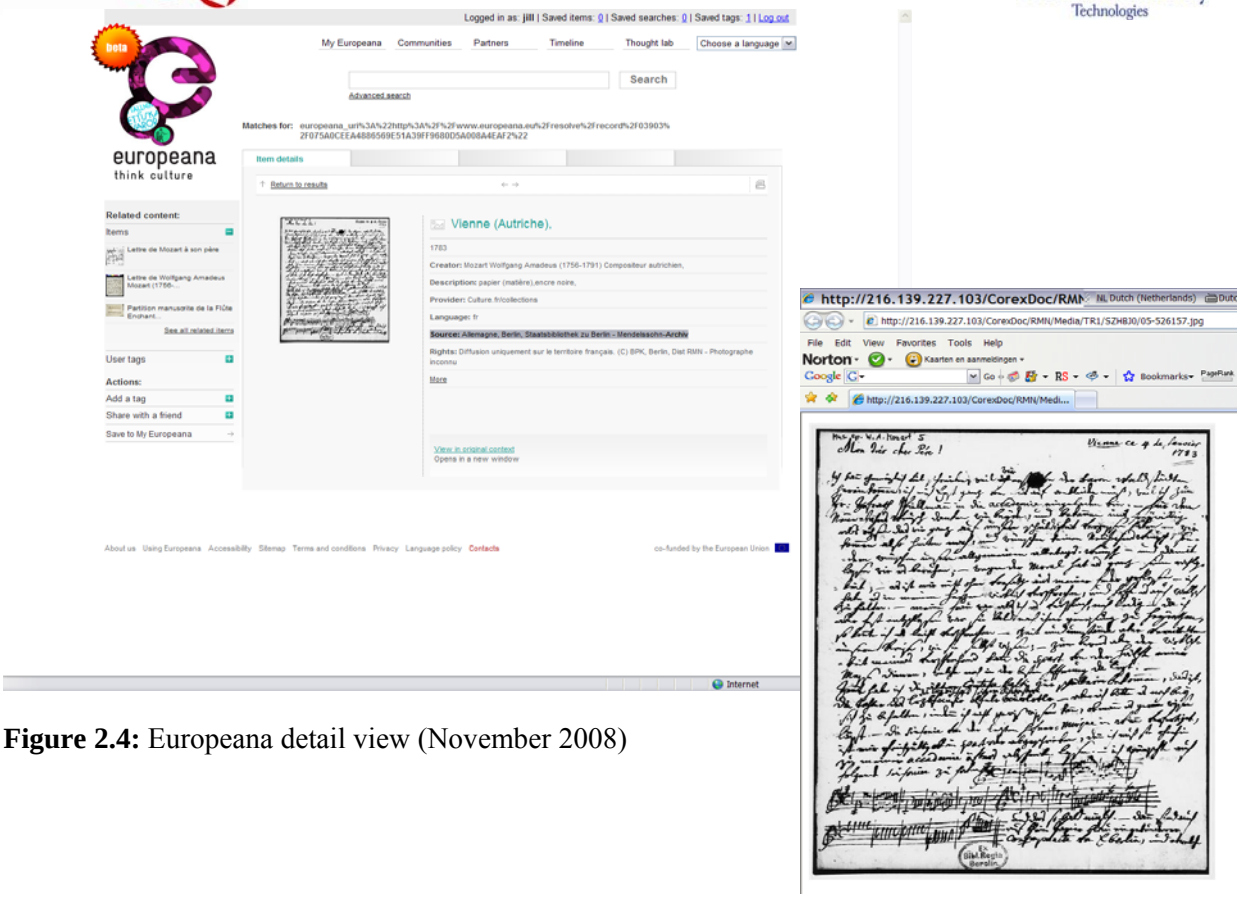


Figure 2.4: Europeana detail view (November 2008)

Figure 2.5: Europeana Timeline View (design July 2008)

2.1.4 European Research initiatives

MonetDB

MonetDB is a open-source database system for high-performance applications in data mining, OLAP, GIS, XML Query, text and multimedia retrieval. MonetDB often achieves a 10-fold raw speed improvement for SQL and XQuery over competitor RDBMSs. Software development is coordinated by the Dutch-based research organisation CWI. MonetDB achieves its goal by innovations at all layers of a DBMS, e.g. a storage model based on vertical fragmentation, a modern CPU-tuned query execution architecture, automatic and self-tuning indexes, run-time query optimization, and a modular software architecture.

Qviz

QVIZ was a two year project funded under the IST umbrella, and was finalised in May 2008. The overarching aim of QVIZ is to research and implement a time-map based search environment for archival information, and to build a collaborative environment for knowledge building within Communities of Practice (CoP). QVIZ basically comprises two main sets of features and functionality, which have been developed and integrated in the project:

- A time-spatial environment – map and time based query interface to gain access to archival resources.
- A collaborative environment – which gives users the possibility to add and share references, for knowledge building together with other users.

The time-spatial environment can perform searches in the archives from simple clicks on the map or from a faceted browser. The faceted query uses different types of data associated with archival resources organized into logical groups. The data used is both the administrative context and user activities in the archival resources. As a natural part of the faceted query, a timeline also gives users the opportunity to find specific information for a particular point in time. Most importantly, the user is given a way to access the document through linkages to the archival portal that holds the content.

Knowledge building within communities of practice is an emerging practice not yet adopted by the archives, but very much needed by the users as a tool for knowledge building in a user to user environment. In QVIZ's collaborative environment, the users can create content themselves, but also work together with different communities of practice. The resources created by users are an essential part of the QVIZ system. They form an expansive knowledge base, built through a web of links used to compile the research results. Users can create groups of references to a specific topic, articles relating to their subject of interest, and summarize content of specific archival references. All of these can be made public or shared with other users [QVIZ 2008].

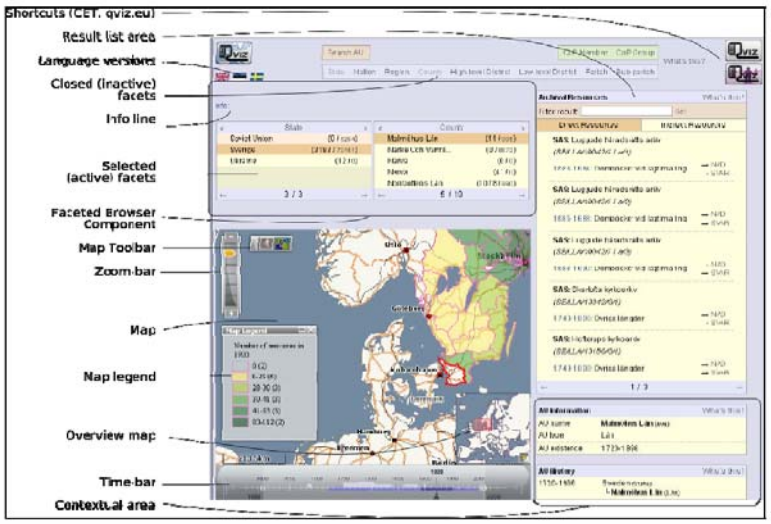


Figure 2.6: Screenshot of the Query Visualization Environment.

BRICKS

The BRICKS Integrated Project[13] (Building Resources for Integrated Cultural Knowledge Services) aimed at establishing the organisational and technological foundations of a Digital Library at the level of a European Digital Memory. A "digital library" in this context refers to a networked system of services over globally available collections of multimedia digital documents, providing a variety of knowledge layers for a variety of users and access modalities. The BRICKS vision was an integrated system that offers functionality for new generation of Digital Libraries, a comprehensive term covering "Digital Museums," "Digital Archives" and other kinds of digital memory systems.

The relevance to MultiMatch is twofold. First, BRICKS is relevant as an instance of distributed service oriented architecture (SOA) set up around Cultural Heritage. BRICKS is also relevant via its community as a focal contact point to an aggregation of professional and players in the CH field.

MICHAEL

The MICHAEL[14] project (Multilingual Inventory of Cultural Heritage in Europe) is facilitating access to cultural heritage information by providing an inventory on the digital collections by cultural institutions. Members of the public are able to use this inventory of the digital collections held by cultural institutions. Each member state runs its own cataloguing programmes with curators closely classifying MICHAEL records following a common ontology.

EASAIER

EASAIER[15] (Enabling Access to Sound Archives through Integration, Enrichment and Retrieval) allows archived materials to be accessed in different ways and at different levels. The system has been designed with libraries, museums, broadcast archives, and music schools and archives in mind. However, the tools may be used by anyone interested in accessing archived material; amateur or professional, regardless of the material involved. Furthermore, it enriches the access experience as well, since it enables the user to experiment with the materials in exciting new ways.

QUAERO

The QUAERO[16] program is a French governmental initiative focussing on the media content production and management chain with the objective to significantly facilitate access to and usage of multimedia content. Search is central to the programme, which will spend a significant effort into the development of very advanced and possibly disruptive technologies in the areas of audio, language, music, image and video processing as well as data coding, content protection, and high performance networks and storage organization.

THESEUS

THESEUS[17] is a research program initiated by the Federal Ministry of Economy and Technology (BMWi), with the goal of developing a new Internet-based infrastructure in order to better use and utilize the knowledge available on the Internet.

ENRICH

ENRICH[18] (European Networking Resources and Information concerning Cultural Heritage) is a targeted project funded under the eContentPlus programme. Its objective is to provide seamless access to distributed digital representations of old documentary heritage from various European cultural institutions in order to create a shared virtual research environment especially for study of manuscripts, but also incunabula, rare old printed books, and other historical documents. It builds on the Manuscriptorium Digital Library that has already managed to aggregate data from 46 collections from the Czech Republic and abroad.

[13] http://www.brickscommunity.org
[14] http://www.michael-culture.org
[15] http://www.easaier.org/
[16] http://www.quaero.fr
[17] http://theseus-programm.de/
[18] http://enrich.manuscriptorium.com

REVEAL THIS

REVEAL THIS[19] (Retrieval of Video and Language for the Home user in an Information Society) develops content programming technology able to capture, semantically index, categorise and cross-link multimedia and multilingual digital content coming from different sources, such as television, radio and the web. Users of the system will satisfy their information needs through personalized semantic search and retrieval, summaries of content and translation of them into their desired language.

BOEMIE

BOEMIE[20] (Bootstrapping Ontology Evolution with Multimedia Information Extraction - will pave the way towards automation of the knowledge acquisition process from multimedia content which nowadays grows with increasing rates in both public and proprietary webs, and will break new ground by introducing and implementing the concept of evolving multimedia ontologies. Driven by domain-specific multimedia ontologies, BOEMIE information extraction systems will be able to identify high-level semantic features in image, video, audio and text and fuse these features for optimal extraction.

PHAROS

The aim of the PHAROS[21] project (Platform for searcH of Audiovisual Resources across Online Spaces -) is to advance audiovisual search from a point-solution search engine paradigm to an integrated search platform paradigm. This platform will be built on an innovative, open, and distributed architecture that enables consumers, businesses and organisations to unlock the values found in audiovisual content.

2.2 Developments in Metadata Interoperability

Metadata interoperability is key to ensuring access to heritage collections from various cultural heritage institutions. However, interoperability is hindered by the diversity of metadata formats and standards that exist in the cultural heritage domain. MultiMatch identified over 40 well-established international standards in a recent survey.[Ireson et al, 2007] Metadata interoperability needs to be established on three levels:

- Syntactic interoperability: metadata can be accessed and processed in the same syntactic format, typically some XML format. RDF is the Web standard with an XML syntax designed for achieving syntactic metadata interoperability.
- Semantic interoperability: metadata can (partially) be interpreted within the same semantic frame of reference. Meaning of metadata of one archive (typically coded in in-house metadata vocabularies) needs to be linked with metadata from another archive. Thus, it requires alignment of archive vocabularies, which are partial as vocabularies differ in scope and perspective.
- Linguistic Interoperability: to allow retrieval across language borders. [Gradmann, 2008]

This paragraph highlights some of the recent advancements in metadata interoperability. D2.1 First Analysis of Metadata in the Cultural Heritage Domain has documented an exhaustive list of metadata schema used in the CH domain. Since the finalization of this deliverable in October 2006, some new practices emerged. Below, we list the main ones:

2.2.1 OAI-ORE

The Open Archives Initiative develops and promotes technologies for archive interoperability: the Protocol for Metadata Harvesting (OAI-PMH) and Object Reuse and Exchange (OIA-ORE) [Sompel and Lagoze, 2007]. OAI-PMH is a mechanism for repository interoperability that can be used to exchange documents according to any XML format as long as it is defined by XML schema. et al Sanderson 2005] OAI uses Dublin Core[22] (DC), the most widely used standard in the cultural heritage domain. The international OAI-ORE effort works towards a solution based on publishing Resource Maps that describe compound objects, referencing resources in their compound object context, and mechanisms to facilitate discovery of Resource Maps . Search and Retrieve by URL (SRU) is a protocol for XML-focused Internet search, which is among the protocols used for Europeana ,Chambers 2007].

[19] http://www.reveal-this.org/
[20] http://www.boemie.org/
[21] http://www.pharos-audiovisual-search.eu/
[22] DC Metadata Element Set: http://dublincore.org/documents/dces/

2.2.2 Convert thesauri for interoperability with the Semantic Web

There are two aspects of metadata interoperability with the Semantic Web. One is providing an interface for Semantic Web agents to access the content portals, the other is using Semantic Web technologies. The use of Semantic Web technologies is often proposed as a way of mapping between metadata schemes without defining specific converters or a "super-scheme". .,van Hage et al 2005] Different organizations are attempting to define standards for specific domains. The EC working group on digital library interoperability defines Semantic Web interoperability with the outside world as one of the goals. In the MultiMatch project, OWL is used as a representation of the internal metadata model, which can also serve as a gateway to the Semantic Web. [Ireson et al., 2007].

While the original vision of the Semantic Web - a layer on top of the current web, which annotates information in a way that is "understandable" by computers - is compelling; there are technical, scientific and business issues that have been difficult to address. One of the technical difficulties is the bottom-up nature of the classic semantic web approach; as each web site needs to annotate information in RDF, OWL, etc. in order for computers to be able to "understand" it. [Antoniou and van Harmelen, 2004]. The essence of a top-down semantic web service is simple - leverage existing web information, apply specific, vertical semantic knowledge and then redeliver the results via a consumer-centric application[23].[Iskold, 2007]

Thesauri are useful for indexing and retrieval on the Semantic Web, but they are often not published in RDF/OWL. Moreover, different organisations use different thesauri. et al Hausenblas 2007, Ireson et al 2007]. A structured method is required to convert thesauri to RDF for use in Semantic Web applications and to ensure the quality and utility of the conversion. Moreover, if different thesauri are to be interoperable without complicated mappings, a standard schema is required.

The Web standard Simple Knowledge Organisation Systems (SKOS) is attractive because it offers syntactic interoperability (through RDF) as well as a limited form of semantic interoperability through its predefined semantic vocabulary relations.

2.2.3 Reference models CIDOC CRM and Getty Crosswalks

Next to interoperability through semantic links between the vocabularies, other approaches towards semantic interoperability include adopting reference models and creating metadata crosswalks.

Reference models. In the eCHASE project, the CIDOC Conceptual Reference Model is employed (a description can be found above). In particular the recent CRM Core proposal is being used as the common model for different multimedia collections.[24] CIDOC CRM has been in development over the last ten years by the museum documentation standards group CIDOC and is in the process of ISO standardisation. CIDOC CRM is becoming increasingly used in the cultural heritage domain. It is capable of modeling the complex objects and relations within its scope, and can be extended to cover many specializations [Sinclair et al., 2005]. The CIDOC CRM community announced in May 2008 that is will strengthen efforts to reach out to cultural heritage organizations, as the model is still perceived as being too complex.[25]

Getty Crosswalks. The Getty Research Institute has produced charts that map several important metadata standards to one another, showing where they intersect and how their coverage differs[26]. Each of these standards can be said to represent a different "point of view" while Categories for the Description of Works of Art provides broad, encompassing guidelines for the information elements needed to describe an art object from a scholarly or research point of view, Object ID codifies the minimum set of data elements needed to protect or recover an object from theft and illicit traffic.

The Network Development and MARC Standards Office of the **Library of Congress** issued a crosswalk between the metadata terms in the Dublin Core Element Set and MARC 21 bibliographic data elements in April 2008. The crosswalk can be used for conversion of Dublin Core metadata into MARC, for instance as a tool for developing XSLT transformations.[27]

---

[23] Examples include: Spock (www.spock.com) Open Calais (http://www.opencalais.com/) and Twine (www.twine.com)
[24] http://eprints.ecs.soton.ac.uk/11567/01/echase.pdf
[25] http://lists.ics.forth.gr/pipermail/crm-sig/2008-May/001113.html
[26] http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html
[27] http://www.loc.gov/marc/dccross.html

The Repositories Research Team Wiki is a useful resource to identify additional mappings and crosswalks. The wiki is aimed at experts in the field of digital repositories. The wiki is maintained by the Repositories Research Team at UKOLN and JISC CETIS.[28]

2.2.4 Atom and tx metadata

Recently, **the Atom Syndication Format** is being used (notably by OAI ORE) for search and retrieval purposes.

Atom is comparable to RSS, but in contrast to RSS, which is really a collection of specifications and de facto practices, Atom is well defined by a single specification, RFC 4287. The primary use case that Atom addresses is the syndication of Web content such as weblogs and news headlines to Web sites as well as directly to user agents, as so-called 'feeds'. A feed contains entries, which may be headlines, full-text articles, excerpts, summaries, and/or links to content on a web site, along with various metadata.[29]

The Atom:feed markup element "is the document (i.e., top-level) element of an Atom Feed Document, acting as a container for metadata and data associated with the feed. Its element children consist of metadata elements followed by zero or more atom:entry child elements. The atom:entry element represents an individual entry, acting as a container for metadata and data associated with the entry. This element can appear as a child of the atom:feed element, or it can appear as the document element of a standalone Atom Entry Document." In addition to common attributes, an entry's defined elements include: atomAuthor, atomCategory, atomContent, atomContributor, atomId, atomLink, atomPublished, atomRights, atomSource, atomSummary, atomTitle, and atomUpdated. Atom is used by Google, Apple and many many others.

OAI-ORE (see above) introduced the notion of a Resource Map, which is a specialization of a named graph that asserts a finite set of resources (the Aggregated Resources), their types, intra-relationships, and relationships with resources external to this finite set (the external resources). A Resource Map Document is a machine-readable representation of a Resource Map. A Resource Map Document can be serialized in different formats, and the purpose of this document is to specify a serialization based on, and compliant with the Atom syndication format. Hereby, a Resource Map Document is an Atom Feed Document with some ORE-specific ingredients. This Atom-based format to serialize Resource Map Documents may be referred to as the Resource Map Profile of Atom.[30]

The metadata element set defined by Atom is reasonably similar to the Dublin Core element set. Informal guidelines exist for limited interconversion between them.[31]

tx metadata is established as a metadata standard for online video with the aim to 'ensure common definitions for basic information such as title, date, author and language and (free) tags. This standard is to be used in video upload forms and video feeds of data coming from each participating site. The standard will allow creation of search and importation tools for (open source) Content Management Systems (CMS) like Drupal, Wordpress, Plone/Plumi etc to easily locate video data in other video databases that use the standard. The first stable version of the standard was released in June 2008.[32] tx metadata is builds on Atom.

Tx metadata standard authors Jamie King and Jan Gerber write *"Because the Atom standard looks toward a future in which it will be adopted by a community of video producers, we consider it appropriate for adoption in the tx: standard. This may seem controversial, especially bearing in mind our 'real world' principle. However, on balance, the case for adopting Atom over RSS is fairly strong. We recognize the shortcomings of Atom: while each of the various web syndication feed formats has attracted enthusiastic advocates convinced of the capabilities of their respective formats, no one would dispute that RSS predominates. But given most video producers do not currently mark up their content in any coherent fashion, the fact that Atom is the best way to create a rigorous, clear and consistent framework for marking up video metadata, means we think it should be used." [King and Gerber, 2008]*

---

[28] http://www.ukoln.ac.uk/repositories/digirep/index/FAQs#Where_can_I_find_metadata_mappings_and_crosswalks_for_difference_metadata_standards.3F

[29] A complete listing of Atom elements can be found online at http://www.atomenabled.org/developers/syndication/atom-format-spec.php#rfc.section.4

[30] Further reading: http://www.openarchives.org/ore/0.1/atom

[31] See: http://intertwingly.net/wiki/pie/PaceEquivalents)

[32] http://wiki.transmission.cc/index.php/Metadata_working_group#Aims

2.2.5 PBCore and EBU Core

The PBCore (Public Broadcasting Metadata Dictionary) was created by the public broadcasting community in the United States of America for use by public broadcasters and related communities. Initial development funding for PBCore was provided by the Corporation for Public Broadcasting. The PBCore is built on the foundation of the Dublin Core (ISO 15836), an international standard for resource discovery (http://dublincore.org), and has been reviewed by the Dublin Core Metadata Initiative Usage Board. The first version of PBCore v1.1 XML Schema Definition (XSD) was released in February 2007.(see http://www.pbcore.org/announcements.html#quickstartupdated).

Over the last few years the European Broadcasting Union (EBU) and its members have been identifying the information required to search and exchange of content. The focus has been the definition of unambiguous radio and television media semantics (e.g. what is a 'programme', a 'media object', a 'title' or a 'bit-rate') and syntaxes proposing logical combinations of these descriptive elements. The main fruits of this effort to inclide EBU Tech 3293-2008 (EBU Core Metadata Set[33]. Tech 3293-2008 is based on the Dublin Core (DublinCore Metadata Initiative) and further refines the EBU core metadata set originally specified for radio archives in a previous version of Tech 3293 with a richer set of syntactically organised attributes. The need for a Dublin Core common base emerged from the requirements of archivists seeking a solution for easy search and retrieval (e.g. over Internet portals) and also for its capacity to interface with archive projects such as Europeana.

2.3 Recent Trends regarding Digitization Standards

2.3.1 Moving Images

In D1.1.1 – State of the Art, the following formats were listed: WAVE, Motion JPEG, H.264, MPEG-2, VC-1, D10. For preservation purposes, two formats seem to prevail: JPEG2000 and DVC. Regarding access, H.264 is gaining popularity.

Preservation: Motion JPEG2000 and the DVC-format
Motion **JPEG 2000** (MJ2), a video stream and file format, was standardized in 2002 as part of ISO/IEC'sJPEG 2000 (JP2) standard with subsequent refinements. This standard has been promoted by digital still camera manufacturers for its unified treatment of still and video compression. For stills, it is clearly of superior quality to its predecessor, JPEG, at any given compression. (ISO/IEC 2002) Motion JPEG 2000 (MJ2) is one potential format for long-term video preservation. The format is attractive as an open standard with a truly lossless compression mode.MJ2 applies JP2 compression to each frame independently. MJ2 is potentially attractive to video archivists not only because it is an open, international standard, but because it has a reversible, mathematically-lossless mode, not just the "virtually lossless" mode of certain other codecs. [Adams, 2002 and Li, 2003]

An alternative is to archive the **DVCAM stream**. This will result in an .avi file 'clone' of the original tape. It will be migrated lossless for preservation. The PrestoSpace wiki Migration Paths for Video Media provides a helpful overview of preservation scenarios.[34]

Access: H.264
H.264 is standard for video compression that has become popular quickly. It is also known as MPEG-4 Part 10, or MPEG-4 AVC (for Advanced Video Coding). It is one of the latest block-oriented motion-estimation-based codecs developed by the ITU-T Video Coding Experts Group together with the ISO/IEC Moving Picture Experts Group (MPEG). The final drafting work on the first version of the standard was completed in May 2003.

H.264 contains a number of new features that allow it to compress video much more effectively than older standards and to provide more flexibility for application to a wide variety of network environments. H.264/AVC experienced widespread adoption within a few years of the completion of the standard. It is employed widely in applications ranging from television broadcast to video for mobile devices. In order to ensure compatibility and problem-free adoption of H.264/AVC, many standards bodies have amended or added to video standards so that users of these standards can employ H.264/AVC. Both of the major

[33] http://tech.ebu.ch/MetadataSpecifications
[34] http://wiki.prestospace.org/pmwiki.php?n=Main.Roadmap

candidate next-generation optical video disc rival formats deployed in 2006 (HD DVD, Blue Ray) include the H.264/AVC High Profile as a mandatory player feature.

Discussions are often held regarding the legality of free software implementations of codecs like H.264, especially concerning the legal use of GNU LGPL and GPL implementations of H.264 and other patented codecs. Consensus in discussions is that the allowable use depends on the laws of local jurisdictions. If operating or shipping a product in a country or group of countries where none of the patents covering H.264 apply, then using, for example, an LGPL implementation of the codec is not a problem: There is no conflict between the software license and the (non-existent) patent license.

Ogg Theora is a open standard video codec, for compressing audiovisual media. It offers multiple qualities and resolutions (up until HD). The compressed video can be stored in any suitable container format. Theora video is generally included in Ogg container format and is frequently paired with Vorbis format audio streams.

The combination of the Ogg container format, Theora video and Vorbis audio allows for a completely open, royalty-free multimedia format. Other multimedia formats, such as MPEG-4 video and MP3 audio, are patented and subject to license fees for commercial use. Like many other image and video formats, Theora uses chroma subsampling, block based motion compensation and an 8 by 8 DCT block. This is comparable to MPEG-1/2/4. It supports intra coded frames and forward predictive frames but not bi-predictive frames

2.3.2 Photographs

HD Photo (formerly Windows Media Photo) was released by Microsoft three years ago and now (last meeting has been held in Poitiers on July 2008) the Joint Photographic Experts Group is going to evaluating and stating a JPEG standard with the name of '**JPEG XR**'. Due to the fact that JPEG XR is supported by Microsoft, we expect it to have a broad application and be used widely in the coming future.

HD Photo is a still-image compression algorithm. It is a file format for continuous tone photographic images. Both HD Photo and JPEG XR support lossy as well as lossless compression (due to the fact that the compression algorithm uses reversible transformations). Windows Vista and Windows XP need no plug-ins to edit the HD Photo files as this format is already included in the two OS libraries.

JPEG XR supports a wide range of colour encoding formats (for example: monochrome, RGB, CMYK and n-channel).. The JPEG XR is under standardization process and it has been designed with large contributions by Microsoft (USA) (through the experience matured with HD Photo) with a clear intention to optimize image quality and compression efficiency and at the same time enabling low-complexity encoding and decoding implementation. This new format offers the ability to decode only the information needed for any resolution or region, a key feature supporting Web imaging applications such as Windows Live Earth, and the option to manipulate the image as compressed data.

JPEG XR introduces support for High Dynamic Range (HDR) photography: providing benefits for both the capture and rendering processes for digital images and improving good imaging results. This is only one of the most challenging applications of the coming standard: we think that this standard can be a wide basis for camera raw formats. In fact, if standardized, the new JPEG XR file format will enable the next generation of digital photography to deliver better pictures with improved compression and more interoperable.[35] The design objectives include high performance, embedded system friendly compression, high compression quality; lossless or lossy compression; support for a wide range of sample formats, etc.

2.4 Cultural Heritage and Web 2.0

In 2006, Time Magazine targeted "You" as its Person of the Year [Grossman, 2006]. The web has become the tool for bringing together the small contributions of millions of people and making them matter. Tim O'Reilly coined the term "Web 2.0" in 2005 and attempted to capture its essence as follows: "Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them" [O'Reilly, 2006].

[35] http://www.microsoft.com/presspass/press/2007/jul07/07-31JPEGXRPR.mspx

In practice, Web 2.0 comprises a suite of technologies for richer user experience, enabling users to easily provide their own Web content and using social-networking facilities [Casey and Savatinuk, 2007]. The 'new web' also fostered the rise of a new economic mode of production, often called commons-based peer production. [Benkler, 2006]. This includes a range of collaborative efforts on the net in which a group of people engages in a cooperative production enterprise that effectively produces information goods without price signals or managerial commands. Examples include open source software and Wikipedia is the obvious example outside of software, but the phenomenon also includes collaborative efforts like the Open Directory Project and Slashdot. Some commons-based peer production efforts are less self-conscious on the part of the users, and emerge more as a function of distributed coordinate behaviour, like del.icio.us or Flickr.

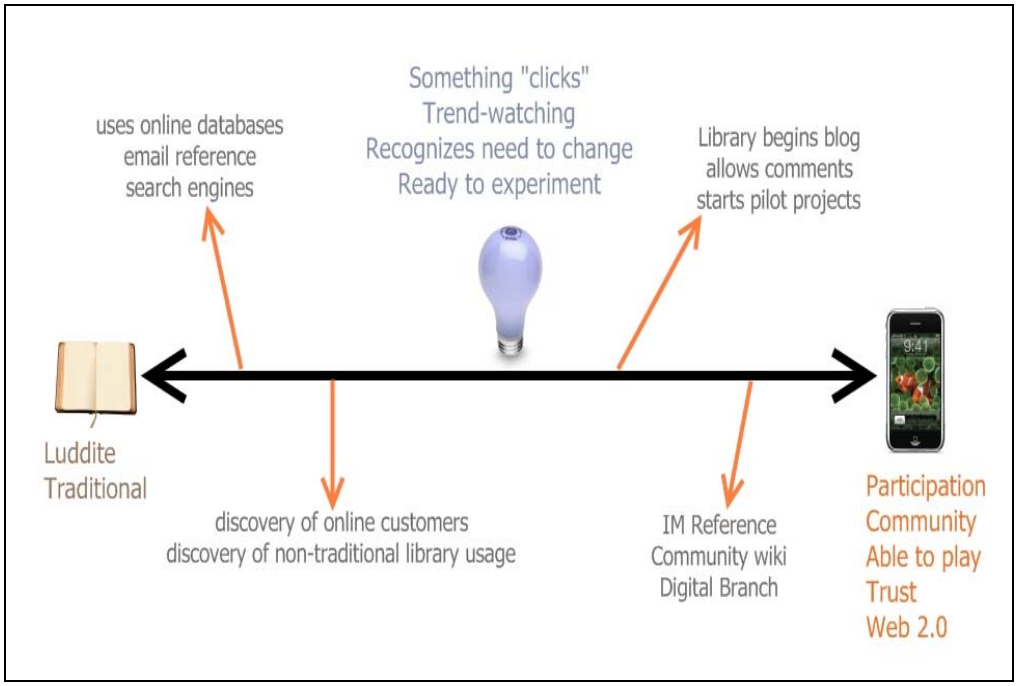


Figure 2.7: Library 2.0: Nina Simon

The cultural heritage sector is also taking advantage of the power embedded in Web 2.0 in many forms. The state of the art of 'e-culture' [Schwarz, 2004] is structured according to three clusters: Social networks, Content distribution and Crowd sourcing.

2.4.1 Social network services and software

A wide variety of social networks services have recently emerged (e.g., Myspace, Facebook, Flickr, Youtube, Ning etc). Many millions of people are spending a considerable amount of time on these websites. Museums are currently exploring how their visibility can improve by linking to and being present at these networks. Nine leading art museums have started a project, ArtShare, where everyone can select works from their collections and have those displayed on their Facebook profile. It is possible for users to add own artworks as well [Berstein, 2007]. More and more museums have Myspace page [Brooklyn Museum, 2008]. Cultural heritage organizations are also exploiting the use of social software, like blogs, Virtual Worlds [Tech Virtual Museum, 2008] and RSS feeds. Also, websites offer podcasts [Global Museum Podcast Directory, 2008] where users can listen to curators, conservators and researchers telling stories about works of art and widgets [Rijksmuseum, 2008] that allow users to view a work from the museum collection in their desktop. Finally, nice examples of online games created in the cultural heritage domain can be found on the website of the Exploratorium, the 24 Hour Museum and the Virtual Museum of Canada.

2.4.2 Content distribution and mashups

Markets are conversations "with the new economy moving from passive consumers to active prosumers" [Levine et al, 1999; McLuhan and Nevitt, 1972]. On the web, this new market has resulted in abundance of

online user-generated content. Ranked amongst the most prominent outlets are eBay, YouTube—streaming 100 million video's daily [Kirkpatrick, 2006], Flickr—serving over 3 billion photo's[36] Cultural heritage organizations are active on all of these platforms. The Bath Postal Museum for example has a store on eBay. Audiovisual Archives such as Sound and Vision in the Netherlands and the French INA have their own channels on YouTube. Large organizations like the Library of Congress and the National Museum of Public Health are teaming with Flickr [Raymond, 2008] to provide better access to their collections. These new channels are really successful to provide access to museum holdings over new channels.

Going one step further, some archives are allowing users to download their material and promote the creation of mashups. Relevant examples in the audiovisual domain include Prelinger Archives in the US and the Creative Archive in the UK. The (legal) availability of such sources is often limited due to intellectual property rights issues. Alternative rights models like Creative Commons are beginning to find their way in the cultural heritage domain enabling users an extremely interactive and rewarding experience [Hoorn, 2006]. The European Commission supports the COMMUNIA network that studies existing and emerging issues concerning the public domain in the digital environment [Communia, 2008]. It is expected that more and more organisations will use open licences [Hatcher, 2007].

2.4.3 Crowd sourcing and semantic tagging

The above mentioned collaboration between Flickr and the Library of Congress is a good example of a phenomenon dubbed as "crowd sourcing;" where users are invited to add additional metadata to objects in their collection; information that can be added to existing catalogue descriptions. [van Hooland, 2006; Bearman and Trant, 2005]. Quite a few initiatives are looking at how knowledge of users can be exploited, including the Powerhouse Museum Bulk Tagger [2008], Steve Museum [2008], LibraryThing [2008], Tate Your Collection [Tate, 2008] and Lignes De Temps from the Centre Pompidou [Puig, 2007]. The Europeana portal will also feature the ability for users to create tags and share tags. The most popular website however, this that of Flickr The Commons. Flickr the Commons invites heritage institutions across the globe to share their pictures with the Flickr Community, and offers all users the opportunity to add tags and comments in order to create more knowledge on the images. Below are two screenshots the Dutch National Archives on Flickr. The National Archives published around 500 images, which were viewed over 600.000 times in just five weeks time. Users left 1.200 tags [Oomen 2008].

Heritage organizations have to examine the reliability of the user-generated contributions, as Web 2.0 celebrates the "noble amateur" over the expert [Keen, 2007]. This dynamic is the centre of a debate that is going on in the library world between experts that feel tagging by non-experts can never be of added value to existing catalogue entries. The main reason is that traditional cataloguing description has the authors' intent as the leading principle, whereas free tagging allows multiple equivalent viewpoints. They fear that free tagging thus will eventually make 'meaning' relative [Hidderly and Rafferty, 2005]. Followers of the free tagging movement agree, but they do not think this is a bad thing. They acknowledge that every user will tag the same object slightly, or even completely different. The result may seem like a mess but it isn't, because automatic clustering and filtering techniques will create meaning and context [Weinberger, 2007; Bowker 2000].

The utility of multiple views is debated by some, but seldom studied in-depth in a realistic use case. This is an emerging area of research, often described as semantic tagging. [Fountopoulos 2007].

There is, at least, a trade-off between high quality but low quantity of description versus high quantity but perhaps low quality of descriptions. Hence, user-generated descriptions are instrumental in discovering hidden gems in the long-tail of CH. We will devote special attention to dealing with subjectivity. First, what is the reliability of more subjective information in user-generated content? Whereas individual authors may be less reliable than the traditional expert views in authoritative heritage descriptions, the consensus views of multiple authors can be surprisingly reliable.

[36] http://www.doeswhat.com/2008/11/06/flickr-3-billion-photos/

Figure 2.8: Flickr the Commons

A case in point is that the accuracy of Wikipedia is close to that of Britannica [Giles, 2005]. Second, how reliable is the tag and link structure on the Web? Whereas freely assigned tags (i.e., folksonomy) are highly subjective, the common tags assigned by multiple authors seem surprisingly objective, and the combined tag clouds give a far richer description of content than traditional controlled vocabulary systems. A case in point are the image labeling games that produce highly accurate descriptors of the content of visual information [Von Ahn and Dabbish, 2004].

2.5 MultiMatch and Moving beyond the State of the Art

The heritage partners in the consortium use well documented digitization standards for migrating their analogue material to digital files. In the way they handle metadata within their respective organization, they rather seem to follow in-house specifications or well established international standards (notably FRBR, Dublin Core, Mpeg-7 for structuring metadata and OAI for harvesting) than to adopt the latest advances in, for example, semantic web standards. This is exemplary for the cultural heritage domain. The ´innovation paradox´, the considerable time-gap between fundamental research and time it takes for practical uptake of technology is very prominent in the cultural heritage domain.

However, slowly but surely, we see a shift towards the adaptation of a new generation of library software amongst the features listed in Chapter 2.1.1. These include improved searching, better user experience, syndication and developer friendlyness by adtoping open standards.

For example, archives are beginning to convert their thesauri to the SKOS standard and are conducting pilots where metadata is represented in RDF. They start experiments (like with Flickr the Commons) that make use of the ´social web´. More and more, archives acknowledge that the only way to legitimize investments in

large-scale digitization efforts is to reinvent their relation with their (increasingly virtual and mobile) user base.

The Minerva EC working group states in their `Handbook on cultural web user interaction that a survey disclosed that 81% of the visitors to the "real" museums used Internet for work or pleasure and 22% of these had previously visited the website of the museum in which they were now for better planning their visit. The information is interesting if we consider that to the question "why did you not consult the museum website before visiting it?", 31% answered that the experience of visiting a museum is spontaneous and free and isn't planned, 28% answered that they already knew the museum very well and 21% got the information necessary through other media [Minerva 2008]. When Europeana.eu was launched in November, 1000 participating institutions jointly supplied access to a total of 4 million objects, located in digital repositories throughout the continent. These are just a few of many statistics that indicate that the web of the future will include a rich and diverse collection of heritage resources, to be reused in all imaginable ways, within an endless number of environments and by a great diversity of users. These objects come from trusted sources, that employ staff that performs quality control over everything that is made public under their supervision.

This is where MultiMatch improves the state of the art.

On the ingest site, what makes MultiMatch stand out is the fact that it can on the one hand handle manually created metadata (directly from archives or harvested using OAI) but also integrate this with information crawled from the web and created by the various extraction techniques. This difference in provenance can be handled by the MultiMatch metadata model, specific metadata elements (such as subject) receive a ´confidence value´ indicating whether this data is populated via ingestion of the content providers data (i.e. manually generated metadata) or via semantic enrichment. Also, the way the MultiMatch datamodel is able to handle dynamic data like feeds makes it stand out from other initiatives. The work on the metadata modeling and the evaluation performed within the first prototype is documented in `D2.2.2 Metadata schema and mapping´. This research will prove to be valuable in future projects in this area; notably because modeling dynamic data and (automatic) semantic annotation processes for efficient retrieval is not trivial. As the CH domain is a complex domain, the MultiMatch schema provides more options then the content and functionality as implemented in the second prototype, to keep options open. The MultiMatch ontology is available in the W3C standard OWL, so in principle the data can be linked to other sources, such as the Linking Open Data project.[37]

On the front end, MultiMatch shows how digitized cultural heritage material can be made assessable online, contextualized by information from Wikipedia, institutional webpages and so on.

MultiMatch offers a combination of: searching within a specific domain and support for multilingual searches. The multimedia search is based on similarity matching and on automatic information extraction techniques. The video search interface for example offers the possibility to easily search through a given programme using the outcome of the speech-to-text engine. Also, MultiMatch features metadata based search, where the user can select one of the available indexes built for a specific metadata field and can specify the value of the metadata field (e.g. the creator's name) plus, possible additional terms.

Concerning multilingual functionality in MultiMatch, users can formulate queries in a given language and retrieve results in one or all languages covered by the prototype (English, Italian, Spanish, Dutch, German, and Polish) according to their preferences. Six separate monolingual index files are maintained. Cross-language searches are performed by a combination of machine translation and domain-specific dictionary components. Users can select the source and the target languages as well as the most appropriate translation among those proposed by the system. This approach is groundbreaking as other sections of this report will evidence. It will provide a valuable input for the community that is developing Europeana.

The overarching concept of supporting the retrieval of cultural objects through different modalities is a major move beyond the state of the art. It offers an unprecedented, real life, vision of the way digital heritage will be made accessible in the future. This will be of importance to other initiatives and organizations in the area that share the vision of the project.

[37] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

References

Adams, Michael D. The JPEG-2000 Still Image Compression Standard, N2412, ISO/IEC JTC 1/SC 29/WG 1. (Dec. 2002).

Ahn, L. von and L. Dabbish. Labeling images with a computer game. In CHI'04: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 319–326. ACM Press, 2004

Antoniou, Grigoris and Frank van Harmelen, A Semantic Web Primer, The MIT Press, Cambridge, Massachusetts, April 2004

Bearman, D. and Trant, J. (2005). Social Terminology Enhancement through Vernacular Engagement – Exploring Collaborative Annotation to Encourage Interaction with Museum Collections. D-Lib Magazine, 11 (9), September 2005.

Benkler, Y. (2006). The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, 2006

Bernstein, S. (2007). ArtShare on Facebook! http://www.brooklynmuseum.org/community/blogosphere/bloggers/2007/11/08/artshare-on-facebook.

Bowker, G. (2000). Sorting things out : classification and its consequences. MIT Press, 2000.

Brooklyn Museum (2007). http://www.brooklynmuseum.org/community/blogosphere/bloggers/2007/11/08/artshare-on-facebook. November 2007. Accessed April 2008.

Casey, M. and Savastinuk, L. Library 2.0: a guide to participatory library service. Information Today, Inc., 2007.

Chambers, Sally. Towards Metadata Interoperability between Archives, Audio-Visual Archives, Museums and Libraries: What can we learn from The European Library metadata interoperability model?, 2007. EDL project D1.1ECP-2005-CULT-38074-EDL.

Dekkers, Makx, Stefan Gradmann and Carlo Meghini (2008). EDLnet D2.5: Europeana Outline Functional Specification For development of an operational European Digital Library.

Fountopoulos, G. I. (2007) RichTags: A Social Semantic Tagging System. Masters thesis, University of Southampton.

Giles, J. (2005). Internet encylcopaedias go head to head. Nature 438(15):900–901, December 2005.

Global Museum's Podcasting Directory (2008). http://www4.wave.co.nz/~jollyroger/GM2/podcasts.htm. Accessed April 2008.

Gradmann, Stephan. Digital Library Interoperability technical and object modelling aspects of Europeana. 2008, http://www.edlproject.eu/conference/downloads/EDLconf_Gradmann.pdf

Grossman, L. (2006). Times Person of the Year: You. 2006 http://www.time.com/time/magazine/article/0,9171,1569514,00.html.

Hatcher, J. (2007). Snapshot study on the use of open content licenses in the UK cultural heritage sector. Eduserv 2007.

Hausenblas, M., Bailer, W. and Mayer, H. Deploying Multimedia Metadata in Cultural Heritage on the Semantic Web, in First International Workshop on Cultural Heritage on the Semantic Web, co-located with the 6th International Semantic Web Conference (ISWC07), Busan, KR, Nov. 2007.

Hidderly, R. and P. Rafferty. Indexing multimedia and creative works. The problems of meaning and interpretation. Ashgate, 2005.

Hoorn, Esther (2006). Creative Commons Licenses for cultural heritage institutions: A Dutch perspective. Institute for Information Law (IVIR), Amsterdam, 2006.

Ireson, Neil. Johan Oomen, Hanneke SmuldersD2.2.1 Semantic Web Encoding: MultiMatch Knowledge Representation and Interoperability with Cultural Heritage Domain Standards. MultiMatch, 2007

Iskold, Alex. Top-Down: A New Approach to the Semantic Web. 2007. http://www.readwriteweb.com/archives/the_top-down_semantic_web.php

ISO/IEC Intl. Std. 15444, Information technology – JPEG 2000 image coding system, particularly Part 3: Motion JPEG 2000 (Sept. 2002, with subsequent amendments).

Keen, A. (2007). The Cult of the Amateur: how today's Internet is Killing our Culture. Doubleday, 2007.

King, Jamie and Jan Gerber. Tx metadata standard, June 2008

Kirkpatrick, M. (2006). YouTube serves 100m videos each day. http://www.techcrunch.com/ 2006/07/17/youtube-serves-100m-videos-each-day.

Levine, R., Locke, C. and Searls, D. (1999). The Cluetrain Manifesto: The End of Business as Usual. Perseus Books Group, 1999.

Li, Jin. Image Compression: The Mathematics of JPEG 2000, Modern Signal Processing 46, pp. 185-221. (2003).

LibraryThing (2008). http://www.librarything.com/forlibraries. Accessed April 2008.

McLuhan, M. and Nevitt, B. (1972). Take Today: The Executive as Dropout. 1972.

MINERVA EC Working Group (2008). Handbook on cultural web user interaction First edition. September 2008.

O'Reilly, T. (2006). Web 2.0 Compact Definition: Trying Again. 2006 http://radar.oreilly.com/archives/2006/12/ web-20-compact-definition-tryi.html.

Oomen, Johan (2008) Towards dynamic access to Audiovisual Archives. SAMT2008, Koblenz, December 3rd 2008.

Puig (2007). Lignes De Temps: Involving Cinema Exhibition Visitors In Mobile And On-Line Film Annotation. Centre Pompidou, Paris, 2007. http://www.archimuse.com/mw2007/papers/puig/puig.html.

QVIZ: Query and context based visualization of time-spatial cultural dynamics. Final Activity Report 2008

Raymond, m. (2008). My Friend Flickr: A Match Made in Photo Heaven. http://www.loc.gov/blog/?p=233.

Rijksmuseum (2008). Rijkswidgets. http://www.rijksmuseum.nl/widget. Accessed April 2008.

Sanderson, R., Young, J., & LeVan, R. (2005). SRW/U With OAI: Expected and Unexpected Synergies D-Lib Magazine 1(2)(Feb.2005)(http://www.dlib.org/dlib/february05/sanderson/02sanderson.html)

Schwarz, M. (2004). What is this thing called E-Culture? http://www.virtueelplatform.nl/article-324-en.html.

Sinclair, P., Lewis, P., Martinez, K., Addis, M., Prideaux, D., Fina, D. and Da Bormida, G. (2005) eCHASE: Sustainable Exploitation of Electronic Cultural Heritage (Poster). In Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, IEE Savoy Place.

Sompel, H. van de, C. Lagoze. ORE Specification and User Guide. http://www.openarchives.org/ore/0.1/toc

Stamou, Giorgos, Jacco van Ossenbruggen, Jeff Pan and Guss Schreiber. (2006). Multimedia annotations on the semantic web. IEEE Multimedia, 13(1):86--90, January-March 2006.

Tate Museum (2008). Write your own label. http://www.tate.org.uk/britain/writeyourown. Accessed April 2008.

Tech Virtual Museum (2008). http://thetechvirtual.org. March 2008.

van Hage, W. R., Katrenko, S. and Schreiber. A. Th.: A Method to Combine Linguistic Ontology-Mapping Techniques, 4th International Semantic Web Conference (ISWC-2005) Galway, Ireland, November, 2005

van Hooland, S. (2006). From Spectator to Annotator: Possibilities offered by User-Generated Metadata for Digital Cultural Heritage Collections. Université Libre de Bruxelles, 2006.

Witten, Ian H., David Bainbridge, Robert Tansley, Chi-Yu Huang and Katherine J. Don. A Bridge between Greenstone and DSpace. D-Lib Magazine September 2005

3. Vertical /Focussed Search Engines

by Carl Ibbotson with contributions from Marco Spadoni, Sam Minelli and Carol Peters

A search engine can simply be defined as a tool designed to retrieve information stored in some system. In the last decade or so, the web search engine has become of particular relevance and prominence, even an individual with the most modest of personal computer skills will be familiar with the search engines provided by Google[38] or Yahoo![39] These search engines allow users to request content from the World Wide Web that meets specific criteria by supplying a set of search terms, usually in the form of words or phrases. In this section, we briefly survey current search engine technology with particular focus on the areas of main interest to MultiMatch: domain-specific or vertical engines, engines specialised for multimedia and multilingual search and retrieval. We also give particular examples on the basis of the partners' own direct experience.

3.1 Generic Search Engines

All the major, current generic web search engines operate in a similar manner. General, broad-based engines aim to index as much of the World Wide Web as possible. They first crawl the web using automated software that follows every page link it finds. They then index and optimize this data into a database, and finally allow users of the search engine to submit queries to this optimized data.

A search results page is then returned to the user; this normally includes a list of web pages with titles, a link to the page and a short description showing where the keywords have matched the content. The popularity of Google's clean, unobtrusive interface and results page has influenced the design of other search engine interfaces, many of which look very similar.

3.1.1 Web Crawling

Due to the immense size of the World Wide Web, and limitations on both bandwidth and CPU time, crawling strategies become important. It has been noted that no search engine indexes more than 16% of the web[40] so choosing which pages to crawl, and when to crawl them are key decisions for a crawler.

Crawlers need to build a metric of importance for prioritizing pages on the Web. How this is done varies between providers. Often, crawling and indexing techniques and system architectures are guarded secrets, but all search engines employ some of the same basic methods. The importance of a page is a function of its perceived quality, and its popularity. Usually measured by how often the page is linked-to from other pages.

Due to the high rate of change of the Web, it is also crucial for a web crawler to sensibly determine how often to crawl a particular web resource. Typically, a crawler will employ a proportional update policy, meaning that pages that have previously demonstrated a high rate of change are generally crawled more often than pages that have shown a lower rate of change.

Large search engines such as Google, Yahoo! or MSN Live[41] have many thousands of machines positioned throughout the world that repeatedly crawl specific areas of the web, constantly providing new data to be indexed and stored. Web crawlers consume a huge amount of infrastructure and bandwidth, and are obviously expensive to run[42].

3.1.2 Indexing

Once web data has been crawled, it needs to be indexed. Different search engines do this in many different ways. Google, for example, indexes the entire page, or sometimes part of it, and often stores additional meta-data about the page, such as titles and headings. AltaVista's indexing strategy involves storing every text word of the page being indexed.

---

[38] http://www.google.com
[39] http://uk.yahoo.com/
[40] http://www.nature.com/nature/journal/v400/n6740/abs/400107a0_fs.html
[41] http://www.live.com/
[42] http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1196112&isnumber=26907

How data is indexed is crucial. One of the most important elements of a search engine is the quality and relevance of the results it returns. When a user enters some search terms, the engine refers to its index of data to provide a result set. There will often be millions, maybe billions, of indexed pages containing the search terms. Returning the most useful and relevant pages to the user is often how search engines are evaluated, and each search engine provider handles ranking the result set in many different ways. Google uses its patented PageRank algorithm[43] to determine the relative importance of a particular document. It works by assigning a numerical weighting to every page it crawls, determined by how often the page is linked-to from other pages. From Google's own website:

> *PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."*

Google's PageRank algorithm, and their extensive infrastructure means their web search engine generates high quality, well-targeted search results, enabling them to gain huge popularity amongst Web users[44]. Accuracy and quality of results appears to be the quality that users value most in a search engine[45], and Google users believe Google has the most relevant results[46].

Other well-documented ranking algorithms such as Hilltop[47] and TrustRank[48] work on related principles. Hilltop gives additional ranking weight to 'expert' sites, those that are built around an individual topic, and therefore gives weight to pages that are linked to from this site. TrustRank gives additional ranking weight to 'trusted' sites, which are selected by hand. Ask.com uses an algorithm based on HITS, which presumes that a good hub is a document that points to many others, and a good authority is a document that many documents point to[49]. Hubs and authorities exhibit a *mutually reinforcing relationship*: a better hub points to many good authorities, and a better authority is pointed to by many good hubs.

Many other search engines have implemented their own page ranking systems, however the workings of such algorithms are often held as company-proprietary secrets to prevent misuse and copying.

In recent years however, all search engines have had to contend with greater amounts of spam content. Consequently, ranking algorithms have become more and more critical, if companies such as Google and Yahoo! are to return relevant results to the user.

The problem often arises with auto-generated fake websites designed purely to exploit the page ranking system's rules, and push particular websites to the top of the search results page, or generate a network of websites solely to host contextual advertisements. It is estimated that currently one third of Google's index has been compromised by machine-generated sites.

While creating junk web pages is so cheap and easy to do, major search engine providers are engaged in an arms race with spammers. Each innovation to the indexing algorithms designed to bring clarity to the web is rapidly exploited by spammers, looking to harvest some classified advertising revenue.

There are signs that the search-engines are beginning to lose the fight. With so many irrelevant results polluting the search results page, Google recently took the approach of boosting Wikipedia results to the top of the search results page. By guaranteeing that a Wikipedia result (which is often relevant and reasonably reliable) appeared on the first page of results, Google has a cheap way of ensuring that it returns at least something of relevance to the original search terms. [50]

---

[43] http://www.google.com/technology/
[44] http://searchenginewatch.com/showPage.html?page=2156451
[45] http://www.seobook.com/archives/001316.shtml
[46] http://www.internetretailer.com/article.asp?id=16570
[47] http://pagerank.suchmaschinen-doktor.de/hilltop.html
[48] http://pagerank.suchmaschinen-doktor.de/trustrank.html
[49] http://www2002.org/CDROM/refereed/643/node1.html

[50] http://www.theregister.co.uk/2007/12/14/googlepedia_announced/

3.1.3 Searching

Once data has been indexed, it can be searched by passing keyword searches to it. Traditionally this has involved simple keyword searches, which are directly matched up to indexed pages and meta-data. AltaVista was the first search engine to allow more advanced queries by allowing the user to use quotation marks to search for phrases, or mark some keywords as mandatory.

Ask.com was an attempt to allow the user to build queries, posed in the form of a natural language question. Ask.com has often being criticised for generating low accuracy search results when compared to other leading search engines with more sophisticated page ranking methodologies, and its popularity has wavered in recent years.

For particularly common user search terms, search engines do not build the result-set afresh each time. Instead the search engine builds the result set once, and periodically refreshes it.

Additionally, most major search engines now offer their services though localised search engines For instance, on the Canada specific version of Google when a user searches for anything, the results will be of web sites with .ca domain extension

3.2 Vertical/ Focussed Search Engines

Vertical Search Engines work in a manner similar to the more broad-based search engines (such as Google and Yahoo!), however vertical search engine crawlers focus on highly refined pages and databases on the Web, and their indexes therefore contain more comprehensive information about specific topics in comparison to broad-based search engines.

Users of vertical search engines are often concerned only with results from a very specific niche (such as a medical database, or a job vacancy database), and are often unconcerned with the avalanche of data that accompanies a search performed on a broad-based search engine. For example, a Web user interested in buying a car would find far more relevant information from a niche search engine, such as Edmunds[51] than on google.com.

One of the problems of the more traditional, broad-based search engine is that the World Wide Web is growing at such an enormous rate, and pages are being updated so frequently that current search engine technology is struggling to continue to provide relevant, up-to-date result sets.

Additionally, a large part of the web remains impossible to index. The 'Deep Web' is a term, which describes sections of the Web that are not part of the 'surface web', and are therefore not able to be indexed. For example, dynamically generated web pages which act as search portals to specialised databases, or pages that are only accessible through dynamically generated links are considered to be in the 'Deep Web'. Because search engines can never link to these pages, they will never appear in search result sets. It is estimated that the Deep Web is several magnitudes larger than the surface web[52].

Vertical/Focussed search engines try hard to access the deep web by crawling it by subject category. Since traditional engines have difficulty crawling and indexing deep web pages and their content, deep web search engines like Alacra[53] (a business information search engine) create specialty engines by topic to search the deep web. Because these engines are narrow in their data focus, they are built to access specified deep web content by topic. These engines can search dynamic or password protected databases that are otherwise closed to search engines.

Because these focussed search engines are indexing specialised databases, with no public write-access, or very specific parts of the web, inaccessible to both spammers and broad-based search engines, problems with spam and auto-generated content are eradicated.

[51] http://www.edmunds.com
[52] http://www.press.umich.edu/jep/07-01/bergman.html
[53] http://www.alacra.com/

3.3 Domain Targeted Search Engines

The aim of the service (http://arianna.libero.it/news/) is to collect, from a set of Internet newspapers and web magazines, all the published articles and to show them to the final end-user, grouping them either by category (Politics, Economics, Sports, etc.) or by "event", i.e. grouping all the articles from different sources that are related with the same piece of news (including follow-ups).

The service is split in two main blocks:

- Data Management Service: an environment whose purpose is acquisition and management of news sources and retrieval and processing of articles. The environment can be thought of as a Web Service to which the data and their attributes are requested;
- Data Deployment Application: an environment whose purpose is querying the Data Management Service, and returning data to the final customer. The environment can be thought of as a Web application.

The most important stages of the pipeline constituting the DMS are:

- The Spider module, that repeatedly visits a list of news websites, several times a day, only retrieving relevant sections;
- The Extraction module, which is in charge of identifying and extracting interesting data (title, body, data, links to pictures) from unstructured pages. Identification is achieved by means of two orthogonal techniques:
  - » Manually crafted per-site sets of regular expressions, built and validated through a web-based user interface, and applied at run-time;
  - » Exploitation of anchor patterns in hub pages to address relevant data in the pointed leaf-pages (articles);
- The Categorization module which, after performing language normalization through a Natural Language Processing engine (tailored for the Italian language), associates each article with a category by means of self-updating Bayesian classifiers, initially trained on well known news sources;
- The Clustering module, in charge of grouping different articles dealing with the same event. This stage exploits the query-by-similarity functionality of the underlying full-text retrieval engine;
- The Indexer and Query Manager modules, build space- and time-efficient indexes and answering user queries.

End-users can make use of service data through the DDA environment in the following ways:

- By browsing static pages containing the most relevant news (in the service Home Page) or containing all the articles grouped by category (the articles Directory);
- By executing a standard, keyword-based query;
- By browsing a pictorial representation of the graph of the news, where nodes are entities (most frequently cited peoples, institutions, companies, cities etc.) and arcs are relationships witnessed by news articles mentioning (at least) two entities at the same time. A user click on a node redraws the graph centred around the selected entity, while a click on an arc returns all articles underlying the relationship between two entities.

When submitting queries, users can choose to sort returned articles either by reverse publication date, or by relevance. The relevance of an article is a function of

- Its affinity to the user query (standard keyword based scoring in title and body),
- Absolute score of the cluster to which the article belongs (a function of the number of articles in the cluster and spread of the cluster),
- Absolute score of the article (a function of the estimated precision of the categorization and importance of the site hosting the article)

A service very similar to Libero WebNews is Google News (e.g. http://news.google.it/). The features of the two services are very similar. However, whilst Libero WebNews currently provide the News Alert functionality only via RSS-feed (and not also via Email as Google News does), it is currently providing news from about 1180 news-sites, with respect to 250 sites claimed by Google News Italy.

3.4 Media Targeted Search Engines

Using text-based search engines to retrieve multimedia content has been simple: Meta-data, or 'tags' are assigned to pieces of multimedia, allowing them to become searchable using standard techniques. For example, youtube.com allows users to upload their videos to the Web and share them with anyone. Before a user uploads their video, they would tag it with appropriate meta-data; for example if they upload a video clip of a boxing match, they may tag it with the words 'boxing', 'fight', 'punch', or whatever other words they considered relevant to the clip. Search engines would search only the Meta data, and treat it as simple text.

There are many web-based search multimedia search engines that serve multimedia content in these ways, Flickr.com, BBC Audio Search, WIND and Google Video are some examples.

IBM's Marvel[54], a image and video search engine, works on a similar principle, but takes it a step further. It has the ability to analyze multimedia content and automatically generate meta-data for that content by comparing it to a library of semantic models.

3.4.1 Multimedia Search Engines

Under the heading of multimedia search engines, one should distinguish between search engines that retrieve multimedia data and those which accept multimedia queries. The first category describes engines that would return documents or pointers on documents of heterogeneous types understanding that the combination of their composing streams is an answer to the query (of any type). The second category is concerned with the form and formulation of the query. It may be interesting to formulate the query using different media. For example, this person (picture) saying something like this (audio and/or text).

While the distinction is interesting, search engines available in practice are of lower complexity. As mentioned above (Section 3.2) many search engines are focused on a single type of media and accept queries specific to that type. Queries are generally formulated using text. Text is not only the simplest media to manipulate and understand unambiguously, it is also the most accessible. A video search engine based on the query-by-example paradigm requires examples to be exhibited. These are not always easily accessible.

A number of search engines may still fall into our first category. These are generally information repositories where a navigation process has been enabled. This includes for example IMDB, the Internet Movie Database. Querying IMDB, one retrieves textual information (e.g. movie synopsis), video excerpts and summaries (trailers), pictures (making of) and structured information (actors, scenes, judgements). From there, Yahoo! Movies, and the INA TV archive can also be put into this category.

Most of the above relies either on structured manually created data (IMDB) or automatically inter-related data (Yahoo! Movies). Links are created over metadata, generally composed of text.

Looking at content-based search engines, all contributions essentially remain in the academic community as prototypes and applications rarely truly meet the general public. When doing so, functionalities are reduced and not engaged into a business process involving risk. This is the case for http://www.MyHeritage.com where one may find look-alike face picture of celebrities ("Find the Celebrity in You™") or Retrievr (http://labs.systemone.at/retrievr/) which allows to query-by sketch in the Flickr image collection.

Looking at academic prototypes, we may non-exhaustively list Gift, Vicode and Web<img>Seek (Univ. of Geneva), Muvis (TUT), Ikona (INRIA), WebSeek (Columbia Univ.), MediaMill (Univ. Amsterdam), Fischlar (DCU), Informedia (CMU), or MARVEL (IBM). The list may be extended by citing almost all participants of the TRECVid benchmark (http://www-nlpir.nist.gov/projects/trecvid) who did develop their own multimodal retrieval systems. These search engines use truly multimodal content-based information to achieve the search process. All are based on low level signal processing (image/audio), language processing, machine learning and data-mining to infer semantic content (both from documents and queries), annotate documents and organize multimedia collections into comprehensive information structures. It is worth noting that the vast majority of these systems take benefit from user feedback and interactions to enhance their performance. However, their performance remains below large public needs (see the last TRECVid

[54] http://domino.research.ibm.com/comm/research_projects.nsf/pages/marvel.index.html

Evaluation: http://www-nlpir.nist.gov/projects/tvpubs/tv.pubs.org.html). Moreover, the "intelligent" strategies involved generally prevent such systems from very large-scale (i.e. planet-scale) applications.

WIND/Libero Image/Multimedia Search Engine

A commercial example of an Italian text-based multimedia search engine is provided by the Libero portal which offers users the possibility of searching among a fair amount of images, MP3 files and videos gathered from the Italian web.

The harvesting strategy is very simple: the engine scans the complete data base of html pages used to build indexes for Italian text search, and extracts links apparently pointing to images, MP3 files or videos. Candidate objects are then fetched, scanned for known magic numbers to make sure they really represent the kind of object the referrer declared, and digests are computed on their contents to help avoid duplicates.

Objects that pass the test are included in the database. Images and videos are further processed to extract fixed-size thumbnails (for videos only the few KB needed to extract some frames are really fetched from the net), and then thrown away to avoid copyright issues.

Indexing considers only textual information associated with every object:

- URL and title of the referring page
- Text surrounding the link
- "File name" of the URL representing the object (the complete path is frequently used for injecting spam)
- Contents of the ALT attribute (where applicable)
- Internal tags (author, title) for MP3 files

Table 3.1: Search features and market availability.

Alinari Search features		
Textual query	Keyword based query	Free input keyword
		Selection from a predefined list
		Selection from a predefined thesaurus
		Selection from a high level ontology
	Natural language based query	Annotation based query: exact term query
		Semantic based query
Visual query	Visual similitude	MPEG7 low level descriptors (see SCHEMA project)
		Statistical analysis
	Visual semantic query	Object detection (see MultiMatch project)
		Environment recognition
		Person detection
		Face recognition
		Mood detection
		Place recognition
		Historical period recognition
Human memory based search	Textual	Dictionary based suggestion (see Google-suggest: "perhaps you were looking for…")
		Semantic Textual suggestions
	Visual	Similitude visual suggestions
		Semantic visual suggestions

User queries are term based. Ranking of results takes into account matches in all indexed fields. In the case of MP3 files, the user may ask to sort results by date, in order to get information on the freshest addition to the web.

An image search service is provided by Google, and is very likely based on the same technology, i.e. search in indexed text "surrounding" the pictures and users perform term-based queries. Of course the coverage is much broader.

Technologies deployed for MultiMatch could improve significantly both engines by allowing content-based retrieval and clustering of results.

Alinari Search Engine
Alinari is currently developing intelligent search features in their site query functionality such as concept suggestion (similar to Google's 'perhaps you were looking for…') and keyword gender independence (male/female) and singular plural independence connected to RSS features: the user sets actively his personal preferences in a tutored context.

3.4.2 Future of Multimedia Searching

Several new types of multimedia search engine are beginning to surface. These are engines that actually search the content of multimedia files, rather than just the meta-data associated with it.

Podzinger[55] is a search engine for podcasts. It allows users to enter text based search terms, and then returns a list of podcasts containing those words. It works by using speech to text technology to convert the audio podcast into a text stream. This text stream can then be indexed and searched in a manner similar to standard search engine techniques. It is therefore possible to locate a podcast based on any single spoken word from the podcast, rather than just a limited set of Meta data tags associated with it.

Blinx has used a similar approach[56], but rather than searching podcasts, Blinx attempts to transcribe and search web-based TV channels. Its effectiveness appears questionable at the moment.

Retrievr[57] is a novel image search engine that features an interface allowing the user to sketch simple pictures, or upload images of their own. These are then matched against Flickr's database of images, and, in theory, similar images are displayed to the user. Retrievr's results would appear to be a little flaky at this stage in development.

Other similar types of multimedia search engine include tv-eyes[58] and singing- fish[59]. A discussion of multimedia and multilingual search interfaces is given in Chapter 8.

3.5 Multilingual Search Engines

Most of the search engines mentioned so far search by simply matching up input search words to indexed meta-data. Searching for "cat" for example will only ever match up exactly to the indexed phrase "cat". Most search engines would then prioritize their results to the locale of the user, but this is not a true multilingual search. By multilingual search, we intend systems that provide both efficient and effective monolingual search but also functionality that permits to search over languages, i.e. entering a query in one language and retrieving results from target collections in different languages.

The development of multilingual search systems is still very much a research question and so far there has not been a lot of transfer of the research results into the application or commercial domains. An important source of literature with respect to the most recent research trends in this area is the website of the Cross Language Evaluation Forum (see www.clef-campaign.org). All the research institutions involved in MultiMatch are active collaborators in the CLEF activity. However, recently, the major search engines are making a concerted effort to implement cross-language functionality in addition to existing language-

[55] http://www.podzinger.com/
[56] http://www.blinkx.tv/
[57] http://labs.systemone.at/retrievr
[58] http://www.tveyes.com/
[59] http://search.singingfish.com/

optimised monolingual functionality. The last two years seem to have seen a reduction of interest in this area by Yahoo! whereas Google currently appears to be making a big investment in translation services.

Yahoo!

Yahoo has offered cross-language search in a few languages since 2005. Yahoo!France and Yahoo!Germany provided a basic multilingual search functionality. The user activated the "Recherche multilingue" or "Suche Translator" option and entered the query in his/her preferred language; the search results included not just the web pages written in that language, but also web pages written in other languages (French, English, German, Italian and Spanish). This functionality was made available in a beta (testing) version and was not particularly intuitive to use; it was also not clear how the results are ranked and no option is provided for specifying in which languages the search should be performed. This development was of course of great interest to MultiMatch. However, at the time of writing it appears to have been discontinued.

Google

In May 2007, Google added a useful new feature to its translation tool - Google Translate. This feature enables you to search for a keyword or search phrase in languages you don't speak yourself and get a quick overview of websites in other languages with the help of machine translation. The Google Translate's search results tool featured twenty three different languages by May 2008 (but the number is rising)- among them Chinese (both traditional and modern), Arabic and Russian. It was released in *beta* status, which means that although it is already publicly available, the tool is still being tested and will continue to be improved over time. Similarly to MultiMatch multilingual search in Prototype 2, search queries are entered in the native language, translated into English and run against Google's index. Any retrieved pages/sites will then be translated from English back into the native language

In August 2008 an extremely interesting development to this service was introduced with the announcement that Google is about to launch a beta test of a document translation service. With the service, the company will connect people who need documents translated with humans who will be paid to do so. The world's most comprehensive set of translation technologies will now be aided by human beings translating documents upon request. Google will offer volunteer and professional translators the opportunity to use Google tools and technologies to translate. In previous columns, we've discussed the need for localization in translation. Google Translation Center will enable users to upload a document, choose a translation language, and select from Google's registry of professional and volunteer translators. If a translator accepts, users will receive the translated content back as soon as it's ready.

As Google prefers to rely on computer algorithms rather than humans, at first glance the Google Translation Center looks somewhat anomalous. However, Google's translation system uses a statistical model that works better the more it can compare the same text in two different languages. The more documents Google has in two languages, the better able it is to match words and phrases from one language to another. By computing statistics over all words and phrases, you get a model of word-by-word and phrase-by-phrase replacements. Machine translation often produces awkward results today, but the impact of having a really large language model should make the sentences flow a lot more easily. The expectation is that Google is planning this service in order to be able to improve it's automatic MT procedures with the unknowing assistance of the world's expert translators.

Additional discussion of multilingual search systems can be found in Chapter 8.2

3.6 Recent Developments in Multilingual / Multimedia Search Engines

Since the MultiMatch project began in May 2006, there have been some important developments in the multilingual / multimedia search domain with the launching of several very ambitious European projects. Here below, we cite the three that we feel are the most significant.

Quaero and THESEUS

The activity of the Quaero and the THESEUS programmes is of great interest to MultiMatch. The Quaero search engine was first announced by Jacques Chirac during the French-German ministerial conference of Reims in April 2005, and was set up by the German and French Economic Affairs Ministers, specifically by the "research and innovation" sub-group. However, since the field covered is extensive and given the differing perception by the two consortia of the thematic priorities, after considerable discussion and delay, it was decided to launch two independent programmes, Quaero and Theseus. Quaero has retained the Franco-

German dimension in that the programme involves German research enterprises and bodies aided by France[60]. In addition, the teams of the Quaero and THESEUS programmes have agreed to maintain a consultation structure and to collaborate on a case-by-case basis when the opportunity arises

Quaero[61]. Following the European Commission approval in March 2008, the Quaero consortium's research and development program is set to receive 99 million euros aid from the French government. Consortium members will contribute an equivalent amount to reach an overall budget of approximately 200 million euros for innovative research projects.

This decision allows the launch of the collaborative research and development program focusing on the areas of automatic extraction of information, analysis, classification and usage of digital multimedia content for professionals and consumers. The research work will concentrate on managing virtually unlimited quantities of multimedia and multilingual information, including text, speech, music, image and video. The industrial partners of the consortium will build business plans leveraging the technologies and tools which will have emerged from the research and development.

The Quaero consortium was created to meet new multimedia content analysis requirements for consumers and professionals, faced with the explosion of accessible digital information and the proliferation of access means (PC, TV, handheld devices).

The consortium is composed of French and German public and private research organizations. The Quaero consortium is coordinated by Thomson. Other large industrial organizations participating are France Telecom, Jouve and Exalead. Dedicated technology suppliers Bertin, LTU, Synapse and Vecsys will contribute and further develop top notch technologies in their respective business domains. French and German public research institutes, coordinated by CNRS are CNRS (INIST, LIMSI, IMMI), INRIA, Institut Telecom, IRCAM, IRIT, LIPN, MIG-INRA, Joseph Fourier University, University of Karlsruhe and RWTH Aachen University. Finally the participation of public institutions BnF, DGA, Ina and LNE demonstrates the strong support of the public sector to the success of the program.

THESEUS [62] is part of the "Information Society Germany 2010 (id2010)" program of the federal government. The five year project has a budget similar to that of Quaero of about 180 million euro. THESEUS aims at the next generation of systems for intelligent search, management, automatic processing and representation of multimedia and multilingual content. In cooperation with leading partners from industry, the media and the research and scientific community, THESEUS is focussed on implementing innovative technologies (search technologies, ontologies, pattern recognition, meta data, translation) in concrete applications with the ultimate objective of developing solutions which are competitive in international markets. At the same time, THESEUS is to contribute toward preserving cultural heritage and maintaining cultural diversity. With the help of THESEUS, cultural institutions in Germany and Europe will be able to prepare their cultural goods and artworks in an innovative and structured way, so as to make them electronically accessible to a wide audience.

The focus of the research program is on semantic technologies, which determine contents (words, images, and sounds) not through conventional methods (e.g., combinations of letters) but which are able to recognize and place the meaning of a content in its proper context. Using these technologies, computer programs can intelligently comprehend the context in which data were stored. In addition, by applying rules and order principles, computers can draw logical inferences from the contents and autonomously recognize and produce connections between various pieces of information from different sources.

THESEUS will be implemented in two phases: In Phase 1, the first solutions and demonstrators are to be developed by 2008. Subsequently, these are to be further evolved through additional partners, which will be largely gained from medium-sized businesses by way of an invitation to bid (Phase 2).

At the current time, 30 research institutions, universities, and companies have joined the THESEUS program with planned projects. The industrial and public research partners are cooperating closely. They are coordinated by empolis GmbH. Also involved are internationally recognized experts of the Fraunhofer Society, the German Research Center for Artificial Intelligence (DFKI), the Research Center for Computer

[60] http://europa.eu/rapid/pressReleasesAction.do?reference=IP/08/418
[61] http://www.quaero.org/
[62] http://theseus-programm.de/

Science (FZI), the Ludwig Maximilian University (LMU) and Technical University (TU) in Munich, the TU Darmstadt, the University of Karlsruhe, the TU Dresden, and the University of Erlangen. The application scenarios are developed from the immediate research results and utilization interests of the leading partners German National Library, empolis, Lycos Europe, SAP, Siemens, as well as the following additional partners involved: Deutsche Thomson oHG, Festo, Intelligent Views, m2any, Moresophy, Ontoprise, Verband Deutscher Maschinen- und Anlagenbau e.V. (VDMA), and the Institute of Radio Technology.

It is clear that the funding and scope of both Quaero and THESEUS are far beyond that of MultiMatch. We will be extremely interested to see their first results.

Europeana[63] – the European digital library, museum and archive – is a 2-year project that began in July 2007. It will produce a prototype website giving users direct access to some 2 million digital objects, including film material, photos, paintings, sounds, maps, manuscripts, books, newspapers and archival papers. The prototype will be launched in November 2008 by Viviane Reding, European Commissioner for Information Society and Media. The digital content will be selected from that which is already digitised and available in Europe's museums, libraries, archives and audio-visual collections**.** The prototype aims to have representative content from all four of these cultural heritage domains, and also to have a broad range of content from across Europe. The interface will be multilingual. Initially, this may mean that it is available in French, English and German, but the intention is to develop the number of languages available following the launch. The project is run by a core team based in the national library of the Netherlands, the Koninklijke Bibliotheek. It builds on the project management and technical expertise developed by The European Library, which is a service of the Conference of European National Librarians. Overseeing the project is the EDL Foundation, which includes key European cultural heritage associations from the four domains.

3.7 Conclusions

Traditional search engines such as Google and Yahoo! are facing greater challenges as the World Wide Web grows faster than their indexing technology can keep up and popularity of the more focussed search engines is rising. Additionally, these search engines are beginning to lose the arms race against spam and fake content.

The publicly available multimedia search engines, which are of particular relevance to the MultiMatch project, currently offer varied levels of results. Podzinger's audio transcriptions appear to be of very high quality. Blinx's appear less so, and it's difficult to imagine using Retrievr in any practical way.

Most of the other Multimedia search engines still rely only on meta-data. IBM's Marvel intelligently generates its own meta-data after analyzing the media, but other search engines rely on a manual tagging process.

There are very few true multilingual search engines to compare. Fotolia.com appears to be one of the few, but its results appear inaccurate. Using the same search terms in different languages should produce the same results, but searching for "cat" in different languages produces completely different result sets with little in common with each other.

MultiMatch's achievements relate well to the current state of the art. Rival multimedia search engines such as THESEUS and the high-profile Quaero are still a long way from completion. Most multimedia searches rely on manually generated meta-data, and those which don't have demonstrated a level of ineffectiveness.

The current state of both multimedia and multilingual search still seems immature. The very few multilingual services available are limited in effectiveness and not particularly user friendly.

Additionally, MultiMatch has introduced new features such as intelligent key-frame generation, and transcript searches that take the user to the appropriate place in the media file. These features are still far from common-place within other search engines.

[63] http://www.europeana.eu/

4. Classification and Information Extraction

by Neil Ireson

Classification (also known as Categorisation) and Information Extraction are part of the Knowledge Discovery (KD) process, which attempts to find "interesting" patterns in data, i.e. those which reveal some underlying meaning (semantics). The KD process incorporates a number of other sub-processes including: Information Retrieval, Topic-tracking, Summarisation and Visualisation. KD was initially the focus of Data Mining research, where the data referred to that found in databases or spreadsheets, more recently, with the increase in computational resources and the availability of a mass of electronic media, the KD process encompasses a wider array of less structured media types, such as text, images, audio and video.

The Classification process allocates an object to one or more categories (or classes). Generally an object is viewed as a member of the category to which it is allocated, however in "fuzzy" or "rough" classification systems an object can also be a partial member of a category. Categories are generally used to contain objects which share a set of properties or attributes. Thus the classification process can be used to filter objects so that when a given category is selected, only objects with the desired properties are viewed or received.

The classification of media objects, such as text, images and videos, is the concern of library classification systems which organise the objects according to some predefined subject structure. For example, the most widely used library classification (taxonomic) systems, at least in the English speaking world, are the Library of Congress Classification and Dewey Decimal Classification systems. However the process of assigning (indexing) an object to a given category (or categories) in the classification is a laborious process involving careful consideration of the object's content. In addition such general classification schemes may not suit the requirements of the individual who wishes to identify and retrieve the classified objects. For specific domains or users alternative classification schemes may better suit their requirements and there may not be a ready mapping between the general and specific classification. Therefore research has focused on automatic approaches to facilitate the process of classification of objects according to their content.

Information extraction (IE) can be defined as the identification of specific instances of semantic elements (entities, events, relationships and their properties) within a given data object (i.e. a text or image). Thus IE can be viewed as the creation of an explicit structured representation (or metadata) from the information implicit in unstructured data. The IE task contrasts with the Information Retrieval (IR) as the result of IR is a sub-collection of objects, which are relevant to a given query; whilst the result of IE is a collection of facts extracted from the objects.

4.1 Pattern Recognition

Although there is a distinction between Classification and IE, IE can be considered as a classification process, the difference being that Classification is used to refer to the categorisation or labelling of an object as a whole, whilst IE refers to the categorisation, labelling or annotation of part of the object. In more general terms both Classification and IE can be considered as pattern recognition tasks; where a pattern is formed from features derived from an object. The recognition task maps (or classifies) a set of features onto a category, thus a media object (text, image or video), or part of that object, which exhibits a given pattern of features can be allocated to a semantic category, label or annotation. Categorisation, labelling and annotation can be considered to be synonymous processes, although annotation is generally seen as providing a more informative description than a simple label or category. Much research is devoted to the construction of automatic semantic annotation systems, due to the fact that manual annotation is a laborious task. This annotation task can be divided into three processes:

1. The processing of the media object to extract low-level feature descriptions.
2. Mapping between the low-level features and high-level of semantic concepts: the difference between these two descriptions of an object is referred to as the "Semantic Gap".
3. Understanding: moving from the annotation of a media object with a set of semantic concepts to a comprehension of the object as a whole (e.g. the narrative of the text or video, or the scene depicted by an image). Such the semantic interpretation may well depend upon the existence of (background) knowledge not contained within the media object.

The first process, feature extraction, is obviously dependant on the media type and will be discussed, below, in relation to each of the media types of interest to the MultiMatch Project (text, image and video. Pattern recognition is concerned with the second process, i.e. closing the Semantic Gap; the general (Machine Learning) approaches applied to the pattern recognition task will be discussed in the next section, with the specific applications in each of the sections on the media types. The third process, understanding, is beyond the scope of this document and the MultiMatch project.

4.2 Machine Learning

Most research into Classification and IE is concerned with the application of Machine Learning (ML) algorithms to the process of detecting classification patterns. The algorithms can be divided into three types, supervised, unsupervised and semi-supervised classification algorithms.

4.2.1 Supervised Classification

Supervised classification is based on the learning of a sequence of input/output pairs. It aims at producing the right result when it is given a new input. Supervised classification is achieved through the labelling of the data by a supervisor. When a new sample has to be added, it is labelled according to the already labelled data. The classification is based either on discrimination or on characterization. Discrimination consists in defining the frontiers between the already labelled data. New samples may then be added to the class they belong to by evaluating their position relatively to these frontiers. Characterization follows a different approach and intends to associate a set of invariants to each class. A new sample will belong to the class having the most similar properties. The following sections give a general introduction to the most widely used ML methods which have been employed in various Classification and IE tasks discussed below.

Decision Tree
The induction of decision trees was one on the original ML techniques developed and has been widely adopted due to its relatively simple implementation and transparency of the classification model. Most of the implementations are based around Quinlan's ID3 and C4.5 [Quinlan, 1993]. The algorithm iteratively partitions the example set according to the values of the most discriminative feature, i.e. the feature which provides the highest information gain.

Rule-based Models
Rule induction methods, unlike the global top-down approach of decision trees, develop a number of "if-then" type classifiers to cover the problem domain (represented by the training examples). These rules are not necessarily exhaustive (i.e. cover all the domain space) nor are they necessarily mutually exclusive (i.e. more than one rule can cover the same space). The "if" section of the rule determines the feature pattern, which constrains the rule coverage in the feature space; the "then" section determines the category to be associated with that part of the feature space. Rule induction algorithms attempt to create rules which are "consistent", i.e., do not cover any negative example and "complete", i.e. covers all positive examples. In practice the consistency and completeness constraints are relaxed to cope with uncertainty, imprecision and noise, in the problem domain and training examples. Rules are thus evaluated according to some measure based on their coverage and predictive accuracy, balancing the trade-off between generality (increased coverage) and accuracy (only covering positive examples).

To generate an individual rule most learner algorithms employ one of the following search strategies.

- Specialisation or top-down algorithms start from the most general rules and repeatedly specialise them imposing constraints in order to avoid covering negative examples.
- Generalisation or bottom-up algorithms start from the most specific rule that covers a given example; they then generalise the rule, relaxing its constraints to extend its coverage of without covering negative examples.

These learning strategies are attempting to generate rules which are either cases of Least General Generalisation or Most General Specialisation. There are other methods of rule induction such as using genetic algorithms [Holland, 1975] which cover the feature space then improve the rule set by combining "good" rules (using a crossover function) and performing local hill-climbing (using a mutation function).

One of the main attractions of rule-induction models is that (as with decision-trees) the model is human interpretable, i.e. that it is possible to determine the semantics behind the domain concept encapsulated by a rule.

Nearest-Neighbour

One of the simplest approaches to classification is to employ *nearest-neighbour classifiers*, also known as *memory-based* learning. The basic concept is to determine the distance between examples, thus an example with an unknown category can be assigned the category of its nearest neighbour, or more usually the most likely category given its K nearest neighbours. Obviously the complexity in the method is in determining distance function. The most straight-forward implementation use a standard Euclidian distance metric, however this assumes a very uniform problem space. More domain specific ML approaches can be applied to learning the appropriate feature weights or combinations. One of the principle difficulties with the application of nearest-neighbour learning is the prohibitive computational complexity when dealing with high dimensional feature spaces and large data sets. The Tilburg Memory-Based Learner (TiMBL), at http://ilk.uvt.nl/timbl/, provides the most widely used implementation of the approach.

Artificial Neural Networks (ANN)

ANN are based on an analogy to their biological counterpart, in the sense that they have simple processing nodes with a high degree of interconnection, processing involves the passing of simple scalar messages, learning occurs via the altering of weights which determine the interaction between nodes. The functioning of an ANN is determined by the topology of the network and learning algorithm applied to the adaptation of weights at the nodes. The most generally used topologies involve an input and output layer of nodes and one or more (hidden) layers. The topology of an ANN (the number of layers and nodes) determines its capacity, i.e. its ability to model a domain; however for complex domains the required capacity can cause difficulty in convergence of node weights.

Support Vector Machines (SVM)

SVM separate the problem domain space using hyperplanes; however one of the most appealing features of this approach is that as well as minimising the empirical error when dividing the example classes, the algorithm also positions the hyperplane such that it maximises the geometric margin between the proximate examples along the hyperplane. These examples are the support vectors; thus SVM are also known as maximum margin classifiers. An important development in SVM was to cope with non-linearity by employing a "kernel trick" [Boser, et al., 1992] which is used to transform the original feature space into a higher-dimensional space using a kernel function. Thus the hyperplane separation in the transformed space can represent a non-linear separation in the original space. The difficulty in implementation then becomes determining the appropriate kernel function for a given domain.

Hidden Markov Model (HMM)

A Markov Process is one in which a system stochastically changes from one state to another, in discrete steps. The change (transition) from the current state into the next state is dependent solely on the current state and not on any previous states. In a regular Markov model, the state is directly observable, and therefore the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly observable, but variables influenced by the state are visible, thus the challenge is to determine the hidden parameters from the observable parameters. There are 3 canonical problems associated with HMMs:

1. To determine the probability of a particular state given the parameters of the model; solved by the forward-backward procedure.
2. To find the most likely sequence of hidden states that could have generated a given state given the parameters of the model; solved by the Viterbi algorithm [Viterbi, 1967]
3. To determine the parameters of the model (state transition probabilities), given a set of observed state sequences; solved by the Baum-Welch algorithm (a special case of the Expectation-Maximisation (EM) algorithm [Dempster et al., 1977]

One of the criticisms levelled at HMM is that in order to make the computations tractable an assumption of conditional independence between each discrete state has to be made (i.e. each state is independent of his ancestors and each observation depends only on current state). This may prove too restrictive for certain problem domains.

Maximum Entropy Model (MaxEnt)
Claude Shannon [Shannon, 1948] introduced the fundamental concept of entropy in information theory to measure the amount of uncertainty (or randomness) there is in a signal or event. MaxEnt modelling is used to determine the probability distribution which maximises the entropy given the known information (i.e. training examples). Applying MaxEnt involves constructing a stochastic model that accurately represents the behaviour of the "random" process by estimating the conditional probability that, given a context (set of features), the process will output a given result (category). The process involved in calculating the model is described in a number of relatively easily digestible tutorials [Berger, 1996; Ratnaparkhi, 1997].

The attractive feature of the MaxEnt model is that, given incomplete information is available (as is the case with IE tasks) inferences, derived from the probability distribution, are made solely on the available information.

Conditional Random Fields (CRF)
CRF can be viewed as a generalisation of the HMM and MaxEnt Model that aims to overcome the independence assumption drawbacks of HMM and the "Label Bias" problem exhibited by other Maximum Entropy Markov-based models. The Label Bias problem can be attributed to the local conditional modelling of each state, as states whose following-state distributions have low entropy will be preferred; despite these previous states possibly having no relation to the observations.

CRF is an undirected probabilistic graphical model where a node represents a discrete random variable, whose distribution is to be inferred, and an edge represents a dependency between the associated random variables. The distribution of each discrete random variable in the graph is conditioned on an input sequence provided by the feature space. A good introduction to CRF is provided by Hanna Wallach [2004].

Boosting
Boosting is a meta-learning approach for improving the accuracy of any given learning algorithm. The Boosting algorithm, which can be seen as a form of Probably Approximately Correct (PAC) learning [Schapire, 1990], iteratively combines (usually by using some majority voting method) weak classifiers (i.e. ones which are at least better than random) into a single accurate classifier. At each iteration the examples are weighted so that those incorrectly classified are "boosted" so that the new weak classifiers focus on resolving the classification error. The most common boosting algorithm is AdaBoost [Freund and Schapire, 1999]

4.2.2 Unsupervised Classification (Clustering)

Perhaps the most problematic practical issue with using supervised classification systems is the need for a set of training examples; unsupervised classification systems remove the need for such a priori labelling of examples. An unsupervised classification process is only given a set of examples; these are then grouped (or clustered) according to the similarity and/or dissimilarity of their features. This process is also known as clustering and can be viewed as attempting to uncover the latent structure within a domain.

The principle issue in clustering is determining the appropriate distance metric to calculate the degree of similarity between two points in the feature space. Using the derived distance metric clustering generally involves minimising distances between examples within a cluster (intra-cluster variance) and maximising distance between examples in different clusters (inter-cluster variance). Clustering either exclusively

allocates examples between clusters or examples can be partially or wholly members of one or more clusters, this is known as fuzzy clustering [Dunn, 1973]. One limitation of the most commonly used clustering algorithms is that either the number of clusters to be provided a priori, such as in the k-means algorithm [MacQueen, 1967] or the size of clusters has to be provided, as with QT (Quality Threshold) Clustering [Heyer et al, 1999].

One approach to removing the need for such a priori information is to use clustering techniques which place the clusters within a hierarchical structure. Hierarchical clustering can be either:

- Top-down beginning with a single cluster and splitting it to maximise some inter-cluster distance, and then continue splitting the clusters until there is one cluster per example.
- Bottom-up being with one cluster per example and combine the most similar cluster, and then continue to combine the most similar clusters until all examples are contained within a single cluster.

However such an approach is computationally expensive, especially when there is a large number of examples, n, to cluster as the complexity is in the order of $O(n^2)$.

A further issue with unsupervised learning is that although it does not require initial user input to create the classification; the output tends to require post classification operations in order to make the results meaningful, such as the allocation of labels or summaries, to the cluster, which is representative of their content.

4.2.3 Semi-supervised classification

Semi-supervised learning is a type of ML technique which makes use of a (typically small amount) of labelled data with a (typically large amount) of unlabelled data for training. Such methods use unlabeled data to either modify or give more weight to hypotheses deduced from the set of labelled data. Zhu [2005] provides a good review of the various approaches to semi-supervised learning.

Expectation-Maximisation (EM)
The goal of EM is to maximize the posterior probability of the model parameters (probability distribution means, standard deviations, and weights) given the data, in the presence of missing data, by applying the following process:

1. Initially estimate model parameters, generally based on some prior (domain) knowledge
2. a) Expectation (E) step: compute an expectation of the likelihood by including the missing (or latent) class variable as if it were observed.

   b) Maximisation (M) step: compute the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E step.
3. Iterate step 2 by using the parameters calculated in the M step to initialise the E step and continue the process until a convergence threshold is satisfied.

The main concern when applying EM is avoiding convergence to local maxima. If the model convergences a local maximum, which is far from the global maximum, the use of unlabeled data is likely to have an adverse impact on learning. One proposed solution to alleviate this possibility is the selection of initial estimates using an active learning approach [Nigam, 2001].

Co-Training
In co-training [Blum and Mitchell, 1998], two classifiers are trained using disjoint features spaces. The features are divided into two class-conditionally independent sets, and a classifier is trained on the available labelled data, using each of the feature sets. Then those unlabelled examples for which one classifier is most confident in its prediction are labelled and added to the training set of the other classifier. The process is continued until some threshold level of accuracy on the training data is reached.

Expansion
Expansion is bootstrapping technique (i.e. one in which a process activates another process which serves the same purpose) which is related to query expansion from Information Retrieval, where terms are added to a query in an attempt to improve precision and recall. The process is initialised with a small set of labelled examples; from these, similar examples are found in the unlabelled data by expanding (relaxing) the feature values of the labelled examples. The similar examples are then assigned labels related to the associated labelled examples; these labels can be weighted according to the degree of similarity. The newly labelled

data is then added to the training set and the process is repeated; with limits imposed on expansion to prevent making spurious inferences on examples too distant from the original labelled examples.

Active Learning

An active learning approach involves selecting the most appropriate sample of unlabelled data to label. The selection of the example can involve the use of a classifier to predict the labels on the unlabelled data to select the examples for which the classifier is most uncertain. Alternatively clustering techniques can be employed to select the most diverse set of examples. Unlike the other semi-supervised methods active learning then relies upon human intervention to label data, however the principle is to minimise the amount of data which needs labelling whilst maximising the quality of that data in term of building the classification model.

4.3 Text

4.3.1 Textual Data

Most of the research into Test Mining has come from the Natural Language Processing (NLP) domain which, for obvious reasons, has focused its attention on written text and transcribed speech. This is known as free or unstructured text, although there is, to a greater or lesser extent, a grammatical structure which can be exploited. Michelson and Knoblock [2005] have reported on some interesting work examining IE of unstructured and ungrammatical text.

Recently there has been more interest in the "mining" of semi-structured texts. In such texts the meaning is partially provided by the structure of the document in which the text appears. The documents may have titles, keywords or summaries and be divided into, possibly titled, sections. There might be internal or external (hypertext) references or text can be contained within tables. This type of document is exemplified by the HTML pages found on the internet, and the interest in being able to extract the information from the text on these pages is driven by the desire to exploit the potential of the billions of pages on the WWW.

Text Classification and IE systems generally presume that the input documents contain text from the domain of interest. However as well as the text providing a potential source of information to answer a given query, it may also contain noise the removal of which would improve the performance of the overall systems. This is often prevalent in web pages which may contain; adverts, menus, site-specific text and links, etc. which do not (directly) relate to the main content of the page. Being able to cleanly extract the relevant text has been highlighted as one of the key challenges for Web content mining [Liu and Chen-Chuan-Chang, 2004].

There are many factors which affect the interpretation of a piece of text, some of these are explicit and obvious such as its language (English, Russian, Japanese, etc.) or source (newspaper, journal article, audio transcript, web page, etc.). The text is also affected by the domain (art, sport, science, politics, etc.) to which it relates. The meaning will also be affected by the intention of the author; this may be to inform (news articles, user manuals, etc.), entertain (literature) or convince (argument, propaganda, marketing, etc.).

4.3.2 Text Analysis and Feature Extraction

The pre-processing of text to extract the relevant features is a necessary phase in all text mining techniques, to transform the text into a representation suitable for processing. Indeed there is often such a dependence on the application of specific pre-processing techniques that the distinction between the pre-processing and text mining technique is arbitrary.

Text Segmentation

The generic term "text segmentation" has analogies in analysis of other media type (i.e. images, video) in that it is a process which attempts to partition the data into coherent regions. For textual data, segmentation is used to refer to a number of different processes, the most basic being tokenisation where a text is partitioned into its atomic units; generally taken to be the word, term or token, although for certain applications (such as language or author identification) the text may be broken down to the character level. It is worth noting that although the process of tokenisation is considered to be a trivial task in Indo-European languages, the process is considerable more complex for Asian languages, such as Chinese, Japanese, Korean, Thai, Vietnamese, Mongolian, and Tibetan, where words cannot be fully identified by typographic features (e.g. spaces).

Similarly the text segmentation process of Sentence Boundary Detection is viewed as a trivial task in Indo-European languages; as boundaries are generally delineated using given characters, such as a full-stop or multiple newlines. The tokens and sentences derived from segmentation are used as input for further lexical and syntactic analysis (see below).

Another process in text segmentation relates to topic detection and tracking (TDT), this can be broadly divided into two forms; the detection of change-of-topic boundaries in a stream of text (such as speech transcripts or newswire feeds) and the partitioning of text into subtopics. Text classification, IE and indeed most other NLP techniques inherently rely on a notion of text documents, therefore the partitioning of a text stream into topic "documents" is a necessary precursor to the application of such techniques. Also the partitioning of long or complex documents into "sub-documents", each containing a coherent subtopic, can be of benefit to NLP techniques as it provides focused input and avoids information overload.

Research into TDT techniques can be divided into the generic machine learning areas of supervised and unsupervised learning. The performance of supervised learning techniques, as is generally the case with such approaches, is reliant on the amount and quality of learning material available, and tend to produce solutions which are not readily portable to other domains. Unsupervised techniques are more domain-independent, mainly relying on the concept of lexical coherence, i.e. topics can be differentiated by their distinct use of vocabulary. In addition to lexical coherence TDT techniques can also determine "cues" which mark the likely transition between topics.

Most of the work in this area has been based around the series of evaluation studies performed as part of the DARPA Translingual Information Detection, Extraction, and Summarization (TIDES) program annually from 1998 to 2004 (see http://www.nist.gov/speech/tests/tdt/index.htm).

- *Semi-structured Documents*

The increasing use of the Internet as a means of communication has provided a large amount of machine readable XML/HTML documents which, as well as containing the text to communicate, contains structural information for the presentation of the text. This structural information can be used to segment the text into meaningful sub-sections [Luo et al., 2004]. This can be seen as an extension to the normal text segmentation process but with the use of HTML tags as "cues" for segment boundaries.

HTML documents, as well as providing additional information for segmentation, add a complexity over free text documents in that when the HTML is rendered the locality of text in the source HTML can be altered. As segmentation relies, to an extent, on the proximity of text to determine cohesion, the final presentation of the HTML must be considered. Thus can be done either by directly analysing the HTML code to extract its structure [Mukherjee et al., 2003], or by utilising the actual visual structure of the rendered HTML page [Kan, 2001; Yang, 2001; Gu et al., 2002].

Lexical Analysis

Lexical analysis provides an interpretation of the meaning behind individual words.

- *Part-Of-Speech (POS) Tagging*

POS tagging is the process of assigning grammatical classes to words in a sentence. The principal difficulty arises because some words can have multiple POS assignments depending upon their contextual use. Its importance stems from the fact that knowing the POS can be useful in subsequent text processing tasks; such as word-sense disambiguation and parsing.

- *Stemming and Lemmatization*

Both stemming and lemmatization attempt to find the base form of a given word (known as the "lexeme" for the word). Lemmatization is a more in-depth process which involves knowing the POS and may also require knowledge of the grammar. Stemming in contrast operates on a single word without knowledge of its context, and therefore cannot discriminate between words which have different meanings depending on POS. Therefore stemmers are less accurate than lemmatizers, they are however, easier to implement and faster. In most applications it is assumed that the use of a stemmer provides sufficient accuracy, however this may be more due to the fact that stemmers are available for a wide range of languages (see Snowball stemmer collection at http://www.snowball.tartarus.org/) and the difficulty in implementing a lemmatizer, rather than any strict empirical assessment of the cost/benefit of stemmers versus lemmatizers.

- *Word-Sense Disambiguation (WSD)*

WSD relates to the problem of "polysemy" where a word can have multiple meanings. For example, given the sentences, "the bank was breached by the water" and "she deposited her money in the bank", WSD determines whether "bank" refers to a river or financial bank. There are two main approaches to WSD; deep approaches and shallow approaches.

Deep approaches presume access to a comprehensive body of world knowledge. However these approaches are not very successful in practice, because of the difficulty in acquiring such knowledge in a computer-readable form (such as the Cyc project [Lenat, 1995], which is now OpenSource). Also there are many oddities introduced by the use of language, such as analogies and idioms, which may deliberately contradict the "proper" use.

Most WSD research focuses on shallow approaches which just consider a words context as defined by its surrounding words, i.e. river bank relates to water, fish, boats, etc. and financial bank relates to money, credit, manager, etc. These approaches define a window of N content words around each word to be disambiguated in the corpus, and statistically analysing those N surrounding words. Two shallow approaches used to train and then disambiguate are Naïve Bayes classifiers and decision lists. In recent research, kernel based methods such as support vector machines have shown superior performance in supervised learning. But over the last few years, there hasn't been any major improvement in performance of any of these methods.

It is instructive to compare the word sense disambiguation problem with the problem of part-of-speech tagging. Both involve disambiguating or tagging with words, be it with senses or parts of speech. However, algorithms used for one do not tend to work well for the other, mainly because the part of speech of a word is primarily determined by the immediately adjacent one to three words, whereas the sense of a word may be determined by words further away. The success rate for part-of-speech tagging algorithms is at present much higher than that for WSD, state-of-the art being around 95% accuracy or better, as compared to less than 75% accuracy in word sense disambiguation with supervised learning. These figures are typical for English, and may be very different from those for other languages.

- *Latent Semantic Indexing (LSI)*

The underlying idea behind LSI is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other [Landauer, 1998]. Thus LSI represents the meaning of a word as a kind of average of the meaning of all the passages in which it appears, and the meaning of a passage as a kind of average of the meaning of all the words it contains.

Syntactic Analysis

Syntactic Analysis is the study of the rules that govern how different words (categorised by their POS; nouns, adjectives, verbs, etc.) are combined into clauses, which, in turn, are combined into sentences. A sentence parsed in order to determine its grammatical structure with respect to a given formal grammar; this transforms input text into a data structure, usually a tree, which is suitable for further processing. Shallow parsing (or "chunking") is an analysis of a sentence which identifies the clauses (noun groups, verbs ...), but does not specify their internal structure, or their role in the main sentence. A frequent use of parsing in IE is to use the parse tree to extract the Subject-Verb-Object pattern from a sentence.

Use of Ontologies

The research on combining ontologies and IE involves both ontology building (generation and population) as an application of IE, and using ontologies to aid in the process of extracting information. In terms of aiding the IE process, given that a concept is present in a text, either because it has been annotated by a user or extracted by an IE system, ontologies can be used to provide "clues" to the other information which is likely to be in the text. Ontologies can also be used to disambiguate, as was mentioned above in WSD, for example given the text contains the word Paris, it is most likely to be a reference to the capital of France, unless the text also contains the geographical place name Texas in which case the ontological can be used to provide the information that Paris is a place in Texas, or if the page contains a Person who is a known celebrity then Paris is more likely to refer to "Paris Hilton", another celebrity. Of course the use of an ontology requires that an suitable and well-formed ontology exists and as was stated above, developing an ontology of reasonable size is an expensive task. However where such ontologies exist, such as in the biological domain, they have been found to be useful in providing information to text processing tasks [Honavar, 2001].

4.3.3 Text Classification (TC)

Text classification, that is the assignment of text documents to one or more categories based on their content, is an important component in many text analysis tasks such as; email "spam" filtering [Drucker et al., 1999], authorship attribution [Diederich et al., 2003], topic identification [Allan, et al., 1998] and (of specific interest to MultiMatch) Web page classification [Dumais and Chen, 2000]. However, much of the initial research into the use of ML for TC has been in the filtering of news stories, primarily because this was the first domain to provide a sizeable "Gold-Standard" corpus for training and evaluation of text classification systems [Lewis, 1997 and Lewis et al., 2004].

The automatic TC process involves: extracting the features from the text, selecting the most discriminating textual features (in its simplest form a set of keywords), allocating a weight to indicate the relative "importance" of the selected features in determining the semantics of the document (for supervised learning this is a measure of the degree to which a feature is indicative of a category) and define a similarity metric to determine the degree to which an object is assigned to a category (based on the combine the feature weights of an object). A good review of the ML approaches used for TC is provided by Sebastiani [Sebastiani, 1999 and 2002].

The feature extraction methods applied in TC tends to be relatively simplistic, in terms of applying the text analysis techniques described above. Textual documents are represented as a vector of terms (words) which are generally reduced to their lexeme (using stemming), and uninformative terms are removed using stop-word lists derived from large corpora (such as the Google stop-word list). However even such simple approaches are language specific. Attempts at applying state-of-the-art text analysis techniques (including parsing [Moschitti and Basili, 2004] and WSD [Kehagias et al., 2003]) have not shown substantial improvement in classification performance over the use of simpler representations.

Given a reasonably sized corpus the number of terms present in the vector representation of the text can be large (i.e. thousands of unique terms). For the application of ML techniques this can be problematic, thus dimensionality reduction (feature selection) methods are employed. The most commonly used approach for supervised learning systems is to select terms which are most indicative of a category; using measures such as Chi-square and Information Gain. An alternative is the use of Latent Semantic Indexing (LSI) to transform the original vector into a space with fewer dimensions [Liu, et al., 2004].

The weighting of the selected features (words or terms) to indicate their importance intuitively should be higher for those features that appear more often but are found in fewer documents. Thus the classic measure is given by the Term Frequency (TF) multiplied by the Inverse Document Frequency (IDF). The calculation of similarity between one document and another, or a document and a given category is determined by the co-occurrence of terms between the documents/category and the weight of those terms Salton and Buckley [1988] examine various approaches to term-weighting.

If sufficient training material is available for a given application domain then supervised ML techniques can be applied to feature selection and/or weighting, resulting in performance improvements over the use of the generic techniques described above. In TC a wide range of ML approaches have been applied including; nearest neighbour classifiers [Masand, et al., 1992], decision trees [Lam and Ho, 1998], Bayesian classifiers [McCallum and Nigam, 1998], Support Vector Machines [Joachims, 1998], rule learning algorithms [Cohen and Singer, 1996], neural networks [Li and Jain, 1998] and boosting [Schapire and Singer, 2000].

As has been stated for many applications a reasonable set of training data is too expensive to create so in order to overcome this document labelling bottleneck, semi-supervised methods have been applied [Nigam and Ghani, 2000; Nigam et al., 2000], however learning text classifiers from unlabelled data is still very much an active area of research.

The application of Text Clustering has tended to use the same basic techniques as text classification for feature extraction, selection, weighting and comparison. Although rather than measures being relative to the given categories, in clustering the measures relate to the categories constructed by the bottom-up or top-down clustering process. It is interesting to note that there has been some work which has shown that the addition of semantic information can aid the clustering process [Hotho, et al. 2003].

4.3.4 Information Extraction

Information Extraction from text, as a research field, has developed out of the more general field of Artificial Intelligence (AI) and more specifically from the area of knowledge representation. The mapping of natural

language texts into more formal conceptual models originated with Roger Schank [1975] and Marvin Minsky's [1975] work in the 1970's. Schank's work formalised texts in terms of "scripts", where concepts within the text are interconnected by dependencies defined by a set of syntactic and semantic rules. Minksy developed a "frame" based representation where, each concept (entities, actions, events) is represented in a frame; the properties of the concept being represented as slots in the frame. The principal difference between to two forms of representation is that events and actions in scripts are ordered; i.e. represent procedural knowledge, whilst frames are linked into a tree or network structure, where a frame can be the value associated to another frame's slot. Such issues of knowledge representation are still important, and the goals of this original work are still fundamental to current research (i.e. the relationship between MUC "templates" and Minksy's frames). However recent IE research has become principally more concerned with the pragmatic process of acquisition rather than the representation of knowledge.

The overall IE process can be divided into a number of sub-tasks; named-entity recognition, coreference resolution, entity relation recognition and event recognition. The primary focus of research in IE has been the utilisation of Machine Learning (ML) techniques to aid in these tasks. The following sections will outline the processes involved in each of the IE sub-tasks and discuss the key techniques applied to them.

IE Subtasks

- *Entity Extraction / Named Entity Recognition*

Entity Extraction or Named Entity Recognition (NER) is the identification of a term or phrase which refers to a specific entity. For example; a person or organization, place name, temporal expression, or certain types of numerical expression. Most of the research into IE has focused on the area of NER as it is the foundation of the other IE tasks; relation and event extraction.

The techniques employed in NER, to an extent, depend upon the entity to be extracted. Some entities, such as temporal expressions, have a relatively common representation and usage across domains. However other entities require more domain specific approaches, this is particularly true of Terminology Extraction, e.g. the extraction of protein or chemical names, which is an important sub-problem in NER. It is worth noting that the extraction of time expressions (TIMEX) is a significant area of NER research as the recognition of TIMEX is necessary for determining the temporal ordering, which is a fundamental task in event recognition. The work in this area has been stimulated by the availability of the 2004 ACE Temporal Expression Recognition and Normalization (TERN) corpus.

There are three basic approaches to identifying entities:

1. Gazetteers or Name lists

A look-up table which matches character strings with entities. Gazetteers work well for stable lists of names (such as days of the week, chemical elements, etc.) but are less useful where the list of names is constantly growing or changing. Even when the names are stable there is the problem of resolving ambiguous usage, for example Rose can be a flower, place name, persons name, colour, etc. There are however a growing number of useful resources being developed such as Getty Thesaurus of Geographic Names (TGN), which contains around 1.3 place names, and Union List of Artist Names (ULAN), which contains around 250000 artist names.

2. Orthography

Orthography NER considers the "internal" character pattern of an entity's lexical representation. This works well for things like dates, phone numbers or postcodes which are readily recognized by their internal format (e.g., DD/MM/YY or chemical formulas). It is however not a technique generally applicable to the extraction of many entity types and thus is used in conjunction with the contextual pattern.

3. Contextual Patterns

Most of the work on NER has focused on the use of contextual patterns, where an entity is identified in the context of the surrounding terms. In the original MUC evaluations some of the best performing systems used hand-coded pattern rules using specific grammars (such as JAPE [Cunningham et al., 2002] which provide syntax for the creation of NER pattern rules. However the creation of such hand-coded rules requires a considerable amount of effort and, as with gazetteers, the performance of rules

tends to be brittle when applied to domains with dynamical changing entity names or name usages. Therefore the majority of work has focused on alleviating the problems of determining contextual patterns for entity identification with the use of machine learning.

- *Coreference Recognition*

Coreference recognition finds multiple references to the same object within in a text. The coreferent objects can be expressed by; the same text, or in a modified version (i.e. James, Jamie, Dr J. Smith, etc.) or as pronouns and designators ("he treated the patient", "The doctor called"). The references can occur both earlier (anaphoric references) or later (cataphoric references) in the text.

- *Entity Relation Extraction*

Relation Extraction identifies the occurrence, and type of relation between two entities, e.g. a person "is_located_at" a city, or gene "codes_for" a protein.

- *Event Extraction*

Event recognition extracts a collection of entities and relations which describe a single event. At the MUC conferences this task was referred to as template filling, while "Event Detection and Recognition" is the term adopted in the ACE program. The simplest approach is to assume that a given segment (sentence, passage or document) of text refers to a single event and fill the templates by combining entities and relations within that segment; resolving any of the co-reference between entities.

Supervised learning methods

The technology currently dominating IE is the supervised learning techniques. The basic approach is to formulate the IE problem as a pattern classification task; training the classification model on a set of pre-labelled positive and negative examples. The positive examples are provided by the labelled (or annotated) entities in the text, the negative examples are provided by the rest of the text. The ML systems can either develop models to identify entire entity in a text or to separately identify the positions defining the start and end of the entity. The pattern used to classify the examples is formed from the lexical, syntactic or semantic features derived from the text using the preprocessing techniques described above. In the training phase examples are extracted from a text by considering a window of features around the entities. ML algorithms are then employed to determine the patterns surrounding an entity which can be used in its identification. These patterns can then be applied to an un-annotated text to determine the likely placement of an entity; if start and end positions are identified then a process of pairing is used to resolve conflicting annotations. There has been a wide range of machine learning algorithms applied to the IE task; in the following sections we will discuss the key approaches.

Despite its general adoption for other tasks, decision tree induction has not been widely used in IE [Sekine, 1998] and [Karkaletsis, 2000], being two of the few examples) as it is less applicable to tasks, such as IE, where features are likely to have non-linear interactions, which adversely effects "greedy" induction processes, and possess a large number of values, which causes problems in determining the discriminative effect of features and limits the transparency of the final tree. Similarly nearest neighbour techniques have not been widely adopted, although Ahn recently examined their use in Event Extraction [Ahn, 2006]; however the work emphasised the approach to the modulisation of the task rather than extraction performance.

A good survey of the initial approaches to the use of rule-based induction for IE is provided by Muslea [1999]. Since then the two main applications of rule-learner to IE have been the LP2 generalisation technique [Ciravegna, 2001] and the uses of Inductive Logic Programming (ILP) [Aitken, 2002]. Simple rules have also been used for the "weak learners" in a boosting approach [Freitag and Kushmerick, 2000].

HMMs have been used widely in text analysis problems due to text, as an ordered sequence of tokens (or textual features), being readily formed as a Markov model. In IE, HMM have been used for the general NER task [Bikel et. al, 1997], as well as specific domains; in particular the biomedical domain [Leek, 1997; Shen, et al. 2003; Bunescu and Mooney, 2004]. In addition other probabilistic techniques have been applied to IE tasks; Maximum Entropy Model (ME) have been used for both Entity Recognition [Chieu and Hwee, 2002; Borthwick, 1998] and Coreference resolution [Kehler, 1997], and Andrew McCallum has championed the use of Conditional Random Fields (CRF) for NER [McCallum and Li, 2003; Sutton et al., 2006] and also for

the extraction of information contained within web page tables [Pinto et al., 2003]. David Ahn has compared the use of CRF for TIMEX extraction [Ahn et al., 2005]; the work also applied MaxEnt to the normalisation of TIMEX statements.

A side from the attraction of using SVM due to their classification and generalisation capabilities, the use of kernel functions allows for a nature discrimination of graph representations as found in parse trees and structured (XML) documents. Therefore SVM have been used widely for the NER task [Isozaki and Kazawa, 2002; Finn and Kushmerick, 2004; Li et al., 2005; Iria, 2006], and specifically for TIMEX extraction [Hacioglu et al., 2005], as well as in coreference [Isozaki and Hirao, 2003] and relation extraction [Zalenko et al., 2003; Culotto and Sorensen, 2004].

Unsupervised/semi-supervised learning methods

Several approaches have applied clustering to IE where a word is characterised by its context and lexical features, for example NER [Lin and Pantel, 2000], relation extraction [Hasegawa, 2004], coreference resolution [Cardie and Wagstaff, 1999] noun phrase deal [Hasegawa et al., 2004] with Gooi and Allan [2004] extending the work to cross-document co-reference.

There are a number of approaches which have applied semi-supervised learning to the NER tasks. These employ bootstrapping techniques by initialising the algorithm with a set of optimised seed patterns which are used to extract a set of Named Entities, these are then marked-up in the unlabelled texts and new patterns are inferred and added to the set of initial patterns [Riloff and Jones, 1999; Collins and Singer, 1999; Etzioni et al., 2005; Nadeau et al., 2006]. Yangarber et al. [2000] use a similar approach, but perform the analysis at the pattern/document level to extract sentences rather than the Named-Entity/pattern level. A similar semi-supervised technique has also been used to extract relations [Brin, 1998].

Finn and Kushmerick [2004] compare a number of Active Learning approaches to IE, although the results are inconclusive a technique which selects documents most dissimilar to those in the labelled set and one which implements a co-train learning like approach improved over the baseline.

4.3.5 Evaluation

For an overview of the history and issues involved in evaluation of IE systems see Lavelli et al. [2004]. There have been a number of challenges which have provided both resources and incentive to stimulate research into Classification and IE.

Reuters: The initial Reuters corpus [Reuters-21578] was the main classification corpus for many years which was both positive in that it provided a means to compare techniques and negative in that it focussed research on a single domain. There is now a new corpus available (RCV1 [Reuters Corpus Volume 1]) which is much larger than the first.

Message Understanding Conference (MUC): was the main testing ground for IE approaches from its start in 1987 to its demise in 1998.

Automatic Content Extraction (ACE): (http://www.nist.gov/speech/tests/ace/) has replaced MUC and continues to organise various challenges for IE tasks.

Pascal Challenge: (http://tyne.shef.ac.uk/Pascal/) The Pascal Challenge on Evaluating Machine Learning for Information Extraction attempted to provide a level “playing-field” on which to assess relative approaches to ML for IE by providing a standard pre-processed corpus [Ireson et al., 2005].

4.3.6 Systems

There are many systems which provide varying degrees of text classification and IE functionality. The following list gives an indication of the most renowned systems which offer resources which are available for research purposes; there are also a number of commercial systems available (see Fan, et al. 2006 for an overview of these systems):

- Armadillo [Ciravegna et al., 2004]
- DIDEROT [Cowie et al., 1993]
- GATE [Cunningham et al., 2002]
- KIM [Popov et al, 2004]
- Know-It-All [Etzioni et al., 2004]

- LingPipe (http://www.alias-i.com/lingpipe/)
- Seeker/Semtag [Dill et al., 2003]
- Snowball and QXtract (http://snowball.cs.columbia.edu/)

4.4 Images

Image analysis is the quantitative or qualitative characterisation of two-dimensional (2D) or three-dimensional (3D) digital images to extract meaningful information. The characterisation of an image is based upon visual features which are extracted from that image, this can then be used to classify images with similar characteristics for applications such as content-based image retrieval (CBIR), which is also known as query by image content (QBIC). Applications may require the classification and retrieval of the entire image as a whole; however images may also be segmented into sub-regions which represent distinct objects within the image.

4.4.1 Feature Extraction

There are four main descriptors for the visual content of the image:

- Colour Features.
- Textural Features.
- Geometrical or Shape-based Features.
- Topological Features.

These features can either be global or local. Global image analysis considers the image as a whole, whilst local analysis first segments the image into several Regions Of Interest (ROI) then determines the properties and features of the ROI.

Colour Features

- *Colour Spaces*

A colour model is an abstract mathematical model describing the way colours can be represented as tuples of numbers, typically as three or four values or colour components. When this model is associated with a precise description of how the components are to be interpreted (viewing conditions, etc.), the resulting set of colours is called a colour space. The choice of a colour space depends on the information to be extracted or on the treatment to be applied.

- *Colour Histograms*

Colour histograms are used to encode the frequency distribution of pixel values either on a whole image or on some region of interest (ROI). Given a finite set of colours, it associates to each colour, its frequency in the image. It is invariant under any geometrical transformation (translation, rotation). When comparing two images or ROI using histograms it is necessary to compute the distance between both histograms using (dis)similarity measures such as Euclidean, χ-square, Kolmogorov-Smirnov and Kuiper distances [Brunelli, 2001]. Classical histograms and most of their derivatives do not take into account spatial distribution of pixels. Nevertheless Blob histograms [Qian, 2000] are able to differentiate pictures having the same colour pixel distribution but containing objects of different sizes. In order to reduce the histogram size, a few representative colours can be selected from the colour space, either using some generic heuristic or by analysing the image. This colour quantisation can be used as a basic descriptor of the image.

- *Colour Moments*

Colour moments have been shown to be both efficient and effective to represent the colour distribution of images [Stricker and Orengo, 1995]. They include the first order moment (mean), the second-order moment (variance) and the third order moment (skewness), thus an image can be described in only nine values (3 moments per colour component).

Textural Features

From a perceptual point of view, a texture may be defined by its ”coarseness”, ”repetitiveness”, “directionality” and “granularity”. However in terms of digital images, the texture of an image or region is defined as a function of the spatial variation in pixel intensities (grey values) [Tuceryan and Jain, 1998]. The analysis of texture is used to determine regions of homogeneous texture, the boundaries between these

regions can then be used to segment the image. Textural classification is also used to associate a region with a textural class (e.g. the material being represented (cotton, sand, etc), or a property of that material (smooth, coarse, etc).

The image analysis applied in the modelling of texture can be divided into three general methods:

- *Statistical Methods*

Statistical methods characterise image texture according to measures of the spatial distribution of grey values (e.g. moments of different orders, correlation functions, related covariance functions).

- *Structural Methods*

The structural methods of texture analysis assume that textures are composed of primitives (called texels). The texture is produced by the placement of these primitives according to certain placement rules. This class of algorithms, in general, is limited in power unless one is dealing with very regular textures. Structural texture analysis consists of two major steps: (a) extraction of the texture elements (texels), and (b) inference of the placement rule. A texture may then be characterized through properties of its texels (average intensity, area, perimeter, etc.) or the texel pattern as defined by the placement rules.

- *Model-based Methods*

Model based texture analysis methods study texture as a linear combination of a set of basis functions. The two main difficulties of such methods are first to find a suitable model to represent the texture (e.g. Fractal Model, Markov model, Fourier filter, Multi-channel Gabor filter, Wavelet transform) and then to compute the accurate parameters which capture the essential perceived characterization of the texture.

Geometrical or Shape-based Features

Using shape descriptors implies being able to extract accurate shapes from an image. Shape descriptors may be based on contour or edge detection together with statistical tools. Such methods are particularly suitable for simple images, which contain one shape easily distinguishable from the background. But better results may be obtained after a segmentation process, which is necessary when dealing when complex images.

Shapes can be described either by their contour or by the region they contain. Moreover they can be either seen from a global or from a local point of view. The former approach, which has been chosen for many shape descriptors, aims at capturing some overall property either of the shape itself (e.g.) or of its contour (e.g. Fourier descriptor). The latter approach is based on local observations on the region or more often on its contour (e.g. inflexion points). Global shape descriptors may be misled when occlusions occur whereas local ones are very sensitive to noise.

- *Region descriptors*

Simple geometrical attributes such as area, eccentricity, bounding box, elongation, convexity, compactness, and circular or elliptic variances are also often used to describe shapes. Although simple to compute, as they can be gathered in attributes vector that may be compared through the use of some accurate distance, their characterisation power is generally too weak to be used in isolation and they are often combined with more complex shape descriptors, such as those provided by geometrical moments.

- *Contour descriptors*

Fourier descriptors are one of the most popular tools to characterise and compare contours. A contour is first sampled into a given number of points. A shape signature function is then applied on the representative points of the contour (e.g. complex shape signature, distance to centroid, area, cumulative angular function, curvature). Such a function produces a set of values, which are encoded through a Fourier transform and then normalized. Other methods include Autoregressive models and Wavelet transforms (particularly suitable for describing high curvature points).

Topological Features

Digital topology deals with properties and features of two-dimensional (2D) or three-dimensional (3D) digital images that correspond to topological properties (e.g., connectedness) or topological features (e.g., boundaries) of objects. Concepts and results of digital topology are used to specify and justify important (low-level) image analysis algorithms, including algorithms for thinning, border or surface tracing, counting of components or tunnels, or region-filling.

4.4.2 Image Segmentation

In order to analyse an image at the level of the objects it contains it is necessary to segment the image so that the image features can be related to the region representing the object. A segmentation process aims at accurately identifying the different areas of an image, either by computing an accurate partition of the image by detecting coherent regions or by detecting the boundaries between regions.

There are three broad approaches which are applied in ROI detection. Affine region detectors which detect regions covariant with a class of affine transformations; for a review of the various methods for detecting these regions see [Mikolajczyk et al. 2006]. The second approach is based on extracting a per pixel salience measure; after grouping pixels of similar saliency a hierarchical representation of salient regions may be obtained [Kadir et al., 2004; Rutishauser et al., 2004; and Walther et al., 2005]. Finally clustering can be applied to ROI as is usual with clustering it is possible to apply three basic methods; generating the clustering bottom-up (starting from a set of seed regions, combine the regions until some stop criteria are reached), top-down (by splitting the image into smaller regions) or a combination of both bottom-up and top-down (several clustering approaches are discussed in Llahi 2005]. The main difficulty in the application of such clustering methods is in deciding how to choose accurate criteria to characterize regions and determining a stopping condition for the algorithm.

4.4.3 Classification and IE

Image Classification and IE can be generally distinguished by processes which categorise the entire scene depicted in the image as oppose to those which categories a ROI or object within that image. Classification of images has been more widely examined due to the fact that image segmentation is not required and thus processes do not have to deal with segmentation inaccuracies, but mainly the difficultly in obtaining annotated images at the region or object level. Recently there has been a number of systems developed which aim to facilitate the process of image annotation [Halaschek, et al. 2005, Petridis et al. 2006, Chakravarthy et al. 2006], such systems are likely to stimulate more research into classification of images at the object level.

The image annotation process associates semantic descriptors, either keywords or ontological concepts, with some visual descriptors of the object contents. A variety of methodologies have been proposed for this process, the simplest approach is to merely consider the co-occurrences between semantic and visual descriptors [Mori 1999], however a number of ML techniques have also been applied to the task including; neural networks [Kosko 1992, Lin 1995, Stamou 2001, Tzouvaras 2003], genetic algorithms [Mitchell 1996], SVM [Vapnik 1995] and HMM [Rabiner 1986, Dugad 1996, Huang 1990].

4.4.4 Evaluation

ImageCLEF: (http://ir.shef.ac.uk/imageclef/) is the cross-language image retrieval track which is run as part of the Cross Language Evaluation Forum (CLEF) campaign.

4.5 Video

One of the features of video analysis is that it brings together a number of media types (image, audio and (via ASR) text) into a single connected setting. Thus video analysis has the opportunity of exploiting the data from these correlated, simultaneous channels, to extract information [Li et al., 2003; Huang et al.,1998 and Sundaram et al., 2000]. In addition there are other features which are specific to the media of video; those that involve the way in which the video frames are linked together using various editing effects (cut, fades, dissolves, etc.). The general video analysis process involves:

- Boundary detection: Segmenting the video stream into shots
- Key-frame extraction: Characterising the content of a shot/video
- Determining what objects are in the shot/video

The primary application of such a process is to allow the index of video in order to make it searchable, for content-based image retrieval systems; however the ultimate goal is to recognise the events portrayed and to understand the narrative of the video.

4.5.1 Feature Extraction

By analysing a video stream in terms of a structured sequence of shots, and then characterising the shots in terms of key-frames, the modelling of video content is reduced to extracting the content of structured still images. This means that the visual features extracted from video are mainly derived from the frame images, which where described above. In addition videos have the features which describe the motion of objects between frames, as well as features relating to the audio channel.

Boundary detection
The identification of the shot boundaries is a key essential step prior to performing shot-level feature extraction and any subsequent scene-level analysis. Shot transitions can be classified as of two types: abrupt transitions (cut) and gradual transitions (fade, wipe, dissolve, etc.). The approaches to detecting these shot transitions either make use of some statistical measure the change in frame features which indicate a transition (a review of several techniques is provided by Boreczky and Rowe [1996], and Dailianas et al. [1995] or use some form of Machine Learning (ML). In general visual features are used to identify the boundaries. However Huang et al [1998] and Sundaram et al [2000] both used a combination of video and audio; based on the idea that the audio should change as well as the video at the shot boundaries.

There are a number of ML approaches to Boundary Detection including nearest neighbour [Kender et al, 1998; Ren and Singh, 2004], neural nets [Ren and Singh, 2004], HMM for both shot boundary detection [Zhang et al. 2006] and higher level topic/story boundary detection [Phung et al. 2002; Chaisorn et al., 2003] and SVM [Feng et al.,2005].

Key-frame extraction
The usual approach to providing a higher level description for a video stream is to extract a set of key-frames which represent a summarisation of the content of the whole stream. The general technique employed is frame clustering [Yeung and Yeo, 1997; Zhuuang 1998; Mundir et al., 2005; Feng et al., 2005], each cluster being centred on a key-frame, thus the key-frames are maximally distinct from one another. The results of applying the clustering technique are dependent upon which features are used, the distance metric employed and the method for determining the number of key-frames (clusters) which sufficiently describe the video. Although clustering is the main key-frame extraction technique, other ML approaches have been applied to the problem, such as genetic algorithms [Avrithis et al., 1999].

Object extraction
The extraction of objects from video applies the techniques described above, for image object identification. As objects can be found in a number of sequential or disparate frames, they can also be used as features in key-frame extraction [Song and Fan, 2005; Lui and Fan, 2005]. Medioni et al. [2001] used object (car) detection with motion analysis to infer the event taking place in the video and thus the behaviour of the actors (drivers).

4.5.2 Classification and IE

As above the semantic classification of objects within a video relies mainly on the techniques applied to still images. However a number of approaches have been applied to the classification of whole videos according to global features using Decision Trees for educational videos [Phung et al., 2002] and news videos [Chaisorn et al., 2003] and more recently the use of SVM to filter videos which contain objectionable content [Jeong, et al., 2006].
Calic et al. [2005] present an interesting paper which discusses the specific issues that relate to the use of semantic information in video, and refers to a number of systems which uses some form of semantics for indexing, classification and retrieval.

4.5.3 Evaluation

TRECVid (http://www-nlpir.nist.gov/projects/trecvid/): The National Institute for Standards and Technology (NIST) have organised a challenge to evaluate video retrieval since 2001.

4.5.4 Systems

MediaMill (http://www.science.uva.nl/research/mediamill/index.php): a semantic video search engine.

aceMedia (http://www.acemedia.org/): knowledge and multimedia content technologies, which provides tools to automatically analyze content, generate metadata and annotation, and support intelligent content search and retrieval services.

4.6 Conclusion and Future Work

This chapter has presented an extensive review of the State of the Art in Classification and Information Extraction, for text, images and videos. Much of this work has focussed on developing the pattern detection algorithms which detect the relevant features in the media type (i.e. words and phrases, textures and areas of interest, slots, etc.). In the future work will continue on improving and refining these algorithms to improve their performance both generically and for specific applications.

Along with the computer science domain, and the world in general, possibly the most interesting challenge and opportunity facing researchers in this domain is the advent of the Internet and World-Wide-Web. To date this has mainly resulted in the provision of a mass of data, requiring the need for processes to work on the large-scale, exemplified by the work on Open Information Extraction [Banko, et al. 2007]. In a recent article [Bhatia and Khalid 2008] a number of possible directions were highlighted for future web-mining including the use of multimedia and multilingual data, in addition the use of the "hidden web", i.e. the databases which are used to generate web pages from user queries, is seen as key. The authors also mention the "wireless web" and "semantic web"; although these are currently less prevalent the wireless web, with the advent of smart-phones, is becoming more commonplace.

Within the MultiMatch project, the use of multimedia and multilingual data is obviously key, and this is reflected in the information extraction techniques used. For textual information extraction the approach adopted (described in D4.4.2) combines the semi-supervised (bootstrapping) approach adopted in the domain-independent KNOWITALL and Open Information Extraction systems. However our domain-specific approach uses domain resources and focused crawling to avoid extracting “out-of-domain” information. In addition, our approach exploits structured-data (in part derived from the hidden web), which provides a "language agnostic approach", which can thus utilise multilingual sources of information. The combination of these two approaches provides the means to gather domain specific information from multilingual sources without the need for annotated training examples. The results of this approach can be most clearly seen in the Faceted Browsing in the MultiMatch User Interface.

Another interesting development which is likely to significantly influence classification and information extraction is the increasing prevalence of Web 2.0 applications which leverage collective intelligence. In some ways applications such as social tagging (of text and images) can be seen as a competitor to the use of automatic techniques, indeed so research indicates that community produced annotations provided more semantically meaningful values than automatically extracted values [Al-Khalifa and Davis 2006]. However in the future the combination of collective and automatic techniques may well provide techniques which exploit the mutual benefit of both approaches. These ideas have been explored with the exploitation of Wikipedia to bootstrap Information Extraction [Fei and Daniel 2007, Fei et al. 2008] and more generally at a recently workshop [CISWeb 2008] and in particular in the paper by Heß, et al. [Heß, et al. 2008]. A future direction of MultiMatch would be to explore the potential benefit of combining Web 2.0 techniques and information extraction in Cultural Heritage.

References

Ahn, D. (2006). The stages of event extraction. Proceedings of the Workshop on Annotating and Reasoning about Time and Events, pages 1–8, Sydney, July 2006.

Ahn, D., Adafre, S. F. and de Rijke, M. (2005). Towards Task-Based Temporal Extraction and Recognition. Dagstuhl Seminar Proceedings of the Workshop on Annotating, Extracting and Reasoning about Time and Events

Aitken, J.S. (2002). Learning Information Extraction Rules: An Inductive Logic Programming approach. In van Harmelen, Prof Frank, Eds. Proceedings 15th European Conference on Artificial Intelligence, pages pp. 355-359, Lyon, France.

Al-Khalifa, H. S. and Davis, H. C. (2006). Measuring the Semantic Value of Folksonomies. In: the Second International IEEE Conference on Innovations in Information Technology, November 17-21, Dubai, UAE.

Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y. (1998). Topic detection and tracking pilot study: Final report. In Proceedings DARPA Broadcast News Transcription and Understanding Workshop (pp. 194–218). Lansdowne, VA: Morgan Kaufmann.

Avrithis, Yannis S., Doulamis, Anastasios D., Doulamis, Nikolaos D. & Kollias, Stefanos D. (1999). A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases Computer Vision and Image Understanding Vol. 75, Nos. 1/2, July/August, pp. 3–24, 1999

Banko, M., Cafarella, M.J., Soderland, S.,Broadhead, M., and Etzioni, O. (2007). Open Information Extraction from the Web. Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)

Beeferman, D., Berger, A., Lafferty, J. (1999). Statistical Models for Text Segmentation. Machine Learning

Berger, A. A Brief Maxent Tutorial http://www.cs.cmu.edu/afs/cs/user/aberger/www/html/tutorial/tutorial.html

Bhatia, MPS and Khalid, Akshi Kumar (2008). Information retrieval and machine learning: Supporting technologies for web mining research and practice. Webology, 5(2), Aricle 55. Available at: http://www.webology.ir/2008/v5n2/a55.html

Bikel, D., Miller, S., Schwartz, R. and Weischedel, R. (1997). Nymble: a High-Performance Learning Name Finder ANLP 1997.

Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In Proceedings of the Eleventh Annual Conference on Computational Learning theory (Madison, Wisconsin, United States, July 24 - 26, 1998). COLT' 98. ACM Press, New York, NY, 92-100. DOI= http://doi.acm.org/10.1145/279943.279962

Boresczky S. and Rowe, L.A. (1996). A comparison of video shot boundary detection techniques, Proc. SPIE 2664, 170-179, 1996

Borthwick, A., Sterling, J., Agichtein, E. and Grishman, R. (1998). Exploiting Diverse Knowledge Sources via Maximum Entropy in Named Entity Recognition, WVLC 98.

Boser, B. E., Guyon, I. M. and Vapnik, V. N (1992). A training algorithm for optimal margin classifiers. In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144-152, Pittsburgh, PA, 1992. ACM Press.

Bunescu, R. and Mooney, R.J. (2004). Collective Information Extraction with Relational Markov Networks Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-2004), pp. 439-446, Barcelona, Spain, July 2004.

Calic, J., Campbell, N., Dasiopoulou S. and Kompatsiaris, Y. (2005). An Overview of Multimodal Video Representation for Semantic Analysis. European Workshop on the Integration of Knowledge, Semantics and Digital Media Technologies, EWIMT 2005, London, UK, November 30 - December 1, 2005

Cardie, C. and Wagstaff, K. (1999). Noun phrase coreference as clustering. In Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, pages 82-89.

Chaisorn, L., Koh, C., Zhao, Y., Xu, H., Chua, T.-S and Qi, T.(2003). Two-level multimodal framework for news story segmentation of large video corpus. 12th Text Retrieval Conference, Gaithersburg, MD, USA, 2003.

Chakravarthy, A., Ciravegna, F. and Lanfranchi, V. (2006). AKTiveMedia: Cross-media Document Annotation and Enrichment. In Poster Proceedings of the Fifteenth International Semantic Web Conference (ISWC2006).

Ciravegna, F. (2001). Adaptive Information Extraction from Text by Rule Induction and Generalisation, in Proceedings of 17th International Joint Conference on Artificial Intelligence (IJCAI 2001), Seattle, August 2001.

Ciravegna, F., Chapman, S., Dingli, A., and Wilks, Y. (2004). Learning to Harvest Information for the Semantic Web. Proceedings of the 1st European Semantic Web Symposium, Heraklion, Greece, May 10-12, 2004

CISWeb 2008. 1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (2008) at the 5th European Semantic Web Conference (ESWC 2008)

Cowie, J., Guthrie, L., Pustejovsky, J., Wakao, T., Wang, J., and Waterman, S. (1993) The Diderot Information Extraction System, to appear in Proc. First PA CLING Conference, Vancouver.

Culotta, A. and Sorensen, J. (2004). Dependency Tree Kernels for Relation Extraction. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics.

Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V. (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, July 2002

Dailianas, A., Allen, R.B. and England, P. (1995). Comparison of automatic video segmentation algorithms, Proc. SPIE Photonics West, 2615, 2-16, 1995.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39(1):1--38, 1977.

Diederich, J., Kindermann, J., Leopold, E., and Paaß, G. (2003). Authorship attribution with support vector machines. Applied Intelligence, 19(1/2), 109–123.

Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Rajagopalan, S., Tomkins, A., Tomlin, J. A., and Zien, J. Y. (2003). SemTag and seeker: bootstrapping the semantic web via automated semantic annotation. In Proceedings of the 12th international Conference on World Wide Web (Budapest, Hungary, May 20 - 24, 2003). WWW '03. ACM Press, New York, NY, 178-186. DOI= http://doi.acm.org/10.1145/775152.775178

Drucker, H., Vapnik, V., and Wu, D. (1999). Support vector machines for spam categorization. IEEE Transactions on Neural Networks, 10(5), 1048–1054.

Dumais, S. T. and Chen, H. (2000). Hierarchical classification of Web content. In N. J. Belkin, P. Ingwersen, and M.-K. Leong (Eds.), Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval (pp. 256–263). Athens, GR: ACM Press, New York, US.

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics 3: 32-57

Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S., Weld, D. S., and Yates, A. (2004). Web-scale information extraction in knowitall: (preliminary results). In Proceedings of the 13th international Conference on World Wide Web (New York, NY, USA, May 17 - 20, 2004). WWW '04. ACM Press, New York, NY, 100-110. DOI= http://doi.acm.org/10.1145/988672.988687

Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D. S. and Yates, A. (2005) Unsupervised Named-Entity Extraction from the Web: An Experimental Study. Artificial Intelligence, 165, pp. 91-134.

Fall, C. J., Törcsvári, A., Benzineb, K., and Karetka, G. (2003). Automated categorization in the International Patent Classification. SIGIR Forum, 37(1).

Fan, W., Wallace, L., Rich, S., and Zhang, Z. (2006). Tapping the power of text mining. Commun. ACM 49, 9 (Sep. 2006), 76-82. DOI= http://doi.acm.org/10.1145/1151030.1151032

Feng, H., Fang, W., Liu, S., and Fang, Y. (2005). A new general framework for shot boundary detection and key-frame extraction. In Proceedings of the 7th ACM SIGMM international Workshop on Multimedia information Retrieval (Hilton, Singapore, November 10 - 11, 2005). MIR '05. ACM Press, New York, NY, 121-126. DOI= http://doi.acm.org/10.1145/1101826.1101847

Finn, A. and Kushmerick, N. (2004). Multi-level Boundary Classification for Information Extraction. In Proceedings of the 15th European Conference on Machine Learning, Pisa, Italy.

Finn, A. and Kushmerick, N. (2003). Active learning selection strategies for information extraction. ECML-03 Workshop on Adaptive Text Extraction and Mining (Croatia)

Freitag, D. and Kushmerick, N. (2001). Boosted Wrapper Induction. AAAI 2000, 577-583

Freund, Y. and Schapire, R. E. (1999). A Short Introduction to Boosting: Introduction to Adaboost. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999.

Giorgetti, D. and Sebastiani, F. (2003). Automating survey coding by multiclass text categorization techniques. Journal of the American Society for Information Science and Technology, 54(12), 1269–1277.

Gu, X.-D., Chen, J., Ma, W.-Y. and Chen, G.-L. (2002). Visual Based Content Understanding towards Web Adaptation, Proc. Adaptive Hypermedia and Adaptive Web-Based Systems, Malaga, Spain, 2002, pp. 164-173

Hacioglu, K., Chen, Y. and Douglas, B. (2005). Automatic Time Expression Labeling for English and Chinese Text. In Linguistics and Intelligent Text Processing, Volume 3406, 2005, 548-559

Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B. and Hendler, J. (2005) Photostuff - an image annotation tool for the semantic web. In 4th International Semantic Web Conference. 2005.

Hayes, P. J. and Weinstein, S. P. (1990). Construe/Tis: a system for content-based indexing of a database of news stories. In A. Rappaport and R. Smith (Eds.), Proceedings of IAAI-90, 2nd Conference on Innovative Applications of Artificial Intelligence (pp. 49–66).: AAAI Press, Menlo Park, US.

Heß, A,, Maaß, C. and Dierick, F. (2008). From Web 2.0 to Semantic Web: A Semi-Automated Approach. ESWC 2008 Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb 2008), Tenerife, Spain

Heyer, L.J., Kruglyak, S. and Yooseph, S., (1999). Exploring Expression Data: Identification and Analysis of Coexpressed Genes, Genome Research 9:1106-1115

Holland, J. H. (1975), Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor

Honavar, V., Silvescu A., Reinoso-Castillo J., Caragea, D., Andorf, C. and Dobbs, D. (2001). Ontology-driven information extraction and knowledge acquisition from heterogeneous, distributed biological data sources, in: Proceedings of the IJCAI2001 Workshop on Knowledge Discovery from Heterogeneous, Distributed, Autonomous, Dynamic Data and Knowledge Sources, 2001.

Hotho, A., Staab, S., and Stumme, G. (2003). Wordnet improves Text Document Clustering. In Proc. of the Semantic Web Workshop of the 26th Annual International ACM SIGIR Conference, Toronto, Canada, 2003.

Huang, J. Liu, Z. and Wang, Y. (1998). Integration of audio and visual information for content-based video segmentation. IEEE Int'l Conf. Image Processing (ICIP98), Special SessioSession on Content-Based Video Search and Retrieval. Oct. 1998. Chicago.

Ireson, N., Ciravegna, F., Califf, M. E., Freitag, D., Kushmerick, N. and Lavelli, A. (2005). Evaluating Machine Learning for Information Extraction, 22nd International Conference on Machine Learning (ICML 2005), Bonn, Germany, 7-11 August, 2005

Iria, J., Ireson, N. and Ciravegna, F. (2006) An Experimental Study on Boundary Classification Algorithms for Information Extraction using SVM

Jeong, C. Y., Han, S. W., and Nam, T. Y. (2006). Automatic Objectionable Video Classification System. Internet and Multimedia Systems and Applications 2006

Joachims, T. (1998). Text categorization with support vector machines: learning with many relevant features. In C. Nédellec and C. Rouveirol (Eds.), Proceedings of ECML-98, 10th European Conference on Machine Learning (pp. 137–142). Chemnitz, DE: Springer Verlag, Heidelberg, DE. Published in the "Lecture Notes in Computer Science" series, number 1398.

Kan, M.-Y. (2001). Combining visual layout and lexical cohesion features for text segmentation. Columbia University Computer Science Technical Report, CUCS-002-01. 2001

Karkaletsis, V., Pailouras, G. and Spyropoulos, C. D. (2000). Learning decision trees for named-entity recognition and classification. In Proceedings of the ECAI Workshop on Machine Learning for Information Extraction, 2000

Kehagias, A., Petridis, V., Kaburlasos, V. G., and Fragkou, P. (2003). A comparison of word- and sense-based text categorization using several classification algorithms. Journal of Intelligent Information Systems, 21(3), 227–247.

Kender, J. R. and Yeo, B.-L. (1998). Video Scene Segmentation Via Continuous Video Coherence, Proc. CVPR '98, pp 367-373, June 1998.

Koppel, M., Argamon, S., and Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. Literary and Linguistic Computing, 17(4), 401–412.

Koster, C. H. and Seutter, M. (2003). Taming wild phrases. In F. Sebastiani (Ed.), Proceedings of ECIR-03, 25th European Conference on Information Retrieval (pp. 161–176). Pisa, IT: Springer Verlag.

Lam, W. and Ho, C. Y. (1998). Using a generalized instance set for automatic text categorization. In Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval, pages 81–89, Melbourne, AU, 1998.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Lavelli, A., Califf, M. E., Ciravegna, F., Freitag, D., Giuliano, C., Kushmerick, N., and Romano, L. (2004). A Critical Survey of the Methodology for IE Evaluation. Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, May 26-28, 2004

Lenat, D. B. (1995). Cyc: A Large-Scale Investment in Knowledge Infrastructure. Communications of the ACM 38, no. 11 (November 1995).

Lewis, D. D. (1997). Reuters-21578 text Categorization test collection. Distribution 1.0. README file (version 1.2). Manuscript, September 26, 1997. http://www.daviddlewis.com/resources/testcollections/reuters21578/readme.txt

Lewis, D. D., Yang, Y., Rose, T. G., and Li, F. (2004). RCV1: A New Benchmark Collection for Text Categorization Research. J. Mach. Learn. Res. 5 (Dec. 2004), 361-397.

Li, D., Dimitrova, N., Li, M., and Sethi, I. K. (2003). Multimedia content processing through cross-modal association. In MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia, pages 604–611, New York, NY, USA, 2003. ACM Press.

Li, Y. and Bontcheva, K. and Cunningham, H. (2005). SVM Based Learning System For Information Extraction. In: Proceedings of Sheffield Machine Learning Workshop. Lecture Notes in Computer Science. Springer Verlag.

Li, Y. H. and Jain, A. K. (1998). Classification of text documents. The Computer Journal, 41(8):537–546, 1998.

Liu, B. and Chen-Chuan-Chang, K. (2004). Editorial: special issue on web content mining. *SIGKDD Explor. Newsl.* 6, 2 (Dec. 2004), 1-4. DOI= http://doi.acm.org/10.1145/1046456.1046457

Liu, L. and Fan, G. (2005). Combined key-frame extraction and object-based video segmentation. IEEE transactions on circuits and systems for video technology. 2005, vol. 15, no7, pp. 869-884.

Liu, T., Chen, Z., Zhang, B., Ma, W., and Wu, G. (2004). Improving Text Classification using Local Latent Semantic Indexing. In Proceedings of the Fourth IEEE international Conference on Data Mining (Icdm'04) - Volume 00 (November 01 - 04, 2004). ICDM. IEEE Computer Society, Washington, DC, 162-169.

Luo, J., Shen, J. and Xie, C. (2004) Segmenting the Web Document with Document Object Model Services Computing, 2004 IEEE International Conference on (SCC'04), pp. 449-452

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297

Maron, M. (1961). Automatic indexing: an experimental inquiry. Journal of the Association for Computing Machinery, 8(3), 404–417.

Masand, B., Linoff, G. and Waltz, D. (1992). Classifying news stories using memory-based reasoning. In Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval, pages 59–65, Kobenhavn, DK, 1992.

McCallum, A. and Li, W. (2003). Early Results for Named Entity Recognition with Conditional Random Fields, Fetures Induction and Web-Enhanced Lexicons, CoNLL 2003.

McCallum, A. and Nigam K. "A Comparison of Event Models for Naive Bayes Text Classification". In AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48. Technical Report WS-98-05. AAAI Press. 1998.

Medioni, G., Cohen, I., Bremond, F., Hongeng, S. and Nevatia, R. (2001) Event Detection and Analysis from Video Streams. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 873-889, Aug., 2001.

Michelson, M. and Knoblock, C. A. (2005). Semantic annotation of unstructured and ungrammatical text. In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005).

Moschitti, A. and Basili, R. (2004). Complex linguistic features for text classification: A comprehensive study. In S. McDonald and J. Tait (Eds.), Proceedings of ECIR-04, 26th European Conference on Information Retrieval Research (pp. 181–196). Sunderland, UK: Springer Verlag, Heidelberg, DE. Published in the "Lecture Notes in Computer Science" series, number 2997.

Mukherjee, S., Yang, G., Tan, W. and Ramakrishnan, I. V. (2003). Automatic Discovery of Semantic Structures in HTML documents. International Conference on Document Analysis and Recognition (ICDAR). 2003

Mundir, P., Rao, Y. and Yesha, Y. (2005). Keyframe-based Video Summarization using Delaunay Clustering

Muslea, I. (1999). Extraction patterns for information extraction tasks: A survey. AAAI 1999 Workshop on Machine Learning for Information Extraction.

Nigam, K. (2001). Using Unlabeled Data to Improve Text Classification. Ph.D. Dissertation, Carnegie Mellon University.

Nigam, K. and Ghani, R. (2000). Analyzing the applicability and effectiveness of co-training. In A. Agah, J. Callan, and E. Rundensteiner (Eds.), Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management (pp. 86–93). McLean, US: ACM Press, New York, US.

Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. M. (2000). Text classification from labeled and unlabeled documents using EM. Machine Learning, 39(2/3), 103–134.

Patwardhan, S. and Riloff, E. (2006). Learning Domain-Specific Information Extraction Patterns from the Web. http://www.cs.utah.edu/~sidd/papers/PatwardhanR06.pdf

Petridis, K., Anastasopoulos, D., Saathoff, C. , Timmermann, N., Kompatsiaris I. and Staab S.,

M-OntoMat-Annotizer: Image Annotation. Linking Ontologies and Multimedia Low-Level Features. Engineered Applications of Semantic Web Session (SWEA) at the 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES 2006), Bournemouth, U.K., 9-11 October 2006.

Phung, D. Q., Duong, T. V., Venkatesh, S., and Bui, H. H. 2005. Topic transition detection using hierarchical hidden Markov and semi-Markov models. In *Proceedings of the 13th Annual ACM international Conference on Multimedia* (Hilton, Singapore, November 06 - 11, 2005). MULTIMEDIA '05. ACM Press, New York, NY, 11-20. DOI= http://doi.acm.org/10.1145/1101149.1101153

Pinto, D., McCallum, A., Wei, X., and Croft, W. B. (2003). Table extraction using conditional random fields. In *Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in informaion Retrieval* (Toronto, Canada, July 28 - August 01, 2003). SIGIR '03. ACM Press, New York, NY, 235-242. DOI= http://doi.acm.org/10.1145/860435.860479

Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D., and Kirilov, A. (2004). KIM – a semantic platform for information extraction and retrieval. Nat. Lang. Eng. 10, 3-4 (Sep. 2004), 375-392. DOI= http://dx.doi.org/10.1017/S135132490400347X

Quinlan, J.R. (1993). C4.5: Programs for Machine Learning. Morgan Kauffman.

Ratnaparkhi, Adwait, R. (1997). A Simple Introduction to Maximum Entropy Models for Natural Language Processing. IRCS Report 97--08, University of Pennsylvania, 3401 Walnut Street, Suite 400A, Philadelphia, PA, May 1997.

Ren, W. and Singh, S. (2004). Automatic Video Shot Boundary Detection Using Machine Learning. In Intelligent Data Engineering and Automated Learning – IDEAL 2004

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Inf. Process. Manage. 24, 5 (Aug. 1988), 513-523. DOI= http://dx.doi.org/10.1016/0306-4573(88)90021-0

Schapire, R. E. (1990). The strength of weak learnability. Machine Learning, 5(2):197--227, 1990.

Schapire, R. E. and Singer, Y. (2000). BoosTexter: a boosting-based system for text categorization. Machine Learning, 39(2/3), 135–168.

Schank, R.C. (1975). Conceptual Information Processing. New York: Elsevier.

Sebastiani, F. (1999a) Machine learning in automated text categorisation: A survey. Technical Report IEI-B4-31-1999, Istituto di Elaborazione dell'Informazione, C.N.R., Pisa, IT, 1999.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. ACM Computing Surveys, Vol. 34, No. 1, March 2002, pp. 1–47.

Sekine, S., Grishman, R. and Shinnou, H. (1998) A Decision Tree Method for Finding and Classifying Names in Japanese Texts, WVLC 98.

Shannon, C. E. (1948). A Mathematical Theory of Communication, Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.

Shen, D., Zhang, J., Zhou, G., Su, J., and Tan, C. (2003). Effective adaptation of a Hidden Markov Model-based named entity recognizer for biomedical domain. In Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine - Volume 13 (Sapporo, Japan, July 11 - 11, 2003). Annual Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ, 49-56.

Song, X. and Fan, G. (2005). Joint Key-Frame Extraction and Object-Based Video Segmentation. wacv-motion, pp. 126-131, IEEE Workshop on Motion and Video Computing (WACV/MOTION'05) - Volume 2, 2005.

Stamatatos, E., Fakotakis, N., and Kokkinakis, G. (2000). Automatic text categorization in terms of genre and author. Computational Linguistics, 26(4), 471–495.

Stricker, M. and Orengo, M. (1995). Similarity of color images, Proc. SPIE, vol. 2420, pp. 381–392, 1995

Sundaram, H. and Chang, S.-F. (2000). Determining Computable Scenes in Films and their Structures using Audio-Visual Memory Models. ACM Multimedia 2000, Oct 30 - Nov 3, Los Angeles, CA.

Sutton, C., McCallum, A. and Rohanimanesh, K. (2006). Dynamic Conditional Random Fields. Journal of Machine Learning Research (JMLR), Vol. 7, 2006.

Tuceryan, M. (1998). Textural Analysis. In The Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau and P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co., 1998.

Tuceryan, M. and Jain, A. K. (1998). Texture Analysis. In The Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co., 1998.

Turney, P. D. and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems, 21(4), 315–346.

Viterbi, A J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 13(2):260–269, April 1967. (The Viterbi decoding algorithm is described in section IV.)

Yang, Y. and Zhang, H. (2001). HTML Page Analysis Based on Visual Cues, Proc. of 6th International Conference on Document and Analysis, Seattle, USA, 2001

Yeung, M., and Yeo, B.-L. (1997). Video visualization for compact presentation and fast browsing of pictorial content. IEEE Trans. Circuits Syst. Video Technol. 7, 5 (Oct. 1997), 771–785

Wallach, H. M. (2004). Conditional Random Fields: An Introduction. Technical Report MS-CIS-04-21. Department of Computer and Information Science, University of Pennsylvania, 2004.

Wu, Fei and Weld, Daniel S. (2007). Automatically Semantifying Wikipedia. Proceedings of the 16th Conference on Information and Knowledge Management (CIKM 2007)

Wu, Fei, Hoffmann, Raphael , and Weld, Daniel S. (2008). Information Extraction from Wikipedia: Moving Down the Long Tail. Proceedings of the 14th ACM SIGKDD international Conference on Knowledge Discovery & Data Mining (KDD 2008)

Zhang, W., Lin, J., Chen, X., Huang, Q. and Liu, Y. (2006). Video Shot Detection Using Hidden Markov Models with Complementary Features. First International Conference on Innovative Computing, Information and Control - Volume III (ICICIC'06), 593-596

Zhuang, Y., Rui, Y., Huang, T. S. and Mehrotra, S. (1998). Adaptive Key Frame Extraction Using Unsupervised Clustering

5. Multilingual/Multimedia Indexing

by Martha Larson and Jaap Kamps

This chapter describes the state-of-the-art in the indexing of cultural heritage (CH) documents in various languages and of various media types. First, we discuss the special characteristics of cultural heritage documents. Second, we discuss the general approaches to indexing that are currently state of the art. Third, we provide additional details of indexing approaches for the four types of media treated in the MultiMatch project: text, images, audio and video. The final section offers an overview that relates the state-of-the-art in multilingual/multimedia indexing to the results that have been achieved over the course of the MultiMatch project.

5.1 Indexing Cultural Heritage Documents

Typical for the cultural heritage domain are collections that have been developed over large spans of time and that are curated by highly trained professionals using systems that have been refined through a tradition of use that pre-dates the digital age. These collections may be stored in databases and not exposed on the internet in a way that makes it easily indexable. Content in the cultural heritage domain includes the entire spectrum of media, from text to audio through video. Finally, users are making a increasingly important contribution to the organization and annotation of cultural heritage content on line.

As stated in Chapter 2,[64] metadata plays a crucial role in providing access to cultural heritage. Cultural heritage institutions have invested enormous effort in gathering information about their precious objects, usually stored separately in library catalogue records, archival inventories, or museum registers. In non-digital collections, these descriptions of CH objects form the main access points for organizing, selecting and retrieving objects. For example, a controlled vocabulary that captures the topical subject of an object by a numerical code, such as DDC [2006] or UDC [2006], can provide subject access to CH objects even across language boundaries. However, combining different CH collections also implies combining different traditions of description, different controlled vocabularies, and different intended audiences in mind. Even when syntactically coded in a uniform format, such as [DCMI, 2006; RDF, 2006; OWL, 2006], the metadata will reflect the provenance of the particular object. Making sense of heterogeneous metadata is one of the greatest challenges for today's cultural heritages institutions.

It is an open problem how to provide uniform access to the myriads of formats in current combined collections, without the need for expensive manual or supervised revision of existing descriptions. There are two current approaches directly addressing this problem: The first approach is to treat the controlled vocabularies as a rigorous ontology, and attempt to define mappings between the different systems (e.g., [STITCH, 2006]). That is, the problem is now translated into a semantic interoperability or ontology mapping problem. The state-of-the-art techniques are far from fool-proof; manual supervision is necessary [Handschuh and Staab, 2003]. Such effort is needed for each mapping covering a single pair of vocabularies. The viability of this approach depends on the number of different vocabularies involved, and on their rigorousness. The second approach is to treat the heritage descriptions as noisy and uncertain, and apply powerful methods from modern text retrieval (e.g., [MuSeUM, 2006]). Specifically, this approach makes very few assumptions on the presence or encoding of particular metadata, but exploits it whenever present. In essence, this is the famous "dumb-down principle" [Weibel, 1995]: although metadata is based on a specific thesaurus or ontology, we can always fall back on the description of the terms in ordinary language.

In collections of digital CH objects, the combination of searching content as well as metadata provides powerful finding aids [Lesk, 2005]. In many cases, the "content" of a digital object will take the form of free text, which either describes the object, or in particular cases, such as that of literary works, constitutes the object. In the case of multimedia, the "content" of the digital object is not text but is rather the so-called "essence," the actual audio, image or text file. The combination of low-level image features with metadata

[64] In Chapter 2, we describe the metadata schemes typically adopted to describe digital objects. Chapter 3 discusses how Information Extraction techniques can be used to create explicit representations, i.e. metadata, from the information implicit in unstructured text. In this Chapter we examine the issues that have to be faced when applying or using manually or automatically assigned metadata for information access.

may be helpful for particular queries, even if it does not generally contribute to improved retrieval performance [Byrne 2003]. In the case of audio archives, representations of the content of documents can take the form of transcripts generated by automatic speech recognition. The discussion of the CLEF CL-SR track in Pecina et al [2008] supports the conclusion that it is preferable not to have to rely entirely on speech recognition transcripts and that, if available, human-generated metadata makes a critical contribution to retrieval in spoken word collections.

Currently, much cultural heritage material is buried in databases and is not exposed on the internet in a way that makes it easily indexable and therefore findable. The work of Byrne [2008a, 2008b] discusses both the high potential of the well-curated data that cultural heritage solutions have accumulated over time and currently store in databases and well as the difficulties inherent in defining a mapping between databases and the RDF triples that are used to represent data in the semantic web. Techniques for exporting data from relational databases into RDF format are currently a subject of ongoing research and no final answer has yet been achieved. In particular, the question remains open whether or not the process should be performed automatically [Byrne 2008].

The metadata describing a cultural heritage collection is rarely static, but continues to develop and grow with the collection. Indexing approaches can take advantage or attempt to foster this growth. Aihara et al. [2008] present a system that supports the creation and sharing of cultural heritage objects. The system is designed for both professionals and casual users. Metadata collected by the system addresses the challenges of (1) dealing with variation in the description of cultural objects (2) connecting different versions of the same object (3) generating multiple metadata representations different user groups that use the cultural heritage object (i.e., experts and non-experts).

On the internet, user communities take root and develop. These communities interact, create content and tag, comment upon and review that content. The structure of these communities is an important source of information. Some communities engage in concerted effort to label web sites that are relevant to their interests with tags that will make them easily retrievable. Cultural heritage collections on-line can also make use of user-contributed annotations such as tags [van der Sluijs, 2008].

In many domains, user satisfaction with information is directly related to its perceived authority and trustworthiness. In the cultural heritage domain, the credibility of the source of the information is of primary importance for expert searchers [Amin, et al. 2008] and is presumably also an important aspect for casual users. Indexing techniques for the cultural heritage domain must represent not only the topical content of information, but also its reliability.

5.2 Indexing Approach

There are two basic approaches to indexing cultural heritage documents in various languages and of various media types. The first approach indexes all document sorts and media types separately, and later integrates the results using distributed indexing techniques and fusion methods similar to those used in distributed IR [Callan et al., 1995]. The second approach is to define a single, complex document type definition that will form the basis for all material to be indexed: documents of various media types (text, audio, image, video, or mixed-content) and accompanying metadata. Despite much progress in searching by content in multimedia databases [Faloutsos, 1996] there is a clear trend toward the combination of various modalities [de Vries et al., 2000; Snoek and Worring, 2005], as mentioned above. Existing generic standards such as MPEG-7 (which is part of the XML family of languages) are able to cater for such a data model by incorporating multimedia content and metadata in a single semi-structured document.

Interestingly, researchers in IR are travelling down a similar path by integrating result ranking in the core of XML databases (e.g., [List et al., 2005]). Such systems radically depart from the standard "document as a bag-of-words" approaches, by preserving the document structure and using region algebras to score individual document components [Burkowski, 1992; Clarke et al., 1995]. The resulting database provides a general framework for complex object retrieval, allowing for a range of retrieval approaches without the need to re-index the collection. The most recent proposals allowing for complex retrieval models can be defined as logical queries on an XML database [Hiemstra and Michajlovic, 2005]. Currently available XML databases or retrieval systems such as the Cheshire [2006], MonetDB [2006], Lucene [2006], and MILOS

[Amato et al., 2004] systems allow - to a greater or lesser extent- - this flexibility. It is an open question how to extend any of the existing systems to the specific demands of cultural heritage retrieval.

5.3 Indexing CH Media Types

5.3.1 Indexing Text

The state of the art indexing methods of cross-language information retrieval use dedicated tokenization methods [Hollink et al., 2004]. Some approaches consider various language-dependent morphological normalization techniques, such as lemmatization or stemming, and other approaches consider language-independent techniques, such as character n-gramming. Although approaches to the indexing of free-text are well studied, it is a major challenge how to preserve the document structure, if available, in the index, and how to ensure that the metadata associated with the documents is indexed in separate fields. The issue of how to index text from multilingual sources is tightly connected with the method chosen for retrieval, and the reader is referred to the discussion on Multilingual/Multimedia Retrieval below for more details about the indexing of multilingual text. As mentioned above, various metadata---both from the original CH documents as well as those automatically assigned by extraction and classification tools---are crucial for providing access to CH documents.

5.3.2 Indexing Images

For indexing images, the state-of-the-art complex object database naturally supports indexing the binary image, features extracted from the image, and the metadata attributed to the images. Highly sophisticated methods have been developed for content-based image retrieval [Smeulders et al., 2000]. Examples are the extraction of salient features of images, such as low-level visual properties of texture, colour, and shape, or various multi-scale robust features. The output of visual feature extractors is typically stored in a dedicated indexing structure separated from the main index. Effective image retrieval methods still heavily rely on metadata, so all available textual information about the images will be carefully indexed. The images may be manually annotated, or semi-automatically derived from the textual context [Barnard and Forsyth, 2001; Jeon et al., 2003]. In some cases, annotations are critical to image retrieval and content-based features make limited contributions [Byrne, 2003].

5.3.3 Indexing Speech and Audio

Speech recognition technology has progressed to a level that is sufficient to make speech recognition transcript derived indexing features effective for text retrieval. Retrieval performance comparable to that achievable on text has been reported in the broadcast news domain [Garofolo et al., 2000]. Typical cultural heritage content differs from broadcast news in that it is not necessarily structured into stories and that the vocabulary used by the speakers and background conditions used for recording are significantly less predictable. In the cultural heritage domain, spoken audio tends include a large proportion of spontaneous speech, which tends to be heterogeneous and unstructured. The cultural heritage domain includes collections containing oral history interviews, lectures, talkshows and studio discussion as well as user generated podcasts are typical.

Providing access to spoken cultural heritage content presents a number of challenges. Speech recognition word error rates for heterogeneous spoken audio content are highly variable [Huijbregts et al. 2007]. In cultural heritage collections such as oral history interviews, particular challenges include spontaneous speech, emotional speech, speech of elderly speakers, highly accented and regionally specific speech, foreign words, names and places [Byrne et al., 2004]. Improving speech recognition performance is an important priority for spoken interview collections [de Jong et al. 2008].

In additional to speech recognition, audio segmentation and spoken content categorization are two other areas related to providing access to oral history collections [Byrne et al., 2004]. Creating appropriate segments is important not only for indexing, but also for display of spoken content results to the user in the interface in a way in which they can be easily skimmed [de Jong et al. 2008]. Providing access to the full scope of the world's languages is an important challenge for speech indexing, since many linguistic resources are required to develop systems for new languages [Goldman, et al. 2005]. Finally, spoken content indexing has a high entry threshold: collection specific systems are time consuming and expensive to develop. Work directed towards affordable access to spoken content collections, e.g., Ordelman et al. [2008],

specifically targets the challenge of "maximizing the potential of the collection while minimizing development costs."

5.3.4 Indexing Video

Traditional approach to indexing video would separate the audio and video streams, where the audio stream is indexed as text using automatic speech recognition techniques, and the video stream is - after shot boundary detection and key frame extraction - converted to content based image features [Hauptmann and Witbrock, 1997]. The integrated use of different sources is an emerging trend in video indexing research. This is a semantically informed multi-modal approach in which the visual, auditory and textual modalities are combined [Snoek and Worring, 2005]. First, a multi-modal approach to content segmentation is proposed; some of the content elements may be converted to text. Then, the different modalities are integrated to enhance the classification accuracy on semantic subtasks such as genre detection, logical units, and named events.

5.4 Moving forward the state of the art of multimedia indexing within MultiMatch

During the course of the MultiMatch project, the state of the art in multimedia indexing was pushed forward in a number of areas. These areas represent the priorities that were set within the MultiMatch project in order to guarantee that the time and resources necessary to achieve substantive scientific progress were available. Because of the tight relationship between multilingual indexing and multilingual retrieval, multilingual research is reported in Chapter 7 below.

Structuring spoken audio An approach to spoken content segmentation was developed in MultiMatch based on TextTiling techniques [Hearst, 1997] applied to speech recognition transcripts. The segments generated with our technique were used as a basic indexing unit and also as structural unit for the display of audio documents in the user interface. The segmentation technique was shown to be robust to the word recognition error levels that characterize transcripts of spontaneous speech, common in the cultural heritage domain. The work is a result of the collaboration between DCU and UvA reported on in [Fuller, 2008] and in D 4.2.2 "Revised Text/Image/Speech/Video Indexing Components for the 2nd Prototype." Subsequently, the technique was demonstrated to be extendable to other languages. Spoken audio segments automatically generated from English to Dutch, German, Italian and Spanish and the audio segments were indexed in the MultiMatch PT2.

Indexing features for spoken audio Investigations into spoken content retrieval have shown that human-generated metadata provides effective indexing features and that the importance of content-based features, i.e., features derived from speech recognition transcripts, may be limited [Pacina et al., 2007]. At UvA, the following research question was posed: "Do we really need speech recognition transcripts for spoken content retrieval?" As our domain of investigation we chose cultural heritage related podcasts, audio series on the internet. We started with the assumption that there are multiple information needs that motivate users to search for podcasts and that it is important to investigate a range of different kinds of information needs in order to determine if any of them specifically profit from speech transcript derived indexing features. The first challenge encountered by our research was the lack of literature concerning the information needs that motivate users to search for podcasts on the internet. To fill this gap, we carried out an extensive survey to identify user information needs in podcast retrieval [Besser et al., 2008; Besser, 2008]. On the basis of the information needs determined in this survey, we formulated a set of queries to test on our experimental podcast corpus. We carried out retrieval experiments using both metadata-based indexing features and speech-recognition based indexing features. In order to perform quantitative evaluation, we carried out human relevance judgments of the documents. We were able to provide a positive answer to our initial research question: speech recognition transcripts were needed for retrieval in the cultural heritage podcast domain. Additionally, our analysis revealed that speech recognition-based features make different contributions to spoken content retrieval depending on the type of query, which is in turn dependent on the underlying information need. This work is reported in [Besser, 2008] and in D7.3 "Evaluation of Second Prototype."

Classification of video Automatic assignment of topic labels to spoken audio is a task that contributes to providing access to spoken audio collections, as in Byrne [2004]. We investigated whether subject labels assigned by archivists can be automatically generated using speech recognition-based features. In particular, we focused on the domain of dual language video, video containing two languages. We were motivated to

focus on dual language video by the existence of interviews with English speakers in the BandG archive that would be useful to a searcher who speaks English, but are effectively lost since they are embedded within Dutch language television documentaries. In order to promote larger scale research effort in this area, UvA and DCU developed and carried out a pilot track, VideoCLEF in the CLEF Cross Language Information Retrieval Campaign [Larson et al., 2008a, 2008b]. We carried out experiments on classification of video transcripts using two approaches. The first made use of training data in the form of speech recognition transcripts from the same domain as the test data. The second assumed a scenario in which no domain-specific training data was available and, instead, training data was collected from external data sources. We were able to show that both metadata and speech transcript-based features can be exploited for automatic generation of subject class labels. Although classification rates are better when domain-specific training data is used, it is also possible to classify video data into subject classes making use of training data collected from Wikipedia. This work is reported on in He et al. [2008], Newman and Jones [2008], 4.4.2 "Semantic Analysis and Classification Component and Documentation for 2nd prototype," and D7.3 "Evaluation of Second Prototype."

Representing complex objects Naked or de-contextualized multimedia files are of limited usefulness to information searchers. If related resources are connected to form objects, searchers can be presented with more directly usable results. Objects that consist of linked files from multiple sources are often called complex objects. In the case of time continuous media such as audio and video, it is helpful for complex objects to be structured and for links between resources to be established on a subdocument level and mediated by way of time codes. Two types of complex multimedia objects of this sort were created in MultiMatch. First, structured podcasts, as mentioned above, were developed and put to use in the MultiMatch PT2. Here, timecodes marked the segment boundaries within the podcast and served to couple the audio essence with segment level term clouds, which served as surrogates in the user interface [Fuller et al., 2008]. At the file level, podcast episodes were associated with their feed metadata and also with metadata describing the feed that contained them. Second, structured video was used in PT2. Videos were represented with their file-level metadata, and also with a structured representation that coupled shot level segmentation, with representative keyframes and speech transcript-based features [Carmichael, et al., 2008]. One of the challenges of indexing complex objects is choosing the correct metadata format to represent them. MultiMatch provided the context for a master's thesis written at the University of Amsterdam dedicated to the issue of metadata for access to web lectures [Kapferer, 2008].

Credibility in internet audio Credibility of information is important for search in the cultural heritage domain. The issue is particularly critical in the case of podcasts, which are often user generated and not published by a source with immediate name recognition among users. During the MultiMatch project, UvA developed a framework, called PodCred, that comprised indicators reflecting the credibility and quality of podcasts. The framework was published in a paper [Tsagias et al., 2008] at the WICOW 2008 workshop (http://www.dl.kuis.kyoto-u.ac.jp/wicow2) and was awarded the prize for the best paper of the workshop. In subsequent work, the indicators in the PodCred framework were mapped to features that can be automatically extracted from podcasts. Extensive experimentation was carried out with surface features and we demonstrated that it was possible to use the PodCred framework to predict user preference of podcasts (i.e., whether or not a podcast is popular). This work is reported in 4.4.2 "Semantic Analysis and Classification Component and Documentation for 2nd prototype."

Topical noise User generated content on the internet is unedited, and contains errors and other aberrations on the document level. However, podcasts and blogs consist not of one document, but of a series of documents. In the case of blogs, users prefer blogs in which the posts are focused on a single topic, rather than jumping from topic to topic. We consider bloggers who stray from the main topical thrust of their blog to be generating blogs with high levels of topical noise. In order to present users with better blog feed search results, UvA developed an approach that calculates a coherence score that reflects the topical focus of a blog. The coherence score was based on previous work at UvA in measuring topical structure in document collections, [2008a, 2008b] also carried out within the framework of the MultiMatch project. In He [2008c] we show that our proposed coherence score can be integrated into the language modelling framework and produced improved retrieval results for blog feed retrieval.

For work carried out within the MultiMatch project on image browsing and retrieval, please refer to Chapters 6 and 7 below.

References

Aihara, K., Yamada, K., Kando, N., Fujisawa, S., Uehara, Y., Baba, T., Nagata, S., Tojo, T, and Adachi, J. (2008) Supporting Creation and Sharing of Contents of Cultural Heritage Objects for Educational Purposes. 2nd International Workshop on Personalized Access to Cultural Heritage.

Amin, A., van Ossenbruggen, J., Hardman, L. and van Nispen, A. (2008). Understanding cultural heritage experts' information seeking needs. Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries.

Amato, G., Gennaro, C., Rabitti, F., and Savino, P. (2004). Milos: A multimedia content management system for digital library applications. In Research and Advanced Technology for Digital Libraries: 8th European Conference, ECDL 2004, pages 14--25. Springer Berlin/Heidelberg, 2004.

Barnard, K. and Forsyth, D. (2001), Learning the Semantics of Words and Pictures, International Conference on Computer Vision, volume 2, pages 408-415, 2001.

Besser, J. 2008. Incoporating User Search Goal analysis in Podcast Retrieval Optimization. Masters Thesis. Saarland University.

Besser J., Hofmann K., Larson M., "An Exploratory Study of User Goals and Strategies in Podcast Search", FGIR Workshop Information Retrieval (WIR2008), Würzburg, Germany, October, 2008.

Burkowski, F. J. (1992). Retrieval activities in a database consisting of heterogeneous collections of structured text. In Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '92), pages 112--125, New York, NY, USA, 1992. ACM Press.

Byrne, K. (2008a). Having Triplets - Holding Cultural Data as RDF. IACH2008, ECDL 2008 Workshop on Information Access to Cultural Heritage, Aarhus, Denmark.

Byrne, K. (2008b). Using RDF Graphs to Combine Text with Structured Fields for Better Retrieval from Hybrid Databases. Doctoral dissertation, University of Edinburgh, Scotland.

Byrne, K. (2008). Having Triplets - Holding Cultural Data as RDF. IACH2008, ECDL 2008 Workshop on Information Access to Cultural Heritage, Aarhus, Denmark.

Byrne, K. and Klein, K., (2003). Image Retrieval Using Natural Language and Content-Based Techniques. DIR 2003, 4th DutchBelgian Information Retrieval Workshop, Amsterdam.

Byrne, W. (2004) Automatic recognition of spontaneous speech for access to multilingual oral history archives. IEEE Transactions on Speech and Audio Processing, Vol 12, Issue: 4, pp. 420- 435.

Callan, J. P., Lu, Z. and Croft, W. B. (1995). Searching distributed collections with inference networks. In SIGIR '95: Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, pages 21--28. ACM Press, New York, 1995. Cheshire. Cheshire3 Information Retrieval Framework, 2006.

Carmichael, J., Larson, M., Marlow, J., Newman, E., Clough, P., Oomen, J., Sav, S., "Multimodal Indexing of Digital Audio-Visual Documents: a case study for Cultural Heritage Data", *Proceedings of the Sixth International Workshop on Content-Based Multimedia Indexing*, pp. 93 – 100, June, 2008.

Clarke, C. L. A., Cormack, G. V. and Burkowski, F. J. (1995). An algebra for structured text and a framework for its implementation. The Computer Journal, 38:43--56, 1995.

DCMI. Dublin Core Metadata Initiative, 2006. http://dublincore.org/.

DDC. Dewey decimal classification, 2006. http://www.oclc.org/dewey/.

de Jong, F.M.G. and Oard, D.W. and Heeren, W.F.L. and Ordelman, R.J.F. (2008) Access to recorded interviews: A research agenda. ACM Journal on Computing and Cultural Heritage, 1 (1). 3:1-3:27.

de Vries, A. P., Windhouwer, M. Apers, P. M. G. Kersten, M. (2000). Information access in multimedia databases based on feature models. New Generation Computing, 18:323--339, 2000.

Faloutsos, C. (1996). Searching Multimedia Databases by Content. Kluwer Academic Publishers, 1996.

Fuller M., Tsagkias M., Newman E., Besser J., Larson M., Jones G J F., de Rijke M., "Using Term Clouds to Represent Segment-Level Semantic Content of Podcasts", 2nd SIGIR Workshop on Searching Spontaneous Conversational Speech (SSCS 2008), Singapore, July, 2008.

Garofolo, J. S., Auzanne, C. G. P. and Voorhees, E. M. (2000). The TREC spoken document retrieval track: A success story. In Proceedings of RIAO 2000: Content-Based Multimedia Information Access, pages 1--20, 2000.

Goldman, J., Renals, S., Bird, S., de Jong, F., Federico, M., Fleischhauer, C., Kornbluh, M., and Lamel, L. Accessing the Spoken Word. International Journal of Digital Libraries 5:287-298.

Handschuh, S. and Staab, S. (2003). Annotation for the Semantic Web. IOS Press, Amsterdam, 2003.

Hauptmann, A. G. and Witbrock M. J. (1997). Informedia: news-on-demand multimedia information acquisition and retrieval. In Intelligent multimedia information retrieval, pages 215-239. MIT Press, Cambridge MA, 1997.

He J., Larson M., de Rijke M., (2008a) "Using Coherence-based Measures to Predict Query Difficulty", 30th European Conference on Information Retrieval (ECIR 2008): Springer, pp. 689–694, April, 2008.

He J., Larson M., de Rijke M., (2008b) "On the Topical Structure of the Relevance Feedback Set", FGIR Workshop Information Retrieval (WIR 2008), Würzburg, Germany, October, 2008.

He J., Weerkamp W., Larson M., de Rijke M., (2008c). "Blogger, Stick to your Story: Modeling Topical Noise in Blogs with Coherence Measures", SIGIR 2008 Workshop on Analytics for Noisy Unstructured Text Data (AND 2008), Singapore, July, 2008.

Hearst, M.A. (1997). Texttiling: segmenting text into multi-paragraph subtopic passages. Comput. Linguist., 23(1):33–64, 1997.

Hiemstra, D. and Michajlovic, V. (2005). A database approach to information retrieval: The remarkable relationship between language models and region models. Technical Report 05-35, Centre for Telematics and Information Technology, 2005.

Hollink, V., Kamps, J., Monz, C. and de Rijke, M. (2004). Monolingual document retrieval for European languages. Information Retrieval, 7:33--52, 2004.

Huijbregts, M., Ordelman, R. and de Jong, F. (2007). Annotation of heterogeneous multimedia content using automatic speech recognition. In Proceedings of SAMT, 2007.

Jelinek, F. (1997). Statistical methods for speech recognition. MIT Press, Cambridge MA, 1997.

Jeon, J., Lavrenko, V. and Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval, pages 119-126, ACM Press, New York, 2003.

Kapferer, T. Metadata Standards for the Sharing of Web Lectures. Ms. Thesis, University of Amsterdam.

Larson M., Newman E., Jones G., (2008a) "Classification of Dual Language Audio-Visual Content: Introduction to the VideoCLEF 2008 Pilot Benchmark Evaluation Task", 2nd SIGIR Workshop on Searching Spontaneous Conversational Speech (SSCS 2008), Singapore, July, 2008.

Larson M., Newman E., Jones G., (2008b) "Overview of VideoCLEF 2008: Automatic Generation of Topic-based Feeds for Dual Language Audio-Visual Content ", Working Notes for the CLEF 2008 Workshop, Aarhus, September, 2008.

Lesk, M. (2005). Understanding Digital Libraries. The Morgan Kaufmann series in multimedia information and systems. Morgan Kaufmann, San Francisco CA, second edition, 2005.

List, J., Mihajlovic, V., Ramirez, G., de Vries, A., Hiemstra, D., and Blok, H. E. (2005). TIJAH: Embracing IR methods in XML databases. Information Retrieval, 8:547--570, 2005.

Lucene. Open-source search software, 2006. http://lucene.apache.org/.

MonetDB. Open source high-performance database system, 2006. http://monetdb.cwi.nl/.

MuSeUM. Multiple-collection Searching Using Metadata, 2006. http://www.nwo.nl/catch/museum/.

Newman, E. and Jones, G. (2008). DCU at VideoCLEF 2008. Working Notes for the CLEF 2008 Workshop, Aarhus, September, 2008.

OWL. Web Ontology Language, 2006. http://www.w3.org/2004/OWL/.

Pecina, P., Hoffmannová, P., Jones, G., Zhang, Y. and Oard, D.W. (2008). Overview of the CLEF-2007 Cross Language Speech Retrieval Track. Advances in Multilingual and Multimodal Information Retrieval (CLEF 2007), vol. 5152: Springer, pp. 737-741, September, 2008.

RDF. Resource Description Framework, 2006. http://www.w3.org/RDF/.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A. and Jain, R. (2000). Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis Machine Intelligence, 22:1349--1380, 2000.

Smith, J. S. (2006). . IBM Research. 2006..

Snoek, C. G. M. and Worring, M. (2005). Multimodal video indexing: A review of the state-of-the-art. Multimedia Tools and Applications, 25:5--35, 2005.

STITCH. SemanTic Interoperability To access Cultural Heritage, 2006. http://www.nwo.nl/catch/stitch/.

UDC. Universal decimal classification, 2006. http://www.udcc.org/.

van der Sluijs K. and Houben, G-J. (2008) Tagging and the Semantic Web in Cultural Heritage ERCIM News 72 - Special: The Future Web, pp. 22-23.

Weibel, S. (1995). Metadata: The foundations of resource description. D-Lib Magazine, 1(7), 1995. http://www.dlib.org/dlib/july95/07/weibel.html.

6. Image Collections Overviews and Browsing

by Stephane Marchand-Maillet and Eric Bruno

This chapter describes the state of the art in the development of image collection browsing and overviewing. This is motivated by the fact that such activities are complementary to search operations and may provide efficient solutions where search tools are deficient due to the lack of representative semantics within the documents.

We also elaborate on why such browsing and overview facilities may provide interesting access means in a context such as that of the MultiMatch project.

6.1 Image Collection Browsing

Many current information management systems are centred on the notion of a *query*. This is true over the Web (with all classical Web Search Engines), and for Digital Libraries. In the domain of multimedia, available commercial applications propose rather simple management services whereas research prototypes are also looking at responding to queries (see Section 6.3 for details and examples).

The notion of browsing comes as a complement or as an alternative to query-based operations in several possible contexts that we detail in the following sections.

6.1.1 Browsing as extension of the query formulation mechanism

In the most general case, multimedia browsing is designed to supplement search operations. This comes from the fact that the multimedia querying systems largely demonstrate their capabilities using the Query-by-Example (QBE) scenario, which hardly corresponds to a usable scenario.

Multimedia search systems are mostly based on content similarity. Hence, to fulfil an information need, the user must express it with respect to relevant (positive) and non-relevant (negative) examples [Smeulders, 2000]. From there, some form of learning is performed, in order to retrieve the documents that are the most similar to the combination of relevant examples and dissimilar to the combination of non-relevant examples.

The question then arises of how to find the initial examples themselves. Researchers have therefore investigated new tools and protocols for the discovery of relevant bootstrapping examples. These tools often take the form of browsing interfaces whose aim is to help the user exploring the information space in order to locate the sought items.

The initial query step of most QBE-based systems consists in showing images in random sequential order over a 2D grid [Smeulders et al, 2000]. This follows the idea that a random sampling will be representative of the collection content and allow for choosing relevant examples. However, the chance for gathering sufficient relevant examples is low and must effort must be spent in guiding the system towards the relevant region of information space where the sought items may lie.

Similarity-based visualization ([Chen, 2000], [Cinque, 1998], [Leeuw, 2003], [Moghaddam, 2004], [Nazakato, 2001], [Nguyen and Worring, 2006], [Nguyen and Worring, 2008], [Rubner, 1999], [Vertan, 2002]) organizes images with respect to their perceived similarities. Similarity is mapped onto the notion of distance (Section 6.2.1) so that a dimension reduction technique (see Section 6.2.2) may generate a 2D or 3D space representation where images may be organized. Figure 6.1 illustrates the organization of 500 images based on colour information using the MDS dimension reduction [Rubner, 1999].

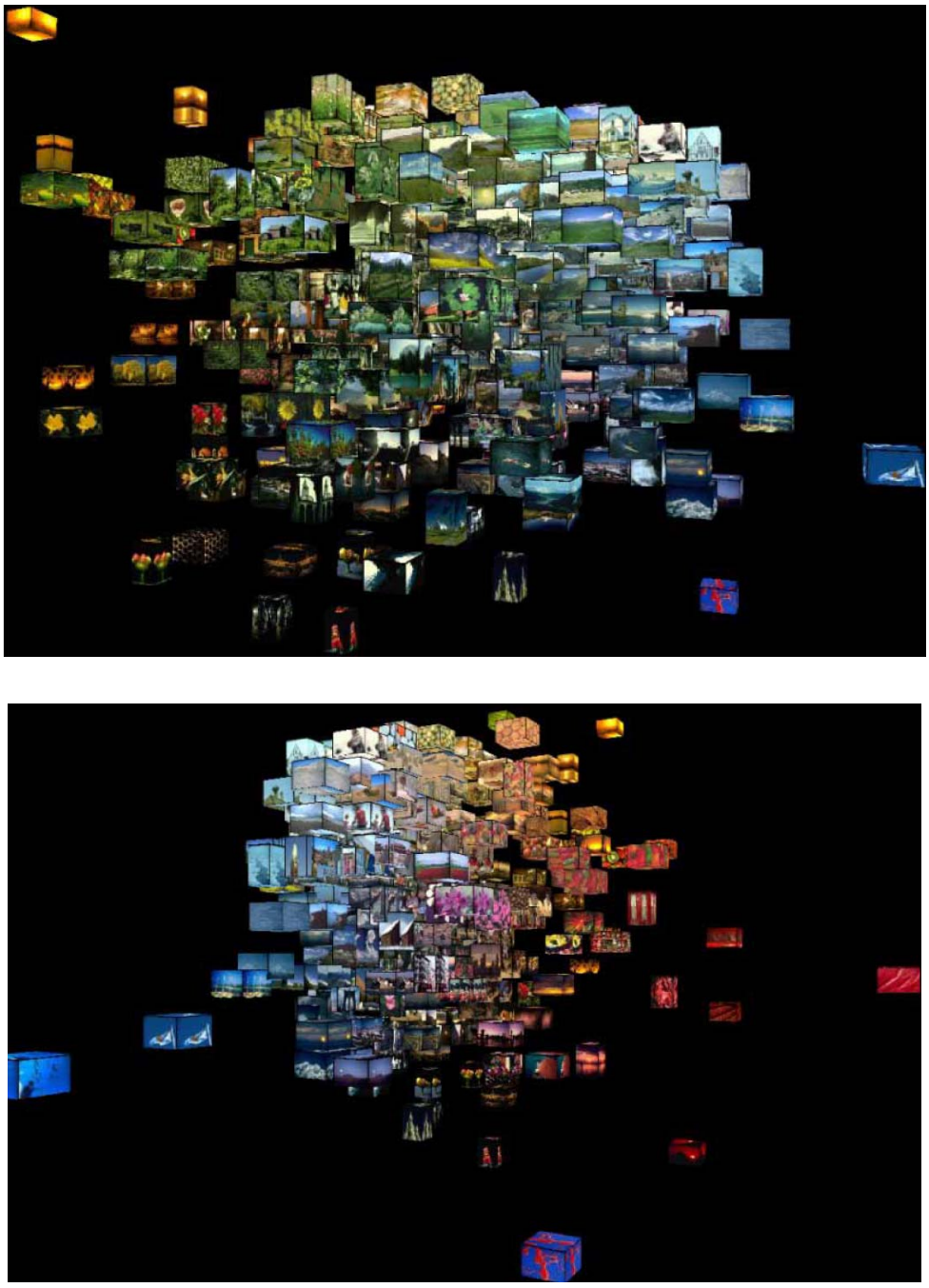


Figure 6.1: Two views of the MDS mapping of 500 images based on colour information

This type of display may be used to capture feedback by letting the user re-organise or validate the displayed images. Figure 6.2 shows a screenshot of the interface of El Niño [Santini et al., 2001].

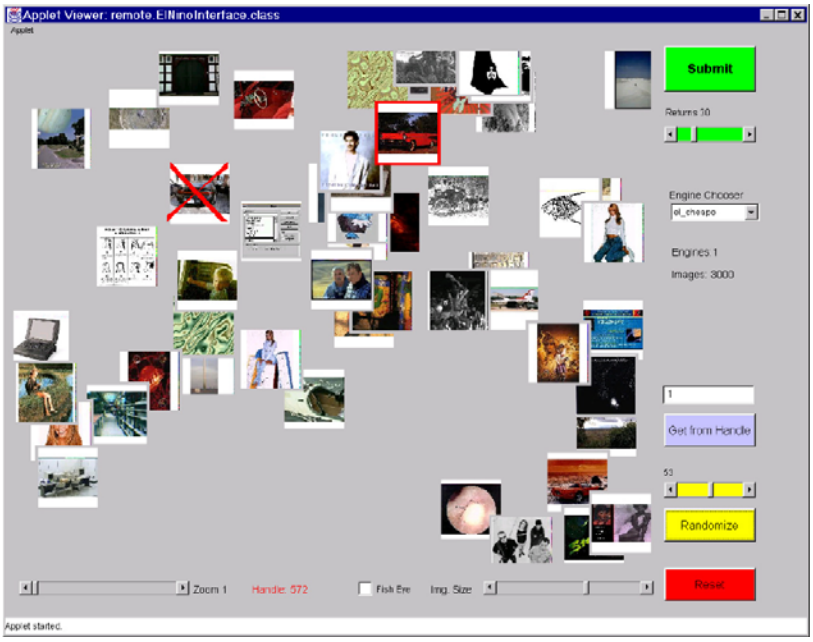


Figure 6.2: Interface of the El Niño system [Santini et al., 2001] where image similarity is mapped onto planar distance

Specific devices may be used to perform such operations. Figure 6.3 shows operators sitting around an interactive table for handling personal photo collections [Moghaddam, 2004].

Figure 6.3: The PDH table and its artistic rendering (from [Moghaddam, 2004])

In Figure 6.4, an operator is manipulating images in front of a large multi-touch display [PerceptivePixel, 2007].

Figure 6.4: Manipulating images over touch.-enabled devices (from [PerceptivePixel, 2007])

Alternative item organizations are also proposed such as the Ostensive Browsers (see Figure 6.5 and [Urban, 2005]) and interfaces associated to the NN^k paradigm [Heesch, 2004].

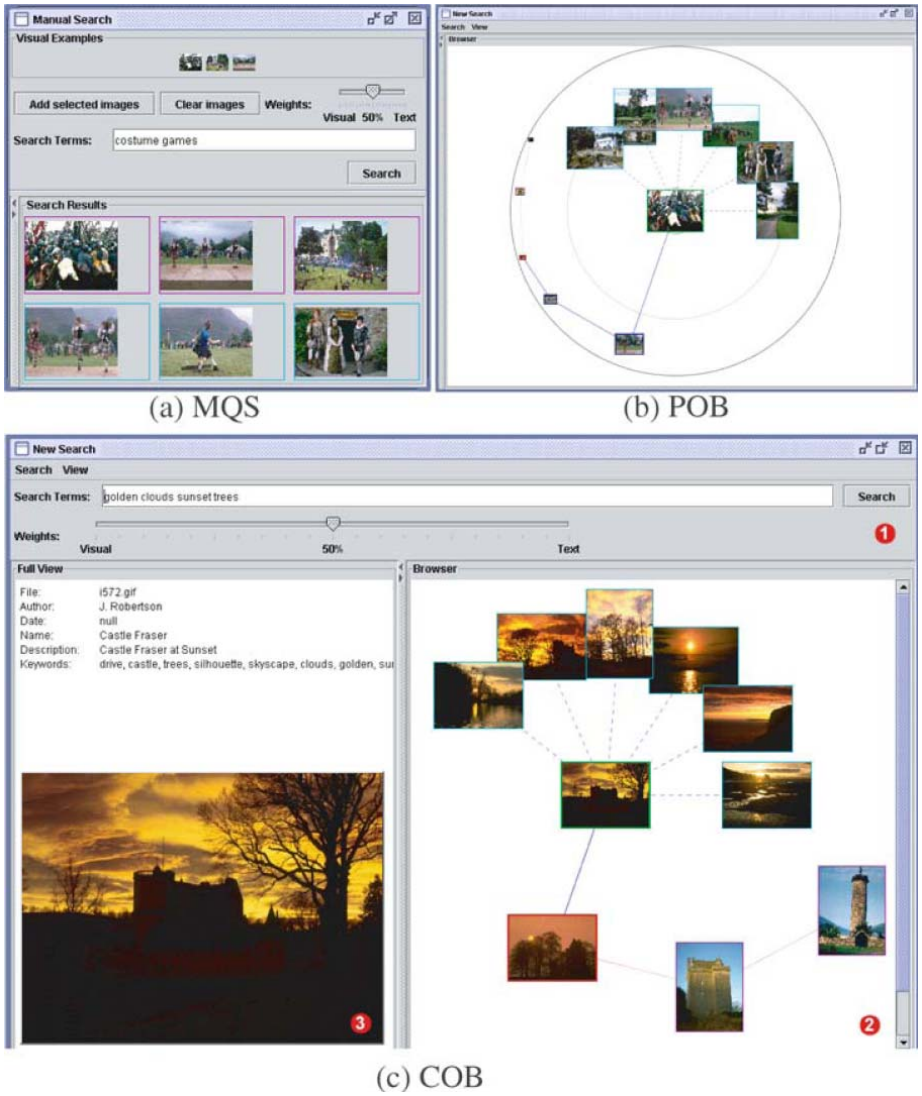


Figure 6.5: The Ostensive Browsers [Urban, 2005]

All these interfaces have in common the fact of placing multimedia retrieval much closer to human factors and therefore require specific evaluation procedures, as detailed in Section 6.4.

Although somewhat different, the development of the Target Search browsers is worth mentioning here. Whereas using QBE-based search a user may formulate a query of the type "show me everything that is similar to this (and not similar to that)" and thus characterize a *set of images*, using Target Search, the user is looking for a *specific image* (s)he knows is in the collection. By iteratively providing *relative* feedback on whether some of the current images are closer to the target than others, the user is guided to the target image. This departs from the QBE-based search where the feedback is absolute ("this image is similar to what I look for, whatever the context"). In that sense, Bayesian search tools may be considered as focused collection browsers.

In this category, the PicHunter Bayesian browser [Cox, 2000] is one of the initial developments. It has been enhanced with refocusing capabilities in [Müller, 1999] via the development of the Tracker system.

6.1.2 Browsing for the exploration of the content space

In the above cited works, browsing is seen as an alternative to the random picking of initial examples for the QBE paradigm. Here, we look at browsing from a different point of view.

In this setup, the user aims at overviewing the collection with no specific information need. Simply, (s)he wishes to acquire a representative view on the collection. In some respect, the above developments may be included into this category as overviews of the sub-collection representative to the query in question.

In [Kustanowitz, 2005], specific presentation layouts are proposed and evaluated (see also Section 6.4). The interface aims at enhancing the classical grid layout by organising related image groups around a central group (see Figure 6.6).

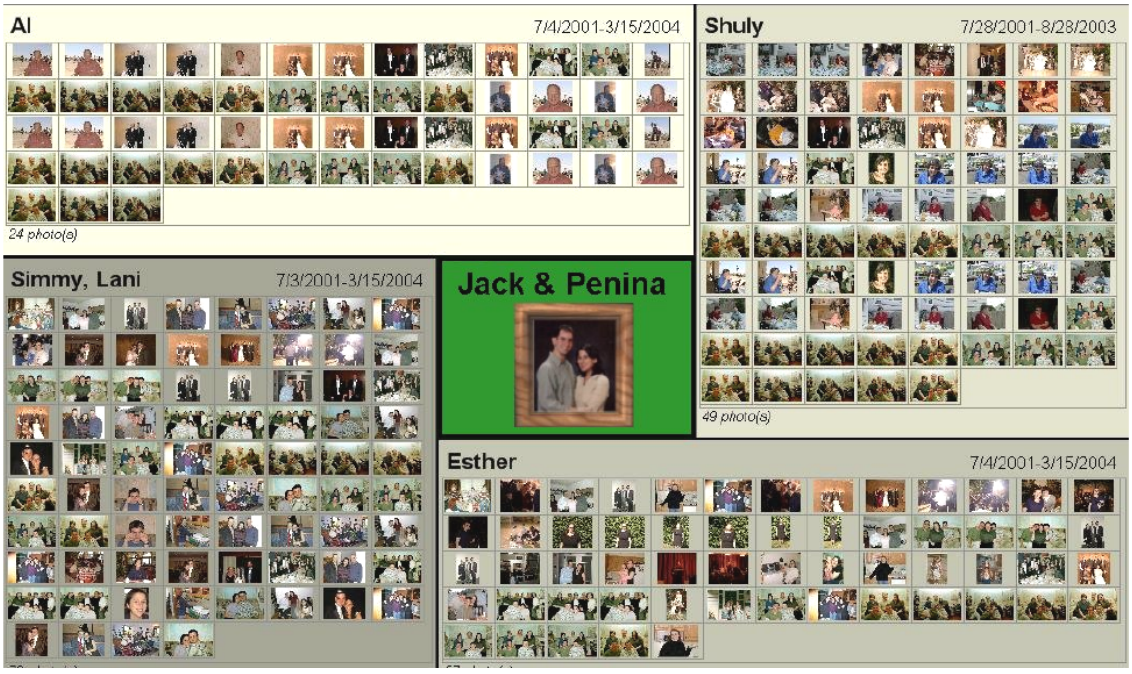


Figure 6.6: Bi-level radial layout [Kustanowitz, 2005].

Somewhat similar is the earlier development of PhotoMesa [Bederson, 2001] which aims at browsing image hierarchies using treemaps.

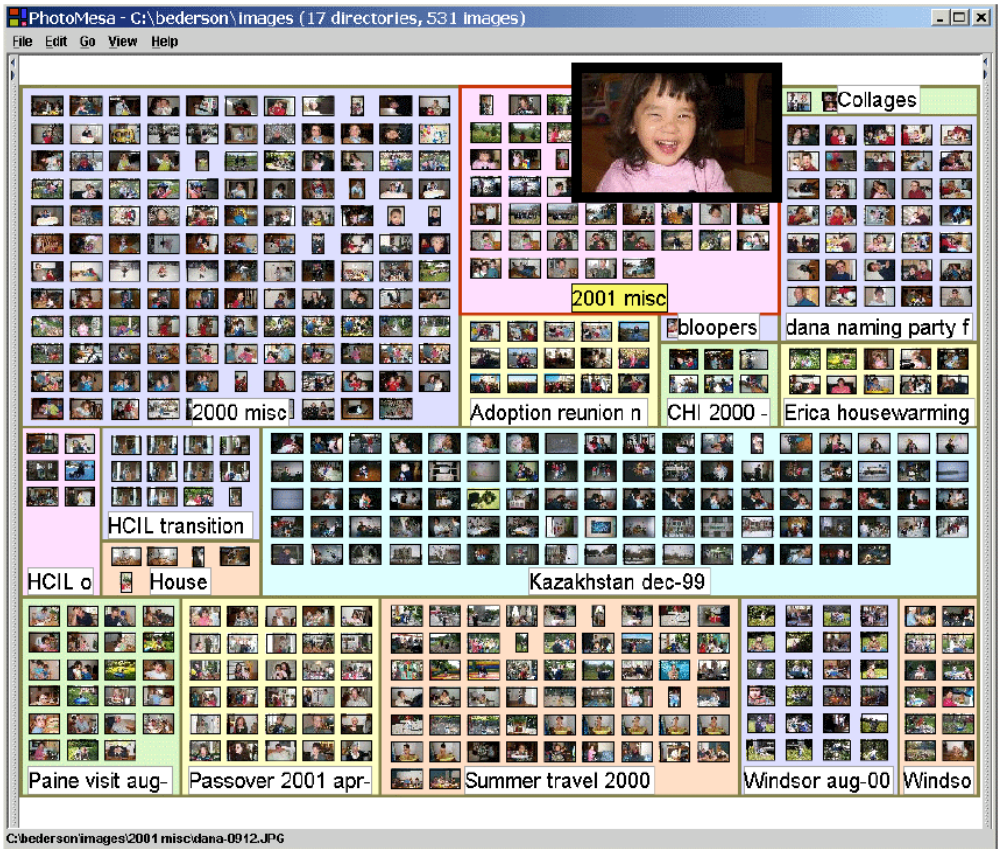


Figure 6.7: Screenshot of PhotoMesa, based on TreeMaps [Bederson, 2001]

Hierarchies are also studied in depth in the Muvis system, both for indexing and browsing via the Hierarchical Cluster Tree (HCT) structure [Kiranyaz, 2008]. In Figure 6.8, an example of hierarchical browsing of a relatively small image collection (1000 images) is shown.

In [Craver, 1999], the alternative idea of linearising the image collection is presented. The collection is spanned by two space-filling curves that allow for aligning the images along two intersecting 1D path. The reason for allowing two paths is that while two neighbouring points on a space-filling curve are neighbours in the original, the converse is not guaranteed to be true. Hence, two neighbouring points in the original space may end up far apart on the path. The use of two interweaved curves may alleviate this shortcoming.

At every image, each of the two paths may be followed in either of the two directions so that at every image, 4 directions are allowed. A browser shown in Figure 6.9 is proposed to materialize this navigation.

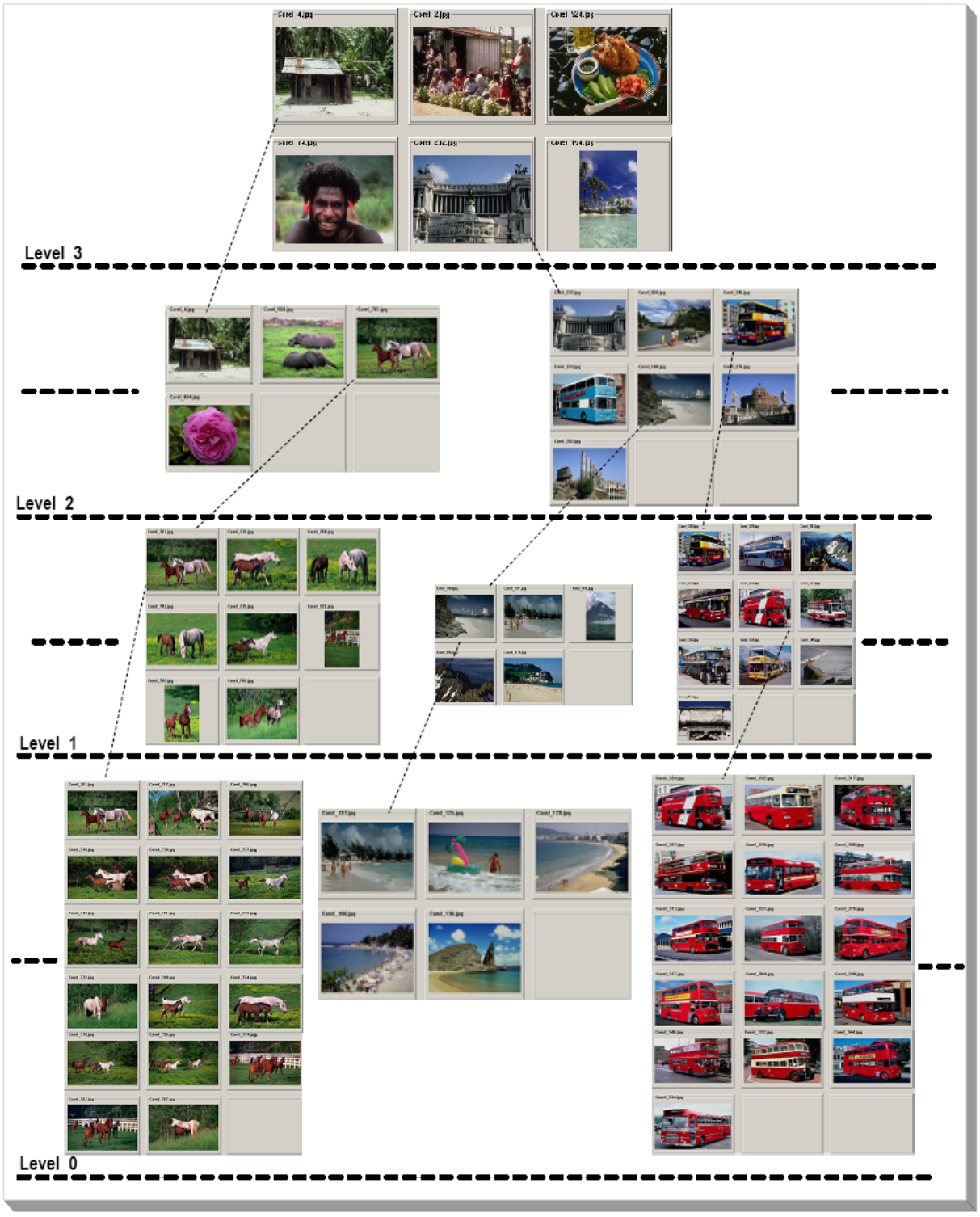


Figure 6.8: An example of HCT-based hierarchical navigation [Kiranyaz, 2008] on the 1k Corel image collection

Figure 6.9:. Multi-linerisation browser [Craver, 1999]

In [Marchand-Maillet, 2005], the principle of Collection Guiding is introduced. Given the collection of images, a path is created so as to "guide" the visit of the collection. For that purpose, image inter-similarity is computed and the path is created via a Travelling Salesman tour of the collection. The aim is to provide the user with an exploration strategy based on a minimal variation of content at every step. This implicitly provides a dimension reduction method from a high-dimensional feature space to a linear ordering. In turn, this allows for emulating sort operations on the collection, as illustrated in Figure 6.10.

Figure 6.10: Image sorting via the Collection Guide (left) random order (right) sorted list

The Collection Guide provides also several multi-dimensional arrangements (see Figure 6.11). However, it is clear that these (as the ones presented in the above section) are conditioned to the quality of the dimension

reduction strategy. In [Szekely, 2007], the underlying data cluster structure is accounted for so as to deploy valid dimension reduction operations (see Section 6.2 below for more details).

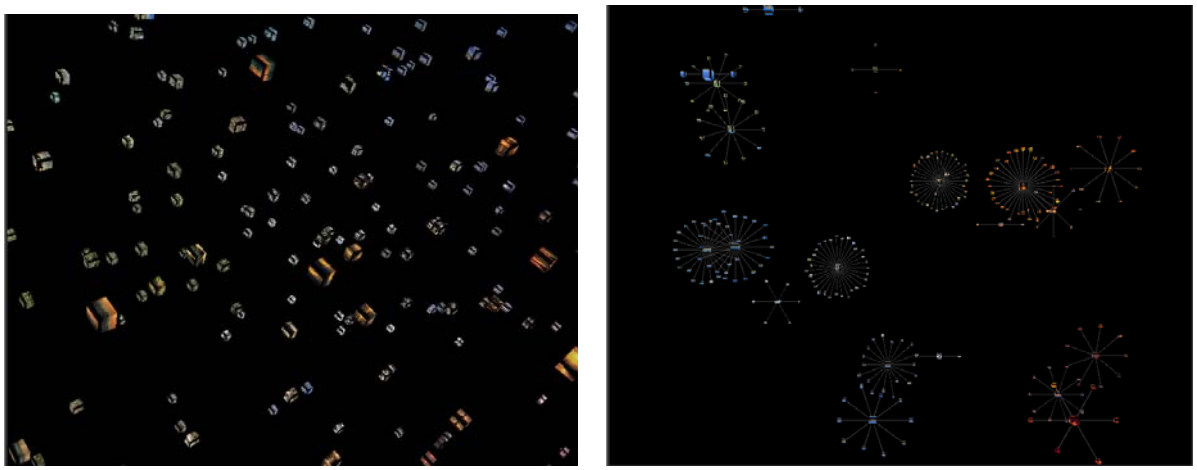


Figure 6.11: Examples of displays provided by the Collection Guide.
(left) generic 3D mapping (right) planet metaphor

Finally, at the border between exploration and search, *opportunistic search* is "characterised by uncertainty in user's initial information needs and subsequent modification of search queries to improve on the results" [Pu, 2003], [Janecek, 2003]. In [Pu, 2003], the authors present a visual interface using semantic fisheye views to allow the interaction over a collection of annotated images. Figure 6.12 displays interfaces associated with this concept.

Figure 6.12a: Displays associated with the opportunistic search mechanism
(from [Pu, 2003] and [Janecek, 2004])

Figure 6.12b: Displays associated with the opportunistic search mechanism (from [Pu, 2003] and [Janecek, 2004])

Faceted browsing [Hearst, 2006], oriented towards search is also at the limit between exploration and querying as it is also for filtering a collection while smoothly and interactively constructing complex queries. Figure 6.13 displays an example application of Faceted Search using the Flamenco toolbox for a collection of annotated images.

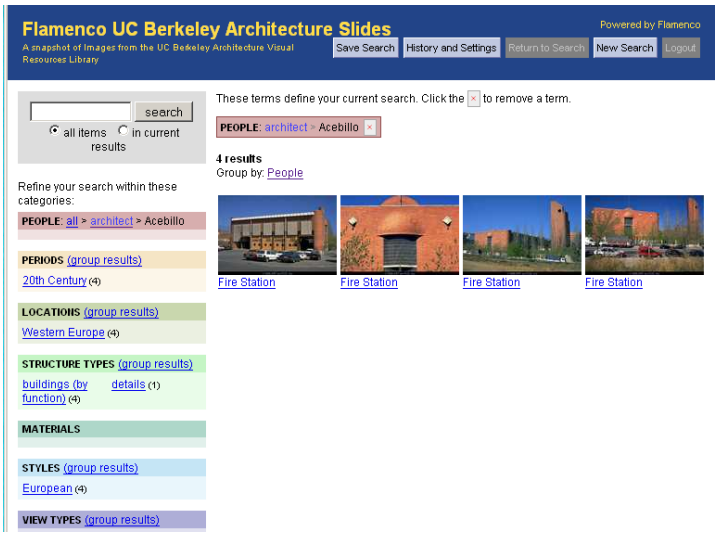


Figure 6.13: UC Berkeley Architecture Image Library (**Flamenco** toolbox)

6.1.3 Browsing to aid content description

While retrieval and browsing are in general passive to the collection (*i.e.* the collection stays as it is), these operations may also be used to enrich the collection content. In [Kosinov, 2003], authors have reviewed and proposed several models that allow for the semantic augmentation of multimedia collections via interacting with them. This follows the line of the Semantic Web and associated domains of knowledge management. In this line, the work proposed in [Schreiber, 2001] relates ontology management and image description.

6.2 Multimedia Space Representation

From a multimedia (image, in our case) collection, one should derive a representation that is both easy to handle via mathematical tools but which also account for the intrinsic meaning (semantics) of the content. From there, operations such as sampling and visualization are made possible. We overview briefly the possibilities in the next sections.

6.2.1 Generic feature space representation

There are well-known image representation techniques in the image compression and retrieval literature [Smeulders, 2000]. Among them, features such as colour, texture and shape emerge as the most global dominant cues for image content characterization.

The task of feature selection is typically associated with data mining. In our context, one may perform feature selection base on several criteria. Typical reported work is based on informative measures associated with predefined features or aims at optimizing a given criterion by the design of abstract feature sets.

Item similarity measurement

Distance measurement depends on the space within which information is immersed. In the case of colour for example, it is known that distance measurements within the RGB colour cube do not correspond to any perceptual similarity. To this end, the HSV, Luv and CIELa*b* colour spaces have been proposed within which simple Euclidean measurement correspond to perceptual distances.

A variety of distance functions exist and may be used for characterizing item proximity [Duda, 2000]. The simplest distance functions that may be used are those derived from the Minkowsky distance (L_k norm) formula. Here, all coordinates are taken equally, meaning that we assume the fact of an isotropic space. If we assume that coordinates are realizations of a random variable with a known covariance matrix, then the Mahalanobis distance may be used. More sophisticated distance functions exist, such as the Earth Mover's Distance [Rubner, 1999].

Collection subsampling

Associated with the concepts of exploration and browsing is the concept of summarization. Summarization is an approach commonly taken for presenting large content and involves a clear understanding of the collection diversity for performing sampling.

The most common way of performing sampling is to use the underlying statistics of the collection. Typically, within the feature space, local density is analyzed. Dense regions of this space will be represented by several items whereas sparse regions will mostly be ignored within the representation. More formally, strategies such as Vector Quantization (VQ) may be used to split the space into cells and only consider cell representatives. *k*-means clustering is one of the most popular VQ techniques.

A geometrical interpretation of VQ is that of defining a Voronoi [Dirichlet] tessellation of the feature space such that each cell contains a cluster of data points and each centroid is the seed of the corresponding cell. This tessellation is optimal in terms of minimizing some given cost function, embedding the assumption over the properties of the similarity measurement function in the image representation space.

A radically different approach is to perform hierarchical clustering on the data. Initial data points form the leaves of a tree called dendrogram. The tree is built upon dependence relationships between data points. In the single-link algorithm, a point is agglomerated with its nearest neighbour, forming a new data point and a node within the tree. The algorithm stops when all points are gathered. Alternatives (complete-link and average-link) preserve the internal structure of clusters when merging.

The dendrogram obtained may then be the base for sampling the collection, as each level of the dendrogram shows a view of the collection. By defining collection samples as closest to the tree nodes at one given level, one obtains an incremental description of the collection.

6.2.2 Dimension reduction

So far, we have considered items as represented by vectors in the feature space. However, two aspects of this mathematical modeling should be inspected. First, we have defined distances and similarity measures irrespectively of the feature space dimensionality. However, it is known that this dimensionality has an impact on the meaningfulness of the distances defined [Aggarwal, 2001]. This is known as the *curse of dimensionality* and several results can be proven that show that there is a need for avoiding high-dimensional spaces, where possible.

Further, typical visualization interfaces cannot handle more than 3 dimensions. Hence, there is a need for consistently representing items immersed in a high-dimensional space in lower dimensional spaces, while preserving neighbouring properties. Dimension reduction techniques come as a solution to that problem. Methods for dimensionality reduction are employed each time high-dimensional data has to be reduced from a high to a low-dimensional space. The principle of the mapping process for methods based on distance matrices is to find the configuration of points that best preserves the original inter distances.

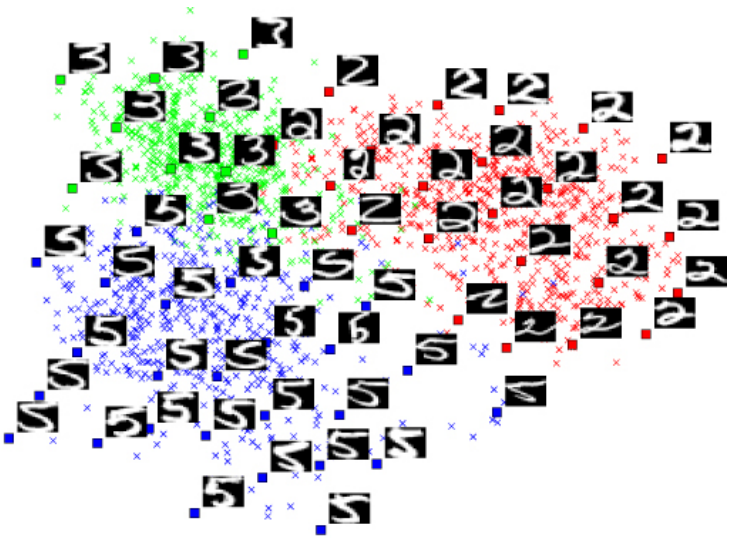


Figure 6.14: Dimension reduction over a database of digit images
(Illustration from http://www.merl.com/projects/dimred)

A number of methods exist. We do not detail the list and principles here but refer the reader to [Szekely, 2007], [Borg, 2005] and [Carreira-Perpiñan, 1997] for thorough reviews on the topic.

6.3 Multimedia Collection Browsers

6.3.1 Extra image browsers

In the above pages, we have reviewed a number of strategies for image collection browsing. We list here other known browsers:

- **Microsoft**'s picture manager (filmstrip mode) is the simplest representation that can be created. It exploits a linear organization of the data. In the context of its usage, linearization is made on simple metadata, which lends itself to the ordering (e.g. temporal or alphabetical order)

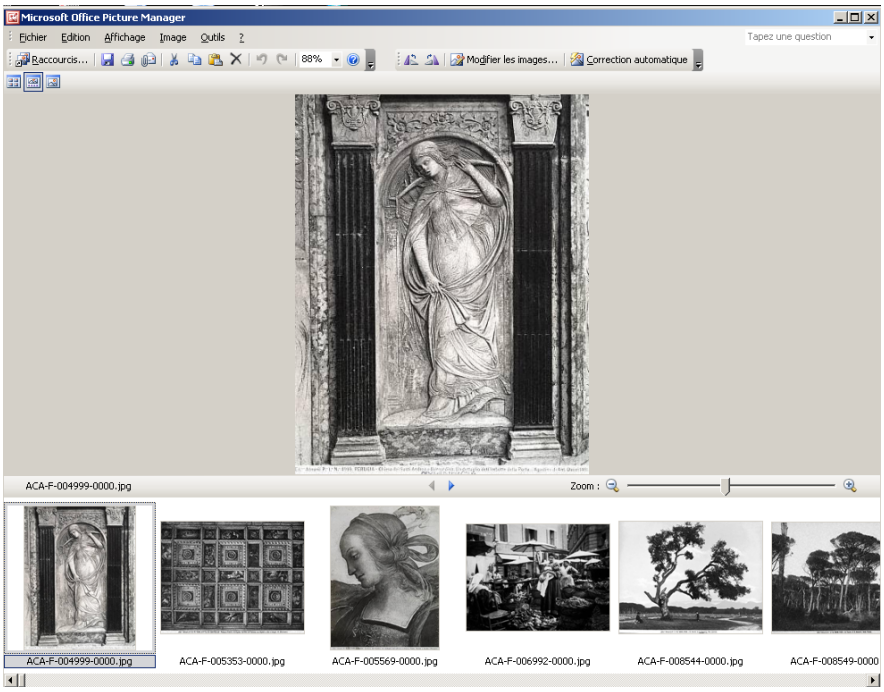


Figure 6.15: Microsoft Picture Manager

– Google's **Picasa** (timeline mode) also exploits the linear timeline to arrange a photo collection. An interesting feature is the near-1D organization whereby groups of pictures are arranged along the path (as opposed to aligning single pictures).

Figure 6.16 Google's Picasa

– **Flickr**'s geotagged image browser exploits the planar nature of geographic data to arrange pictures.

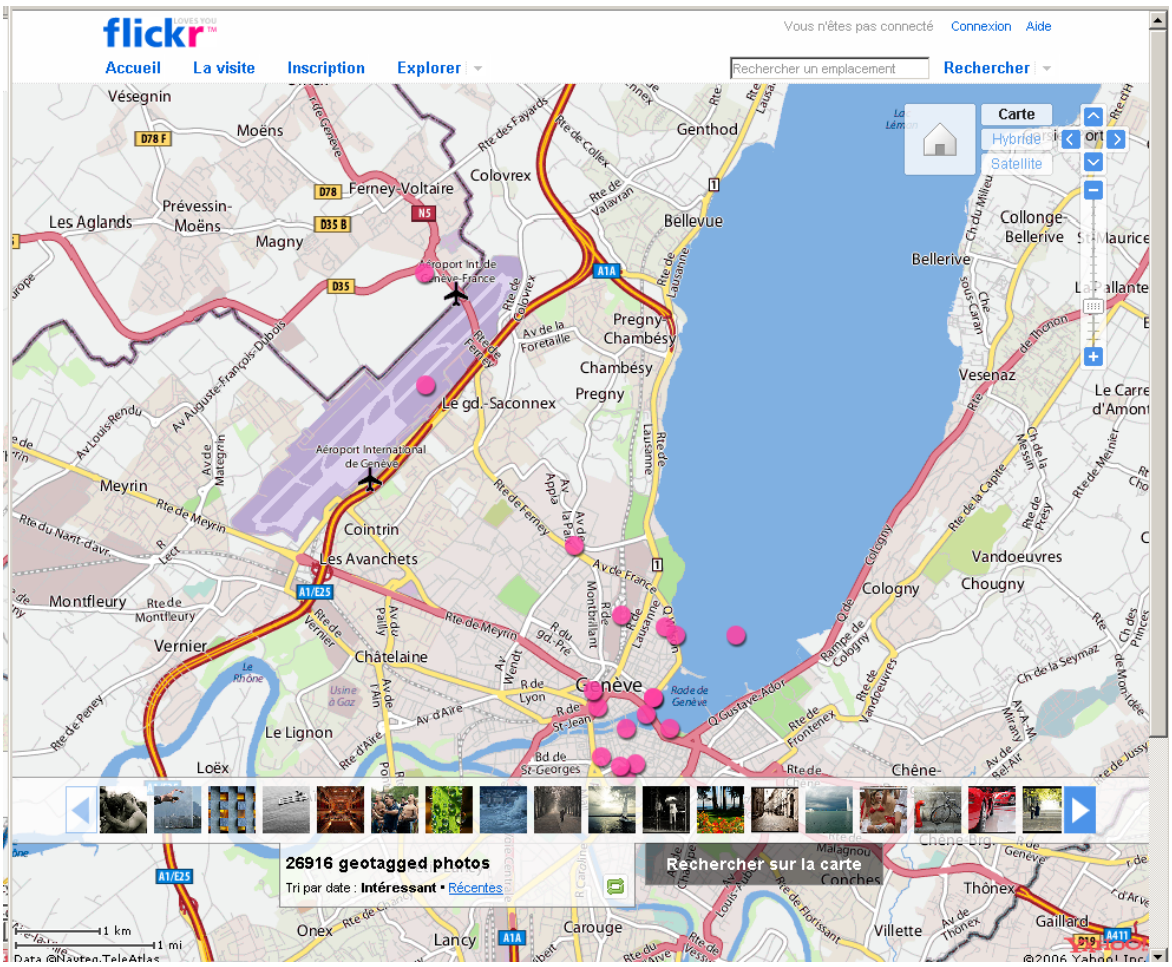


Figure 6.17 Flickr's Geotagged Image Browser

6.3.2 Related patents

Image browsing is of high commercial interest since it provides a added value over a collection of data. The following are some US patents related to image browsing.

6233367 Multi-linearization data structure for image browsing

6636847 Exhaustive search system and method using space-filling curves

6907141 Image data sorting device and image data sorting method

7003518 Multimedia searching method using histogram

7016553 Linearized data structure ordering images based on their attributes

7131059 Scalably presenting a collection of media objects

7149755 Presenting a collection of media objects

6.3.3 Other media

Browsing may clearly apply to media other than images. This section complements examples already provided in detail in MultiMatch Deliverable 1.1.1 (Section 7). Hence, while not detailing underlying strategies we give examples and pointers to multimedia browsers that we think provide interesting browsing functionalities.

ViCoDE (Video Collection Description and Exploration – [Bruno, 2008]) is a video search engine interface implementing the QBE paradigm and allowing some exploration functionalities.

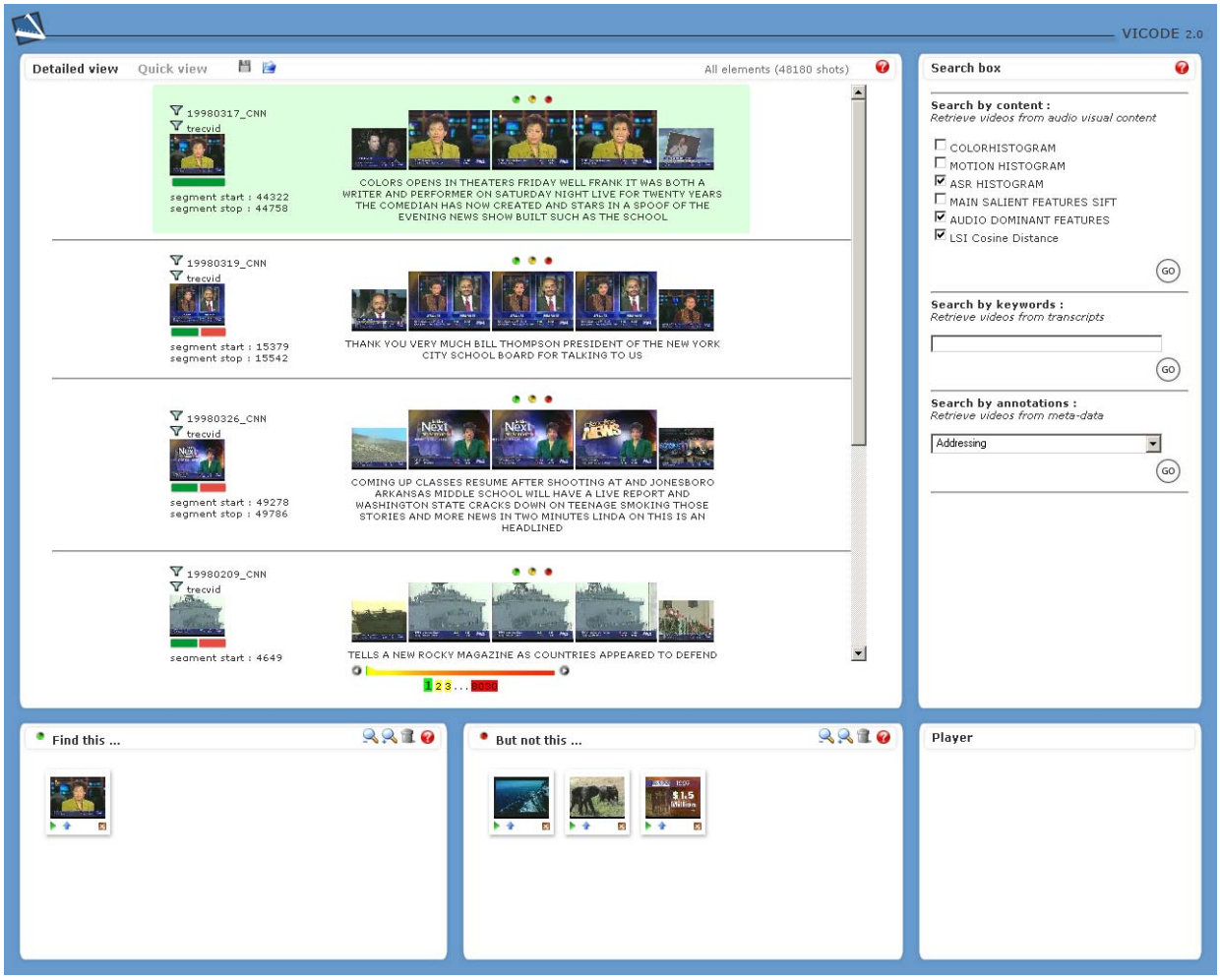


Figure 6.18 ViCoDE - Video Collection Description and Exploration

The **MediaMill** browsers [Worring, 2007] allow time and similarity based video exploration. They have been tailored to the TRECVid challenge (interactive task) and thus are relevant for news content exploration.

Figure 6.19: MediaMill Browsers

Islands of music [Pampalk, 2003] use Self Organising maps to arrange music pieces into a planar landscape, then used for browsing.

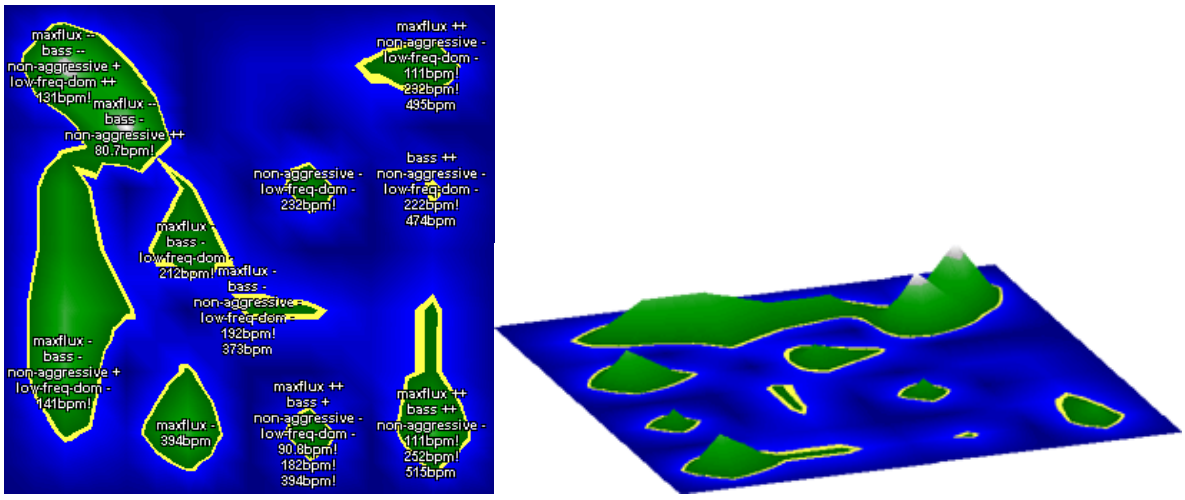
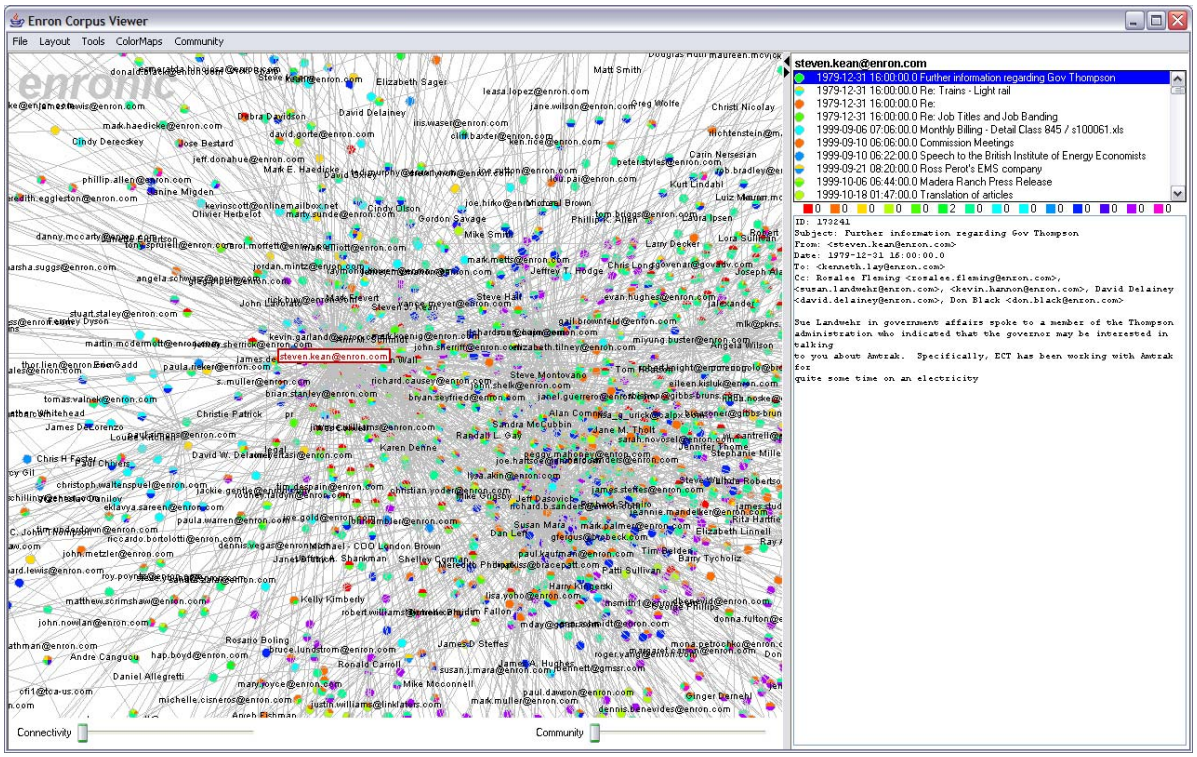


Figure 6.20: Islands of Music

Enronic [Heer, 2004] Email collection browsing, investigation. As emails represent communications between humans, this work is related to the domain of Social Network Analysis.



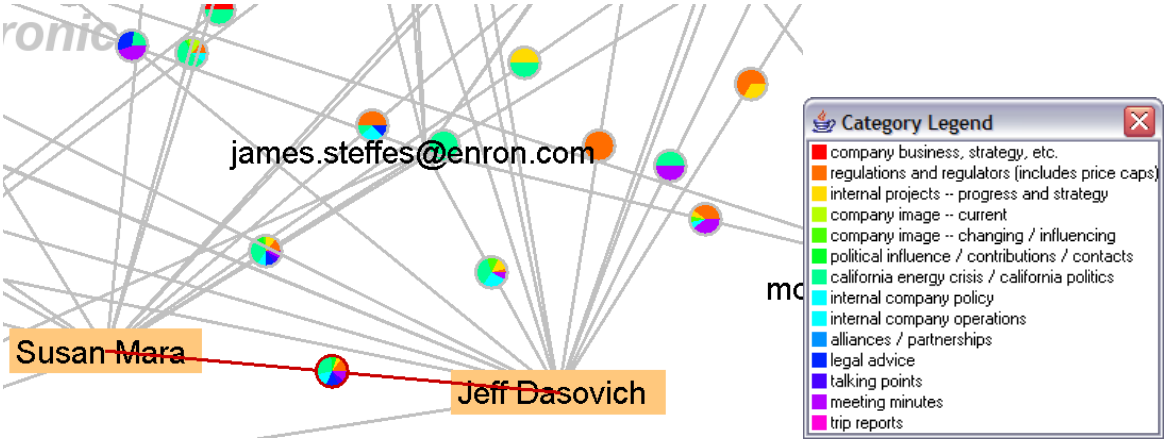


Figure 6.21: Enronic

Scatter/Gather [Hearst, 1996] is an early work on clustering retrieval results for their exploration by categories.

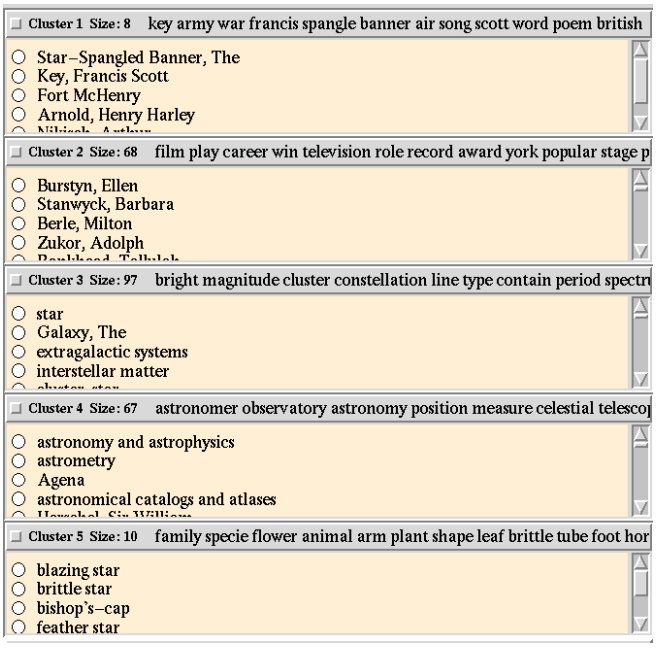


Figure 6.22: Scatter / Gather

6.4 Evaluation

In [Chen, 2002], it is analyzed how browsing and the more general fact of providing an efficient interface to an information systems is often listed as one of "Top-ten" problems in several fields (e.g., Information Retrieval [Croft, 1995], visualization and virtual reality). A new top-ten list of problems in the domain is created including benchmarking and evaluation.

Firstly, the majority of browsing tools proposed in the literature organize their content using low-level features such as colour or texture. [Rorissa, 2007] demonstrates via several user studies that this is relevant and that features may indeed be used as a basis for visualization and hence browsing.

There are numerous efforts to benchmark information retrieval as a problem with a well-posed formulation. When including the quality of the interface or performance of the interaction with the information system, things are however less clear. The fact of embarking human factors in the context make the formulation less definite and prevents the automation of the performance measures (see e.g. [Ivory, 2001]).

Several attempts to propose evaluation protocols and frameworks have nevertheless taken place ([Black, 2002], [Rodden, 1999], [Urban 2006b]). Some particular aspects such as zooming [Combs, 1999] and presentation ([Kustanowitz, 2005], [Rodden, 2001]) have been the focus of attention for some works.

While systematic retrieval performance evaluation is possible using ground truth and measures such as Precision and Recall, having reliable performance evaluation of interfaces and interactive tools requires long-term efforts and heavy protocols. It is certainly an area where developments should be made to formally validate findings. It is often a strong asset of private companies which carefully invest in user-based testing in order to validate tools that are simpler but more robust than most research prototypes.

6.5 MultiMatch Information Browser

In the course of the project, we have built on the analysis made in this state of the art and constructed an enhanced information browser as complement to the main search interface. We have followed the idea that searching and relevance feedback help in exploring local portions of the information space, whereas browsing should help the user to obtain both a global overview of the information space at hand and provide the user with a clear and efficient browsing strategy.

We have therefore mixed the idea of the Collection Guide with that of linearization and faceted browsing to obtain an information browser starting from a specific document and linking, out of several possible

dimensions, to other documents close to that dimension. By clicking on any of the non-central documents would bring it at the central place with its associated context.

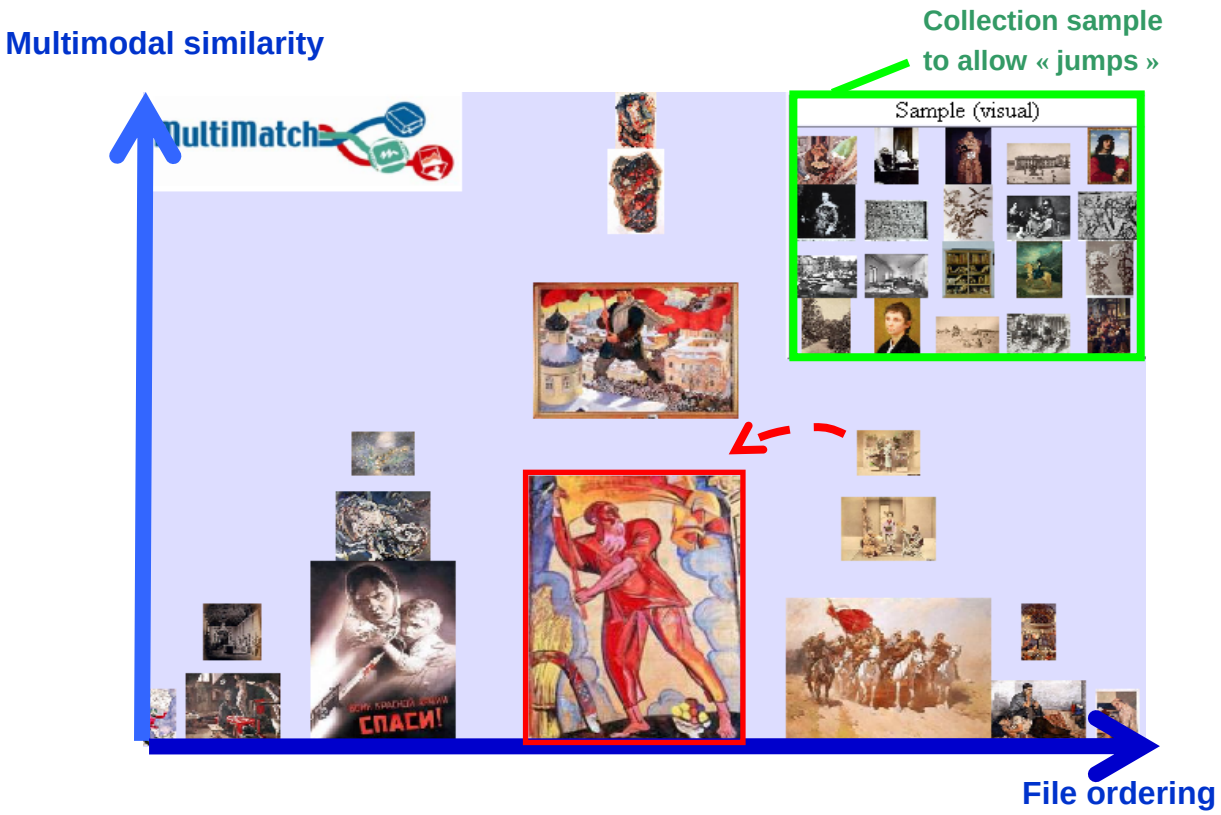


Figure 6.23: MultiMatch Information Browser

Ordering set over horizontal and vertical dimensions may be modified and adapted at will. They may come from natural ordering (e.g. timeline over creation dates, alphabetical ordering of creator's name, piece title) or be created using the Collection Guide methodology [Marchand-Maillet, 2005] out of content or metadata features (e.g. multimodal similarity).

6.6 Concluding Remarks

Image browsing comes as a complement to query-based search. This is valuable, due to the imperfect nature of content understanding and representation, due principally to the so-called semantic gap. Browsing is also interesting to resolve the problem of the user's uncertainty in formulating an information need. Opportunistic search and faceted browsing are example of principles and applications that bridge search and navigation.

The above analysis shows that, as a complement to classical retrieval systems, browsing and navigation should be differentiated. It is suggested here that *browsing* is directed towards an objective (information need) and thus indirectly relates to searching and acts at the *document scale*. As such, browsing is seen as assistance within similarity-based search systems, where the QBE paradigm is often deficient.

Browsing should be differentiated from *navigating* where the aim is the understanding of the collection content. Navigation-based systems thus use an absolute (global) modeling of the collection and include a global notion of similarity (i.e. that is driven by generic feature). This is to be opposed to browsing systems, which use a notion of similarity based on the context of the neighbourhood of the sought items (i.e., the interpretation of the collection is made at the light of the sought items).

Image collection browsing imposes focus on user interaction and thus the interface design and evaluation. This refers to the work done by the Human Factors (HCI) community, which is somewhat regrettably not sufficiently inter-weaved with the Information Retrieval and Management community.

Finally, while browsing and navigation may be seen as an extension and complement to searching in image collections, it can also be applied to other media such as audio (music, e.g. [Pampalk, 2003]) and video (e.g. [Worring, 2007], [Ciocca, 2007]). These temporal media offer a temporal dimension that directly lends itself to exploration and thus makes browsing an obvious tool to use.

References and Relevant Literature

Aggarwal C.C. et al (2001). On the surprising behaviour of distance metrics in high dimensional space. In Proceedings of the 8th International Conference on Database Theory (ICDT'2001), 2001.

Ahlberg C. and Schneiderman B. (1994). Visual Information seeking: tight coupling of dynamic query filters with starfield displays. In CHI'94, pp 313-317, ACM Press, 1994.

Bederson D. (2001). PhotoMesa: A zoomable image browser using quantum treemaps and bubblemaps. In ACM Symposium on User Interface Software and Technology, CHI Letters, volume 3, pp 71-80. 2001.

Black JA et al. (2002) A method for evaluating the performance of content-based image retrieval systems based on subjectively determined similarity between images. In: Proc. Int. Conf. on Image and Video Retrieval, LNCS 2383, pp 356–366.

Borg I and P.J.F. Groenen. (2005). Modern Multidimensional Scaling: Theory and Application. Springer, 3rd Edition, 2005.

Börner K. (2000) Extracting and visualizing semantic structures in retrieval results for browsing. ACM Digital Libraries, San Antonio, Texas, June 2-7, pp. 234-235

Bruno E et al.(2008). Design of multimodal dissimilarity spaces for retrieval of multimedia documents. *To appear in IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2008.

Carey M. et al. (2003) Info navigator: A visualization tool for document searching and browsing. In Proceedings of International Conference on Distributed Multimedia Systems, September 2003.

Carpineto C. and Romano G (1996). A lattice conceptual clustering system and its application to browsing retrieval. Machine Learning, 24:95-122, 1996.

Carreira-Perpiñan M. (1997). A review of Dimension Reduction Techniques. Tech Report number CS-96-09. University of Sheffield, 1997.

Chang M.et al. (2004). Collection understanding. 4th ACM/IEEE-CS Joint Conference on Digital Libraries JCDL '04. ACM Press, New York, NY, 334-342, 2004

Chen C. et al. (2000). Content-based image visualization. In IV'00: Proceedings of the International Conference on Information Visualization. Pp 13-18. IEEE Computer Society.

Chen C. and Börner K. (2002). Top ten problems in Visual Interfaces of Digital Libraries. In Visual Interfaces to Digital Libraries, K Börner & C. Chen Eds, Springer Verlag, LNCS 2539, 2002.

Chen J.Y. et al. (2000). Hierarchical browsing and search of large image databases. IEEE Transactions on Image Processing. 9(3):442-455, 2000.

Cinque L. et al. (1998). A multidimensional image browser. Journal of Visual Language and Computing. 9(1):103-117, 1998.

Ciocca G. and Schettini R. (2007). Hierarchical Browsing of Video Key Frames. ECIR 2007: 691-694.

Combs T.T.A. and Bederson B.B. (1999). Does Zooming Improve Image Browsing? Proceedings of Digital Library (DL 99) New York: ACM, 130-137.

Cox I.J. (2000). PicHunter: Theory, implementation, and psychophysical experiments. IEEE Transactions on Image Processing, 9(1):20-37, 2000.

Craver S. et al. (1999). Multi-linearisation data structure for image browsing. In SPIE Conference on Storage and Retrieval for Image and Video Databases VII, 155-166, January 1999. [US Patent 6233367]

Croft, W.B. (1995), What do people want from information retrieval? D-Lib Magazine, November 1995.

Demartines P. and Herault J. (1997). Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. IEEE Transaction on Neural Networks, 8(1):148-154, 1997.

Duda R. et al (2000). Pattern Classification. Wiley-Interscience; Second Edition.

Hearst M and Pedersen J (1996), Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, *Proceedings of the 19th Annual International ACM/SIGIR Conference*, Zurich, August 1996

Hearst M. (2006), Design Recommendations for Hierarchical Faceted Search Interfaces. ACM SIGIR Workshop on Faceted Search, August, 2006.

Heer J (2004). Exploring Enron: Visualizing ANLP Results. In *Applied Natural Language Processing. InfoSys 290-2. University of California, Berkeley.*

Heesch D. and Ruger S. (2004). Three interfaces for content-based access to image collections. In Proceedings of the International Conference on Image and Video Retrieval (CIVR 04). LNCS volume 3115. pp 491-499. Springer Verlag, 2004.

Hinton G. and Roweis S. (2002). Stochastic Neighbor Embedding. Neural Information Processing Systems. 15:8333-840, 2002.

Ivory Y. and Hearst M. (2001). The state of the art in automating usability evaluation of user interfaces. ACM Computing Surveys, 33(4):470-516, 2001.

Janecek P. and Pu P (2003). Searching with semantics: An interactive visualization technique for exploring an annotated image collection. In Workshop on Human Computer Interfaces for Semantic Web and Web Applications (HCI-SWWA '03), November 2003.

Janecek P and Pu P. (2004). Opportunistic Search with Semantic Fisheye Views EPFL Technical Report: IC/2004/42.

Kang H. and Shneiderman B (2000). Visualization methods for personal photo collections: Browsing and searching in the PhotoFinder. *IEEE International Conference on Multimedia and Expo (III) 2000*, pages 1539–1542, 2000.

Keim D.A (2002). Information Visualization and visual data mining. IEEE Transactions on Visualisation and Computer Graphics, 7(1):100-107, 2002.

Kiranyaz M and Gabbouj M (2008). Hierarchical Cellular Tree: An Efficient Indexing Scheme for Content-based Retrieval on Multimedia Databases, IEEE Transactions on Multimedia (in Print)

Kustanowitz J. and Shneiderman B. (2005). Meaningful presentations of photo libraries: rationale and applications of bi-level radial quantum layouts. *ACM/IEEE Joint Conference on Digital Libraries*, pages 188–196, 2005

Kosinov S. and Marchand-Maillet S (2003). Overview of approaches to semantic augmentation of multimedia databases for efficient access and content retrieval. In 1st International Workshop on Adaptive Multimedia Retrieval (AMR2003), 2003.

Laaksonen J. et al (2000). PicSOM- Content-based image retrieval with self-organising maps. Pattern recognition Letters, 21(13-14):1199-1207, 2000.

Leeuw W. and Liere R. (2003). Visualization of multidimensional data using structure preserving projection methods. In Data Visualization: The State of the Art. FH Post and GM Nielson and GP Bonneau eds, pp 213-224. Kluwer 2003.

Le Saux B. and Boujemaa N. (2002) Unsupervised Robust Clustering for Image Database Categorization, IEEE-IAPR International Conference on Pattern Recognition (ICPR'2002), Quebec, Canada, August 2002.

Li J.X. (2004). Visualization of high-dimensional data with relational perspective map. Information Visualization, 3:49-59, 2004.

Marchand-Maillet S. and Bruno E (2005). Collection Guiding: A new framework for handling large multimedia collections. Seventh International Workshop on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib'05), Cortona, Italy.

Moghaddam B et al. (2004). Visualization and user-modeling for browsing personal photo libraries. International Journal of Computer Vision, 56(1-2):109-130, 2004.

Müller W. et al (1999). Hunting moving targets: an extension to Bayesian methods in multimedia databases, In *Multimedia Storage and Archiving Systems IV (VV02)*, Vol. 3846 of SPIE Proceedings, Boston, Massachusetts, USA, 20-22 September 1999. (SPIE Symposium on Voice, Video and Data Communications)

Nakazto M. and Huang T.S. (2001). 3D MARS: Immersive virtual reality for content-based image retrieval. In IEEE ICME'01. IEEE Computer Society, 2001.

Nakazato, M. et al. (2003) ImageGrouper: A group-oriented user interface for content-based image retrieval and digital image arrangement. Journal of Visual Languages and Computing 14, 363–386 (2003)

Nguyen GP and Worring M. (2005). Scenario optimization for interactive category search. In ACM-MIR'05. ACM Press, 2005.

Nguyen GP and Worring M. (2006). Interactive access to large image collections using similarity-based visualization. Journal of Visual Languages and Computing (to appear).

Nguyen GP. and Worring M. (2008), *Optimization of interactive visual similarity based search*, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP). Vol. 4, Issue 1, February 2008.

Pampalk. E (2003). Islands of Music - Analysis, Organization, and Visualization of Music Archives. Journal of the Austrian Society for Artificial Intelligence, Vol. 22, No. 4, pp 20-23.

PerceptivePixel, (2007). Introduction video at http://www.perceptivepixel.com

Pu P. and Janecek P. (2003). Visual interfaces for opportunistic information seeking. In C. Stephanidis and J. Jacko, editors, 10th International Conference on Human -Computer Interaction (HCII '03), pages 1131{1135, Crete, Greece, June 2003.

Rodden K. et al. (1999). Evaluating a visualization of image similarity as a tool for image browsing. In INFOVIS'99. IEEE Computer Society, 1999.

Rodden K. et al. (2001). Does organization by similarity assist image browsing? In ACM-CHI'01. pp 190-197, ACM Press, 2001.

Rorissa A et al (2007). Exploring the relationship between feature and perceptual visual spaces. Journal of American Society for Information Science and Technology (JASIST), 58(10): 1401-1418 (2007).

Rubner Y (1999). Perceptual Metrics for Image Database Navigation. PhD thesis, Stanford University, 1999.

Rubner Y. et al. The Earth Mover's distance as a metric for image retrieval. International Journal of Computer Vision, 40(2):99-121, 2000.

Santini, S et al. (2001). Emergent semantics through interaction in image databases. IEEE Transactions on Knowledge and Data Engineering, 13(3):337-351, 2001.

Schreiber A.T. et al. (2001). Ontology-based photo annotation. IEEE Intelligent Systems, 16(3):66-74, 2001.

Székely E. et al. (2007). Clustered Multidimensional Scaling for Exploration in IR. In 1st International Conference on the Theory of Information Retrieval, Budapest, Hungary, 2007.

Smeulders AWM et al. (2000). Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12): 1349-1380, 2000.

Stan D. and Sethi I. K (2003). *e*ID: A system for exploration of image databases. Information Processing and Management Journal, 39(3):335-361, 2003.

Tenebaum J.B. et al. (2000) A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500):2319-2322, 2000.

Urban, J et al. (2006a) An adaptive technique for content-based image retrieval. Multimedia Tools and Applications, 31(1):1-28, 2006.

Urban, J. and Jose, J. M. (2006b) Evaluating a workspace's usefulness for image retrieval. *Multimedia Systems* 12(4-5):pp. 355-373.

Vendrig J. et al (2001). Filter image browsing: interactive image retrieval by using database overviews. Multimedia Tools and Applications, 15(1):83-103, 2001.

Vertan C. et al. (2002). Browsing image databases with 2D image similarity scatter plots: update of the IRIS system. International Symposium on Communications, 397-402, 2002.

Wang J.Z. et al. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(9):947-963, 2001.

Worring, M. (2007) The MediaMill semantic video search engine. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages -. Honolulu, Hawaii, USA, April 2007

Yang J. et al. (2006). Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis. IEEE Symposium On Visual Analytics Science And Technology, pp 191-198, 2006

7. Multilingual/Multimedia Information Retrieval

by Gareth J.F. Jones with contributions from Martha Larson and Stephane Marchand-Maillet

In common with many areas of language processing, the origins of information retrieval (IR) research are to be found in the exploration of techniques for electronic English language text archives. A number of successful models for information retrieval have been, and continue to be, developed with English language documents as their primary research focus.

However, English language document collections, and electronic text documents in any language, represent only a minority of the information sources that a user may wish to search to satisfy their information need. The need to expand the scope of IR research beyond English text has been recognised in the last 15 years. Increasing amounts of work have been conducted and reported which explore non-English IR, cross-language information retrieval (CLIR), multilingual information retrieval (MLIR) and multimedia information retrieval (MIR). This work has greatly increased understanding of the issues of multilingual and multimedia information retrieval and access. A range of techniques have been proposed, explored, evaluated and refined. However, the techniques are imperfect and many challenges remain to improve effectiveness and to extend the scope of retrieval tasks. This will require a deeper investigation of the issues and problems than has been carried out so far together with the introduction of novel techniques.

When efforts to expand the horizons of IR began it was not at all clear what retrieval methods should be adopted for these new tasks in order to achieve the greatest IR effectiveness. It was found that established IR methods transferred well to other languages, and linguistic media, speech and scanned text images. The reason for this result should perhaps not be too surprising given the rigor and care taken over the years to ground these models in sound theoretical analysis, and the extensive experimental evaluations that have characterized this work. Significant issues arise with respect to translation between search topics and documents for cross-language and multilingual information retrieval. For MIR, there are significant issues related to the definition of retrieval units, i.e. what should we look for in an image or video, and the accuracy with which features can be detected automatically once they have been defined.

This chapter continues in the next section with a brief review of the relevant details and indexing assumptions of text IR. Section 6.2 describes experimental work with non-English test collections, this is extended in Section 6.3 which gives results for cross-language and multilingual IR. Section 6.4 introduces multimedia IR and highlights some relevant experimental work. Finally, Section 6.5 draws conclusions from existing work and looks toward future applications and challenges.

6.1 Probabilistic Models and Feature Indexing

IR systems seek to satisfy a user's *information need*. Current IR systems attempt to do this by locating *relevant* documents from within which the user can extract the required information. Potentially relevant documents are selected and returned to the user based on a retrieval model taking the user's query as input. The retrieval model can make use of whatever information is made available about the documents from among which it is seeking to locate the relevant ones. Document information is most typically based on simple extracted attributes such as words present in a document, but may include phrases or other extracted features; additionally features may be annotated with functional details such as their part-of-speech or semantic details such as those representing a geographic place or a time. While such annotations are not generally used within retrieval models which are normally based on word-level features, they can be useful for document browsing interfaces using maps or timelines, or for more advanced retrieval applications such as question answering systems which usually include some degree of language processing to locate the answer to a user's questions from within the available documents.

Document retrieval models fall into two broad classes of Boolean and best-match, the latter being the dominant modality of searching in current IR research. Boolean retrieval uses search queries constructed using Boolean logic to select documents which match these criteria from the available collection; the documents are returned to the user unsorted. The user must then browse among the returned documents

either randomly or using some potentially useful criteria such as the date of creation, author, or document source. The requirement for complex search queries and the absence of content-based ranking means that it is unattractive to the majority of users of search engines who lack the enterprise to construct complex queries and desire the simple way of determining which documents are most likely to be of interest to them provided by ranked best-match IR. Over the years, many best-match ranked retrieval models have been proposed and evaluated. The most popular models being: the vector-space approach [Salton & Buckley, 1988], the probabilistic model [Robertson & Sparck Jones, 1976], and more recent methods based on statistical language modelling [Ponte & Croft, 1998].

If we had a complete model of each document, describing all potentially important features, with a correspondingly detailed model of the information need expressed by the search request, we might expect perfect retrieval with all relevant documents having higher probabilities than non-relevant documents. Alas such document models do not currently exist, and indeed the expression of information need in the search request is often an insufficient or inaccurate expression of the user's information need. Due to these deficiencies, retrieved ranked document lists generally interleave relevant and non-relevant documents. The objective of research in ranked IR is to improve the reliability of these imperfect relevance probability estimates.

Every document can be assumed to be a unique event, and in general, we take it that the description of each document used for retrieval is similarly unique. A problem arises with this modelling assumption, since it is difficult to assign probabilities to unique events. A solution comes in the form of decomposing document descriptions into their non-unique components or attributes, whose association with relevance can be estimated. These attributes can be used in combination to synthesise a relevance probability estimate for each unique document. The derivation of the early form of this practical probabilistic model (the "binary independence model") is described in van Rijsbergen [1979], and the more recent extended form of the model (well known as the "Okapi BM25"' model) in Sparck Jones et al. [2000a]. In the BM25 model the likelihood of relevance for a document *j* is computed based on the sum of the *combined weights cw(i,j)* of the independent attributes *i* which occur in both the document and the current search request. *cw(i,j)* values are computed based on the classic IR attribute weighting features of across document collection frequency (the *collection frequency weight cfw(i)*) of attributes *i*, the *within document frequency* of an attribute *i* in the document *j*, and an adjustment of the weight to compensate for document length [Robertson & Walker, 1994].

In general for current IR systems, each document is modelled as a simple "bag-of-words" which lists the attributes occurring within the document and their frequency of occurrence. The degree of match between a document *j* and the search request is then simply computed as a matching score *ms(j)* of the sum of the weights of the attribute in common between the request and the document. A list of documents ranked by matching score is then returned to the users. Documents are thus represented within the IR system as (assumed) independent attributes. The models used for ranked retrieval tell us nothing about the language of these attributes or even the media of the documents. Of course, much of the experimental work that established the effectiveness of this model has been carried out using English text collections often taken from general news or agency sources, but in theory there should be no reason why they cannot be used effectively for other languages, media or data sources.

Several well established techniques are typically applied for automatic indexing of English language text documents. These include removal of frequent *stop words*, such as those in van Rijsbergen's list [van Rijsbergen 1979], *suffix stripping*, using a method such as the Porter algorithm [Porter 1980], standardisation of spelling, and conflation of synonyms. Whatever preprocessing is applied, the features used for retrieval are still independent attributes derived from the document. Combined with enhancements such as relevance feedback and pilot searching using large additional document collections, these methods have shown effective retrieval in many evaluation tasks undertaken in the last 10 years or so.

The following sections look at the adaptations required for the application of IR methods to non-English documents, cross-language and multilingual information retrieval, and the effectiveness for multimedia information retrieval.

6.2 Non-English Information Retrieval

A key consideration when developing an IR system for a new language is the selection of the most suitable set of attributes to be used to index the documents. The lexical and structural differences between languages mean that the distributions of attributes within individual documents and across collections will vary between different languages. However, since the standard IR models make no explicit language dependent assumptions about these distributions, there is no reason to suppose that, with appropriately selected indexing units, they should not work effectively for any language.

From a linguistic perspective English actually provides a good starting point for the investigation of indexing methods and retrieval models. The basic word units of the language are easily identified, and the types and degrees of inflection of individual words are relatively simple compared to those of many other languages. There are of course many exceptions to these apparently simple rules of inflexion, and ongoing debate over the basic units of meaning, but generally these concerns can be safely ignored or handled by explicit exception lists for the purposes of IR indexing. Some other languages have similar properties to English while others introduce new issues which must be addressed for effective retrieval. This discussion outlines some of the features relating to indexing and retrieval of a range of representative languages.

From an IR perspective, languages such as French, Italian and Spanish can be addressed using adaptations of the techniques used for English. Thus for each language, we need to develop a suitable set of high frequency stop words that can be removed safely without affecting retrieval effectiveness, suffix stripping algorithms to conflate words to common stems, and appropriate synonym dictionaries [Wechsler, Sheridan, & Schäuble, 1997]. Standard IR methods using this approach have been shown to be effective in comparative evaluations of non-English IR tasks, for example within the Cross-Language Evaluation Forum (CLEF) workshop series [Savoy, 2004].

More complex issues are introduced by languages such as German and Dutch which are highly declensional with a rich system of inflections and cases [Braschler & Ripplinger, 2004]. In addition, in common with other Germanic languages, such as Swedish, and other languages such as Finnish, there is free compounding of words to express concepts developed from the component words. In these cases, although words are still the building blocks of the language, they are frequently combined into noun compounds without spaces. If one of these noun compounds appears in a search request and a document, there is a very good chance that this is a relevant document. However, the generative nature of the compounds means that often no match will be found for a search compound within the document set, even if the similar concepts are being described This can lead to many potentially relevant documents being missed, since they do not contain the compound in exactly the form used in the request. The general approach to this problem is to develop methods for compound splitting; these techniques may rely on the use of a compound dictionary or language specific rules for identifying word units within compounds, or a combination of both methods [Braschler & Ripplinger, 2004]. Of course, in addition to the decompounding of these concatenated words, indexing of these languages also benefits from the application of effective stemmers and removal of stop words.

Different issues arise in the case of east Asian languages such as Chinese and Japanese. The written form of these languages uses ideograms of Chinese origin. There are many thousands of these characters which usually have some meaning associated with them. Most words are formed by bringing two characters together. The meaning of the word is usually related to those of its constituent characters. Shorter words consisting of one character can express simple concepts and occasionally longer words more complex ones. While Chinese is restricted to a single character set, in the case of Japanese three additional character sets are in common usage: *hiragana* whose role is similar to function words and verb suffices in English, *katakana* which are used to transliterate Western concepts, e.g. *computer* appears phonetically in Japanese katakana as *ko n pu ta*, and *romaji*, for Western characters sometimes used for numbers and proper nouns. The major concern when indexing languages of this type is the observation that there are no spaces between the words of each sentence. The text must thus be segmented into suitable representative units prior to indexing. Further since the ideogram character set is itself so rich, there is a question of what the best units for retrieval actually are.

A number of approaches have been explored for indexing these languages. The most basic method is simply to take each character as an indexing unit, a slightly more elaborate one is to use overlapping n-grams of characters of varying lengths, while the most complex strategy is to apply morphological analysis to identify

the most likely word break points. A number of experiments using various Chinese and Japanese test collections exploring different approaches to segmentation have been carried out with inconclusive results, for example Huang & Robertson [1997] and Jones, Sakai, Kajiura, & Sumita [1998]. All the above approaches produce a good level of retrieval effectiveness.

6.3 Cross-Language and Multilingual Information Retrieval

Retrieval involving more than one language is broadly classified into two areas: cross-language information retrieval (CLIR), and multilingual information retrieval (MLIR). CLIR is concerned with the retrieval of documents in one language using search requests in another language, e.g. Dutch requests used to retrieve Italian documents. MLIR extends this to retrieval from a collection where documents are uniquely present in one language, but the collection overall covers documents in multiple languages, e.g. using an English request to retrieve from a collection with documents in English, Dutch, Spanish, and Italian. In practice, more complex situations are clearly possible. A single document may contain material in more than one language, and individual documents may be repeated in different languages within a collection. From these definitions it can be argued that CLIR is really a subset of MLIR. This section introduces research questions posed by CLIR and MLIR, and outlines the main solutions that have been proposed and explored to date.

6.3.1 Cross-Language Information Retrieval

The principal question that arises in the context of CLIR is: how should the language barrier between the search requests and documents be crossed? Should search requests be translated into the language of the documents, should the documents be translated into the language of the request, or both? Further, what is the best approach to carrying out this translation?

Request Translation vs. Document Translation

There are well rehearsed arguments for and against request or document translation, with the main issues relating to translation cost, at what stage it is carried out, its effectiveness for retrieval, the available translation and computational resources, and the storage implications.

Generally it is held that translating requests when they are entered will be fast enough, since they are likely to be short, not to interfere with interactive searching. Unfortunately, short requests often have minimal formal linguistic structure, and further because they are short, there is little information of the context in which the request words have been selected by the user. These factors mean that it will often be difficult to perform reliable deep linguistic analysis when attempting to perform translation of the request. One consequence of this is that it can be difficult to select the contextually appropriate translation of polysemous words. A further implication of attempting to translate short requests is that the mistranslation of individual words can have a significant impact on retrieval effectiveness. However, since the document collection to be searched will not have been translated, and is therefore accurate, redundancy effects are often found to help to ameliorate translation errors even for short requests. It is further frequently argued that, since deep linguistic analysis of a request may not be possible (or if possible may not be desirable, if it is likely to be unreliable), and since we are only seeking to transfer the words into another language, shallower translation methods may be better for request translation CLIR.

Consider now the alternative approach of document translation. Documents are generally much longer than search requests, and the content will often be linguistically well structured with large amounts of contextual information available. Thus translation of documents using formal linguistic analysis is potentially more accurate than it is for requests. This may not be the case for web content where content is often more informally structured without formal sentences. However, even in this case the amount of contextually related material in the document may assist in accurate translation. While they may generally be translated more accurately than short requests, translated documents will nevertheless contain a number of errors arising from incorrect analysis of the source text and limitations of the translation dictionaries. These errors will inevitably impact adversely on retrieval accuracy for CLIR. However, adopting document translation does mean that no translation has to take place when the search request is entered, so the retrieval stage itself is computationally faster and cheaper. Also, the search request is now accurate, with no possibility of translation error. A major disadvantage of document translation is the very high cost of translating all the documents. Although, since translation is done in advance of retrieval and only has to be done once, it can really be regarded as part of a very expensive indexing process. However, there are storage implications

which arise from the need to maintain a separate search collection in each request language into which the documents are translated.

Experimentally both request and document translation have been shown to be effective, with at least one study showing that combining the retrieval output of both methods used independently can produce the best overall retrieval effectiveness [McCarley, 1999].

One way to address the problem of storage is to translate all documents into a single "pivot" language, most probably English, and then to translate the requests into this same language when they are entered. This has the disadvantage that since both the requests and documents are being translated, translation errors will be compounded with a consequential impact on retrieval effectiveness. Pivot languages can also be used when resources are not available to translate directly between the request and document languages [Gollins & Sanderson, 2001]. In this case they can be used for translation of both requests and the documents into the pivot language, or for sequential translation of either the requests or documents into the language of the other.

Translation Methods for CLIR

Another widely debated issue in CLIR is how the translation should be carried out. The issues here relate both to the actual best means of translation for CLIR, were a perfect translation resource to be available, and the most appropriate method, where technical and resource limitations mean that real translation systems are currently far from perfect. Broadly speaking the three translation strategies that have been explored for CLIR can be categorised as: dictionary-based, comparable corpora, and machine translation.

Most early work in CLIR advocated the use of bilingual dictionaries for topic translation, with a variety of elaborations to improve their effectiveness for this task [Hull & Grefenstette, 1996]. In its simplest form, this approach replaces each word in the search request with all possible translations of the word in the document language appearing in a bilingual dictionary. As well as including the appropriate translation, if it is available in the dictionary, this simple method often introduces many contextually inappropriate translations of this word. These incorrect translations have been shown to significantly degrade CLIR retrieval effectiveness relative to monolingual IR for the same set of requests and documents. It has been demonstrated that dictionary-based CLIR performance can be improved by using careful phrase translation and relevance feedback both prior to and after translation of the request [Ballesteros & Croft, 1998].

Given the problems with ambiguity arising from the use of bilingual dictionaries, and the gaps which occur with regard to their coverage of domain specific vocabulary items, alternative methods have been explored which align comparable corpora in the different languages [Sheridan & Ballerini, 1996]. Related terms appearing in this aligned content are used to translate requests in a context specific way. One of the problems with this strategy is that suitable related corpora are often not available for alignment. A widely explored way to overcome this problem is to use content from the internet [Nie, Simard, Isabelle, & Durand, 1999]. In this approach, large numbers of web pages are collected and aligned, and then used for request translation. Nie et al. demonstrated that an improvement in retrieval effectiveness can be obtained by using the aligned web documents in combination with a bilingual dictionary.

Perhaps the most obvious solution to crossing the language barrier between requests and documents is to use a standard commercial machine translation system. Indeed for CLIR using document translation, machine translation would appear to be the only realistic option given the huge amount of ambiguity that the other translation methods would introduce. Certainly I'm not aware of work which attempts to translate whole document collections using a different method. The arguments in favour of machine translation for CLIR centre on the potential for accurate translation of the words, appearing in the request or the document, which can be achieved by bringing sophisticated translation resources to bear on the task. Current machine translation systems often produce rather unnatural prose output. However this is not a problem for CLIR where we are only interested in the reliable translation of words with good relevance selectivity. The arguments against machine translation for CLIR are based on the previously stated issues of poor linguistic structure in search requests, which can render them difficult for formal linguistic analysis using machine translation, with consequential translation failures and inappropriate translation of words. Dictionary limitations can also result in translation problems for both requests and documents. This latter issue is likely to pose particular challenges for domains and their associated specialist topics which will often be outside the general purpose vocabularies used for developing the standard versions of commercial translation

systems. Specialised dictionaries can be available to adapt machine translation to specific domains, but these are only likely to be available commercially for domains where the financial returns are deemed likely to justify the significant investment required to develop them.

An experiment at Toshiba performed a comparative evaluation of progressively more sophisticated request translation strategies ranging from simple bilingual dictionary lookup, to part-of-speech tagging, sense disambiguation, and full machine translation for an English - Japanese CLIR task [Jones, Sakai, Collier, Kumano, & Sumita, 1999]. Perhaps surprisingly given the arguments against machine translation for CLIR, the best retrieval effectiveness was found using full machine translation. This result was observed for both natural language request statements, and requests modified to disrupt the linguistic structure by removing the function words prior to translation. More recent experiments have shown that a combination of machine translation and the BM25 ranked retrieval model combined with relevance feedback produces among the best reported effectiveness for the CLEF CLIR tasks [Jones & Lam-Adesina, 2001] [Lam-Adesina & Jones, 2003]. Analysis of the retrieval behaviour of individual requests showed that there is sensitivity to the failure to translate important words, usually previously unseen proper nouns. For example, failure to translate phonetic loan word proper nouns rendered in katakana in Japanese if they are not present in the translation dictionary significantly degrades retrieval effectiveness. This will often be a problem for bilingual dictionaries as well; although, the impact on retrieval performance may be masked by translation ambiguity issues. However comparable corpora should be able to capture these domain specific translations, as long as they include documents covering the appropriate related topics in their training set. It should be noted that in all cases the documents used in these experiments were taken from published news corpora, and the results may not extend to material that is not formally published and/or is outside the topics encountered in national and international news stories.

Many papers have been published describing CLIR results in more recent years. The references included here are generally those which first introduced or advocated a particular translation approach for CLIR, in each case subsequent work has often extended these methods. While machine translation shows good results when available, bilingual dictionaries and aligned corpora are an important translation resource for CLIR with language pairs for which well developed machine translation tools are not available, and most likely where structural and domain issues render machine translation systems less effective, although this latter points remains to be illustrated in practice. There are direct bilingual dictionaries available between most major languages pairs, and even for minority languages there are bilingual dictionaries to major languages such as English, while the expanding amounts of electronic text available from many sources mean that corpus-based methods will become an increasingly important resource for translation in CLIR.

6.3.2 Multilingual Information Retrieval

In MLIR the IR system is expected to respond to a search request in one language by generating a ranked list of potentially relevant documents in multiple languages. Similar to CLIR, MLIR can be approached using either a request or document translation strategy. The challenges of MLIR include similar translation issues to CLIR; however it also introduces a significant new problem which arises because the documents in each language will often be in separate collections. In a practical system, document collections may be geographically distributed with no option to merge them into a single collection. However, even if the documents can be combined into a single physical collection, the fact that they are in different languages means that semantically related search terms cannot be conflated, and effectively the collection will still behave as separate, language specific, sub-collections. The major difficulty that arises for MLIR is how to take the separate outputs from searching individual collections and merge them into a single output list for delivery to the user, which reliably ranks relevant documents higher than non-relevant ones. For this reason, MLIR is often seen as being akin to monolingual distributed IR, where separate search collections are stored and searched independently for practical or commercial reasons; lists retrieved from the individual collections must then be merged to form a single ranked list output [Callan, 2000]. There are also potential issues or retrieval effectiveness arising from the separation of the overall ``virtual'' document collection into multiple smaller collections since the term weights may be less accurately estimated within the smaller collections.

In MLIR the merging problem arises since ranked lists from the separate collections will be generated using different indexing strategies, and, as discussed earlier, the features will have varied distributions for the individual languages. This means that the document matching scores from the retrieved ranked document

lists will generally be incompatible. For example, documents retrieved from a collection with higher average matching scores will tend to be favoured in the merged list. Thus the list may be biased towards certain collections and hence languages, regardless of the actual relative likelihood of documents retrieved from these collections being relevant. If this problem is overcome, a further concern is that the matching score profiles of the lists may be different. Hence the lists cannot be merged in a simple reliable way. In general for distributed IR, difficulties of list merging vary depending on the number of differences between the IR systems used to compute the separate lists, and potentially the cooperation between the maintainers of the separate search engines. The separate retrieval engines may reliably make all statistics of their collections available to the merging algorithms, they may make some subset available (potentially of questionable reliability) or they may make no information available beyond identifying the documents and their retrieval rank [Callan, 2000]. The amount of information available from the separate collections affects the complexity of the merging strategy that can be adopted. If the separate retrieval systems use different retrieval ranking algorithms then the scores will clearly be incompatible, but even if an identical retrieval strategy is used for all the collections, the matching scores will be incompatible due to the different values used to estimate the term weights or other ranking parameters. In MLIR, these issues are compounded by problems arising from variations in the properties of the languages. For document translation MLIR, if the document index data are located physically together, the index files can be combined to form a single search collection. This removes the need for merging of separate lists. However, if the collections are distributed or query translation is being used, some method of merging must be adopted.

A variety of list merging algorithms of varying complexity have been proposed for distributed IR. A number of these have been applied for MLIR with varying degrees of success. The simplest approach involves ignoring the score incompatibility problem, and simply merging the ranked lists using their raw scores. More complex methods involve ranking the separate collections in terms of their estimated likelihood of containing relevant documents, combining these collection matching scores with the matching scores of individual documents to form a composite score, and using this combined score to generate the final merged document list. These methods have been shown to be effective for monolingual distributed IR [Callan, 2000]. Unfortunately, they have not proved so successful for MLIR, where it has been difficult to improve performance beyond that achieved using the simplest methods [Lam-Adesina & Jones, 2003; Savoy, 2004].

In our experiments for the CLEF workshop MLIR task in 2003, we translated all the documents from their original languages of French, German and Spanish into English using machine translation. We then compared retrieval effectiveness of various list merging strategies with that for a single collection formed from the translated documents. Overall we found that the single collection method worked best indicating that all the merging strategies fell short of the performance that could potentially be achieved using these document sets [Lam-Adesina and Jones, 2003]. Once again our results showed that the BM25 Okapi probabilistic model produced among the best retrieval effectiveness for this task. Of course it will not always be possible to translate the entire retrieval collections and then combine them. More recent experiments using the CLEF 2003 MLIR tasks have shown that list merging can produce good retrieval results [Si and Callan, 2006]. However, merging remains an important ongoing concern for MLIR requiring further investigation.

6.3.3 Multilingual Web Retrieval

In recent years significant effort within the information retrieval research community has focused on the development of effective methods for retrieval of web content. This has gained momentum since the late 1990's, but is still a young area of research, and although many important results have already been attained, open problems remain that require further research, as is observed in Melucci and Hawking [2006].

Given its world-wide coverage, it is no surprise that the Web is inherently multilingual. Dominant world languages are all well-represented on the Web. Some multilingual Web content is created by translation between languages but, predominantly, documents appear in the languages they were originally authored in. The result is a heterogeneous body of information in which is content available in one or more languages with no guarantee that it will be duplicated in another language. The importance of developing approaches to improve access to multi-language Web collections has been recognized by the international research community, which has established exercises such as the Web track at CLEF, which promotes synchronization between researchers working in the area by developing systematic tasks, test-suites and evaluation of web content [Sigurbjörnsson et al. 2005; Balog et al. 2006].

Not only is the content of the Web multi-lingual, the users who wish to access this content are also polyglots [Sigurbjörnsson et al. 2005]. Especially in Europe, many users are able to make use of information presented to them in a range of languages. These users are quick to make use of the passive knowledge that they may have of a specific language, especially in cases when they realize that the information that they need is not available in another language. In addition, machine translation techniques offer a huge potential to support users in making use of information in languages that they do not understand at all.

Like classic information retrieval, web retrieval attempts to provide a user with information that satisfies an information need. However, many users undertake web search to find a particular URL or to perform a particular transaction rather than to find information [Broder 2002]. Also, frequently users like to browse in web collections, which means that web retrieval research must also focus on the question of providing information to a user who has no clearly formulated information need, but instead requires an overview of an area. One particularly challenging task is to provide web retrieval techniques that will support users who are browsing with the purpose of discovery of new subject areas that they were previously unaware of, or who are interested in finding new connections between topics that they are already familiar with.

Other differences between retrieval in digital libraries containing text documents and search in Web content concern the difference in the nature, structure and volume of information available on the Internet, as discussed by [Baeza-Yates and Ribeiro-Neto 1999]. On the Internet, data is changing constantly. Because it is produced by a variety of sources, both professional and informal, Web data is fundamentally heterogeneous and its quality is variable. The amount of data available on the Internet is unrivalled in volume, constituting a particular challenge for Web search. Finally, data available on the Internet is distributed, meaning that before it can be indexed it must be gathered. Gathering of data requires a web crawler to discover and fetch web content so that it cab be indexed for searching. A challenge for the future is to design and implement web crawlers, whose efficiency stems from their intelligence, i.e. their ability to crawl only that material that will later be relevant to the information needs of the users of the search engine they were designed to feed. This issue is important for the harvesting of content for the MultiMatch search engine where crawled content should be drawn from the broad domain of cultural heritage.

Web retrieval can exploit normalization and term extraction techniques that have been developed for classic text retrieval, but also makes use of characteristics particular to Web content. Web retrieval makes critical use of the fact that web pages do not exist as isolated entities, but are connected to each other via hyperlinks. The most well known exploitation of this link structure is the PageRank algorithm which formed the starting point for the development of the Google search engine (see Chapter 3). A future direction for Web retrieval is to make full use of the structural information provided by the tree structure of XML documents and by the information contained in the XML tags. Fielded indexes that index path-tagged terms have demonstrated great potential and the future will surely see optimization of such techniques.

Alongside search engines which accept free text queries from users and deploy automatic methods to determine relevant websites, search engines based on hand crafted web-categories are also being developed. Such search engines supply users with high quality information, but suffer from the disadvantage that they do not provide wide coverage since the classification of sites into categories has to be done by hand and is very time consuming [Baeza-Yates and Ribeiro-Neto 1999]. A research direction for the future is to pursue approaches that will deliver the benefits of category-based search, but with reduction or near-elimination of human effort.

The Internet has witnessed the development of a profusion of search engines, each deploying its own crawler and its own search strategies. For this reason, the results delivered by one search engine have a great potential to complement the results returned by another. The bundling of search engine results is another important area of investigation for researchers involved with web retrieval.

The Internet is characterized by the existence of user communities which create content and interact with one another. The structure of these communities is an important source of information. Some communities engage in concerted effort to label web sites that are relevant to their interests with tags that will make them easily retrievable. Log files of user behaviour is another source of information. Patterns of previous searches can be used to refine future searches. For some types of searches, it is critical that a search engine returns

reliable information to the user. Although every query deserves a reliable result, travel and medical queries can be particularly critical. For this reason, it is important to analyze the quality and the authority of web pages and for search engines to be aware that content providers may be trying to trick them into indexing pages that are not truly authoritative. (add probably several relevant citations which cover these points).

In sum, techniques required to tackle the challenge of web retrieval encompass, but extend approaches to text retrieval. Understanding how users formulate their information needs into queries for web search and exploitation of the particularities of web content are both necessary if web retrieval technology is to advance into the next generation. Web retrieval research stands to gain by embracing the multilingual nature of the Internet and leveraging complementary sources of information in multiple languages.

6.4 Multimedia Information Retrieval

The current expansion in archives of digital multimedia content is creating the need for tools to automatically search and retrieve material from these collections. Similar to the work on multilingual text documents, recent years have seen a rapid increase in research exploring Multimedia Information Retrieval (MIR). Multimedia archives comprise material in one or more of audio or visual media, often accompanied by some form of manual electronic text annotation or metadata. Retrieval from these collections raises a number of issues with respect to both the indexing and retrieval processes. Multimedia content can be either static, in case of individual digitized images such as photographs or paintings, or temporal, comprising audio and/or video content. The static or temporal nature introduces various concerns with respect to the presentation to the user and browsing of retrieved content.

Indexing and retrieval methods for MIR depend on the media under consideration. Let us consider these in order of increasing complexity. Electronic text material available for MIR can either take the form of metadata or direct transcription of content. Metadata may describe the content in some way, e.g. the names or roles of the characters appearing in an image, or the events taking place in a video. Transcriptions of linguistic content may be generated manually or automatically. For example, the close captioning often broadcast with TV sources can be captured and used as a high quality transcription of the content for the purpose of retrieval and browsing.

Existing IR research has focussed very much on linguistic content, and so can in general be applied directly to manually annotated material associated with multimedia content. The usefulness of manually entered descriptive metadata will depend on the quality of the data, and its usefulness in addressing an individual need. Thus, while the visual content of an image may make it relevant to a particular request, if the descriptive metadata is not pertinent to the aspect of this item which makes it relevant, then the MIR system will fail to locate it. Therefore, the effectiveness of MIR will clearly be affected by the accuracy and richness of the annotation. Additionally, the complexity of the retrieval methods used for textual annotations may be influenced by their form; if the annotations are highly structured, this may be taken into account in the retrieval algorithms adopted.

Of more interest within recent and current research, is MIR based on automated annotation of the content. The following sections consider indexing and retrieval for first spoken documents, and then image and video data.

6.4.1 Spoken Document Retrieval

In many situations it is uneconomic or impractical to manually transcribe the spoken contents of multimedia documents, and thus transcriptions must be generated automatically using speech recognition technologies. Forming transcriptions in this way using current speech recognition tools has a number of limitations. The most significant issue is that, like machine translation systems used for CLIR, these tools make mistakes; incorrect words can be inserted into the transcription, correct words deleted, or one word incorrectly substituted for another one. These errors arise for a number of reasons relating to both the natural language data and the tools themselves. Speech recognition is inherently challenging for a number of reasons including the following: the speech may be poorly articulated, it may not follow expected linguistic patterns, it may be captured using poor quality equipment, there may be high levels of background or environmental noise, or there may be crosstalk where more than one speaker is talking at the same time. The accuracy of a speech recognition system is limited by the effectiveness of its acoustic models to accurately recognise the

sound patterns of the current speaker, and of its language models to predict their use of word patterns. Current speech recognition transcription systems are also correctly described as "large vocabulary", where only the words within a predefined vocabulary can be recognised correctly; other so called "out-of-vocabulary" words will be transcribed incorrectly by definition. In general, the overall accuracy of an automatically generated document transcript will depend on the extent to which the speech deviates from the trained parameters of the speech recognition system and the quality of the input speech signal.

The effect of recognition errors is to produce a "noisy" transcription which will have some similarities to the output of a machine translation system. The characteristics of the errors however are likely to be somewhat different. A machine translation system can determine its output, although it may experience problems with the naturalness of the word patterns generated, or be subject to limitations in the richness of the available vocabulary or linguistic structures. By contrast, a speech recognition system must do its best to transcribe the data presented to it. Automatic transcriptions often include apparently random insertion and deletion errors. A potential problem for both machine translation and speech recognition though is how to appropriately handle input words outside their vocabulary.

Research into spoken document retrieval (SDR) began with a number of projects in the early 1990s. These examined various approaches to automatically indexing the spoken contents and were evaluated using locally developed test collections [Glavitsch & Schäuble, 1992; Jones, Foote, Sparck Jones, & Young, 1996]. When these projects started, the potential of IR techniques derived from experience with electronic text documents to transfer successfully to errorful spoken document index files was very much an open question.

It is a feature of speech recognition that the hardest words to recognise accurately are often short function words. Of course, these are generally not useful for retrieval, and hence SDR systems can still operate with good reliability in the presence of relatively high word recognition error rates. A further issue is that since important words within a document are often repeated, even if the word is recognised incorrectly when it occurs in one place, it may be correctly recognised elsewhere in the document. Whilst errors of this type will degrade the overall quality of term weights, the documents will still be retrieved. This distortion of term weights can result in some distortion of the ranked retrieval list, relative to that which would be achieved with a perfect document transcription, but overall high levels of retrieval effectiveness can still be achieved.

Interest in SDR increased significantly in the mid-1990's and a track was introduced at the annual TREC series in 1997. For the first time researchers were able to work with a common SDR test collection. The SDR track ran for 4 years, each conference increased the document collection size or the complexity of the retrieval task. During this time speech recognition technologies continued to advance. Using the best available transcription systems, achieving recognition average word errors rates of around 20% with a vocabulary of around 65,000 words, together with the BM25 model and retrieval enhancement techniques, such as relevance feedback and merging with in-domain large contemporaneous text collections, TREC SDR participants demonstrated similar overall retrieval effectiveness for manual and automatic document transcriptions [Johnson, Jourlin, Sparck Jones, & Woodland, 2001] [Garofolo, Auzanne, & Voorhees, 2000]. The success of the TREC SDR track indicated, at least for a task where the transcription system can be well trained for the domain of the document collection, in this case broadcast news, that SDR is effective using current speech recognition technologies.

More recently the Cross-Language Speech Retrieval (CL-SR) task at CLEF in 2005 and 2006 has explored speech retrieval for a more challenging document collection in a cross-language framework. Each document consists of multiple fields consisting of: an automatic transcription made with a large vocabulary automatic speech recognition system adapted to the domain of the data, a number of keywords assigned automatically based on these transcriptions, manual assigned keywords, a short manual summary of the document and a manually assigned list of proper nouns appearing in the actual audio of the document. This document set thus poses the challenges of SDR, but also the combination of multiple fields for effective retrieval. The optimal way of doing this is not obvious as explained in [Robertson et al, 2004]. Cross-language experiments carried out by the participants in the CLEF tasks show that speech retrieval behaves similarly to standard text retrieval in cross-language tasks; that is problems of translation between search requests and documents result in a reduction of retrieval effectiveness of between 10% and 20% [White et al, 2006].

6.4.2 Image and Video Retrieval

Whereas it is natural to use the same indexing units for spoken content and written linguistic content, the appropriate mechanism for indexing and retrieving from visual media is much less clear. Visual content can include natural scenes either in static images or moving video, as well as other image content, for example scanned or overlaid textual material.

Considering first the more straightforward case of textual content in images. The first stage in automatically indexing this material is to identify zones or regions in the image containing text. The text in these zones is then recognised using an optical character recognition (OCR) process. After this, it can be indexed using a standard retrieval approach derived from experience with electronic text documents. Unfortunately, similar to speech recognition systems, OCR systems make mistakes; although the errors in this case are often of a different form. Instead of making whole word recognition errors, as is the case for speech recognition, OCR systems typically make errors in the recognition of individual characters. Each of these errors will usually introduce a new word into the indexing vocabulary of the collection. These words will not be useful indexing terms, since they will not match correctly with terms appearing in typed search requests, and they will also have disproportionately high collection frequency weights, since they are very rare within the document collection. A simple way to resolve this problem might be to attempt to correct automatically the spelling of these words using a dictionary. However, it is not always clear what the correct word should be. Indeed sometimes a word not present in the dictionary will actually have been correctly recognised by the OCR system, and attempting to correct OCR errors in this way may replace these accurately recognised words with incorrect words taken from the dictionary. As a consequence of this problem, "correcting" the OCR output with a dictionary may lead to a degrading of retrieval effectiveness. Another issue, similar to spoken document recognition, is that the accuracy of the output of an OCR system will be related to the difficulty of the recognition task. OCR accuracy will depend on the quality of the printing, the fonts used, and the contrast between the print and the paper. For example, modern laser printed output with a simple font is easier to recognise than older mechanically printed documents for which the paper may be yellowing with age. Significantly more difficult to recognise accurately is handwritten text, for which accuracy will obviously depend on how clearly it has been written, as well as the other factors affecting printed text [Rath, Manmatha, & Lavrenk, 2004]. Interestingly, while relevance feedback has been shown to be very effective for SDR [Johnson et al., 2001], the differences in error types encountered between OCR and speech generated transcripts, mean that it does not transfer to scanned text documents in a simple way and correction techniques must be applied to make it effective for this task [Lam-Adesina & Jones, 2006].

A much less well defined task is the retrieval of multimedia documents based on non-linguistic visual content. When examining a visual scene, we might want to identify any number of different features. For example, we may wish to recognise the individuals appearing in the image, the place where the scene is taking place, the objects in the picture, or perhaps the events being depicted. Identifying these features is very difficult. Indeed doing this in a robust way outside a very narrow pre-defined domain is currently not possible. Much visual media can be interpreted in a seemingly unlimited, often subjective, number of ways. This type of intelligent analysis will be beyond analysis of visual features alone, often requiring knowledge outside that available in the visual content itself. Of course, texts can frequently be interpreted in many ways as well, but for retrieval purposes, word level indexing has generally been shown to be effective without needing to determine any particular interpretation of the text. In the case of images, not only are attempts at recognising features unreliable, there is no obvious parallel means of selecting indexing units for open domain retrieval. Current video media retrieval systems either focus on very narrow domains, for example identifying pictures of predefined named individuals, or seek to index images using low-level features, such as colour or texture. Indexing images using such low-level features is perhaps comparable to identifying the letters in a text document without determining what the words are. A detailed summary of much of the work carried out in developing image and video retrieval technologies is described in [Smeulders et al, 2000], Much research is currently devoted to the segmentation of images into meaningful regions or to detect objects without extensive training to identify specific object types.

The difficulty in indexing images and of specifying search queries for them means that retrieval of visual media inherently requires more user interaction than text retrieval. For MIR systems, a user will typically initiate a search either using a text request which will locate some potentially relevant images or video based on their textual annotation, or they will select a sample image and request the retrieval system to "find me

more like this", in response to which the system returns images with similar colour and texture profiles to those of the example. The user is then able to provide feedback on the images retrieved using this initial query, after which further searches are carried out, with feedback after each one, until the user's information need has been satisfied. Such "more like this" searches are typically based on generic MPEG-7 low-level image features of: global colour, regional colour, texture and edges within the image. Some current research is extending this to explored interactive use of objects to enable users to select a combination of standard image features and detected objects in building more complex queries for feedback [Sav et al, 2006].

A significant challenge for MIR is the combination of the visual features with the textual metadata to provide an overall search output. Simple approaches to this are based on a simple data fusion strategy of forming separate ranked lists for each feature and then adding them in a weighted scalar sum. This is a simple strategy but can be effective, although it is important to assign the correct weights to each feature list. This also true of data fusion for text only retrieval, but is probably more crucial for MIR where the importance of individual features will be quite different for individual queries. For example, for one query colour of the query image may be important in finding relevant documents, whereas for another query a combination of metadata text and image texture may be important. A method to automatically select query dependent optimal features weights is introduced in [Wilkins, Ferguson and Smeaton, 2006].

While the above late fusion mechanism proves effective, it is important to define early approaches whereby the relevant features are combined at an early stage, thus enabling truly multimodal query. Important shortcomings however are the heterogeneity and normalisation of the features to combine. [Bruno *et al*, 2006] propose a distance-based learning strategy to combine multimodal feature at query time. Features are homogenized by considering relative distances rather than absolute values. A new representation space is thus created by an appropriate choice of pivot-like points. Efficient non-linear learning techniques (SVM, KFD) may then operate interactively within such a feature space based on user feedback to isolate portions of population relevant to the query.

The discussion so far really assumes that retrieval is of images with may be annotated with textual metadata. For video retrieval some additional processing and modelling is often required. Video is typically composed of events or scenes which are composed of a sequence of camera shots. Standard video processing typically first locates the shot boundaries, points at which the camera changes. Some camera changes are easy to locate others, such as gradual fades, can be problematic. Once shots have been identified, the next stage in video processing is typically to identify a single representative frame or "keyframe" for the shot. Retrieval for the shot then proceeds exactly as for static image retrieval using the keyframe. This of course assumes that a keyframe can be located which sufficiently represents the shot, such that it contains features which represent aspects of the shot that are going to appear in query images for which the shot is relevant. For some shots temporal features of the shot may be important in describing it, and in order to use this the temporal aspect of the image must be captured in some way.

Video shots are editing entities that may not be fully appropriate for video retrieval. A concept more advanced and probably more suited than the shot for searching is that of the *story*, somewhat close to the textual *topical* segmentation. In that case, the partition must be done according to semantic criteria gathered from a multimodal inspection of the streams (see e.g. Janvier *et al* [2005]). Semantic units are then said to be more appropriate for gathering relevance feedback than simple shots. The challenge here is to form a proper characterisation of the temporal evolution of the semantic information from multimodal features.

Since 2001 the TRECVID workshop has provided standard document collections for researchers to explore indexing and retrieval tasks for video data [Smeaton, Kraaij, & Over, 2004]. Tasks undertaken in TRECVID include: automated shot boundary detection, story boundary detection, visual feature recognition, locating named individuals or events in video, and interactive searching of a video archive. TRECVID is proving instructive in the development and evaluation of MIR technologies, but perhaps the clearest message so far is the large amount of work that remains to be done to achieve mature MIR systems.

6.4.3 Hybrid Searching for Multi-field Documents

The foregoing discussion has assumed that searching is based on a simple best-match ranked retrieval strategy. However, as has been mentioned a number of times documents are often accompanied by a range of metadata fields, such as date of creation, author, publisher or publication venue. A common approach to

exploiting this data in the search process is simply to fold it into the main document text field and use the attributes as search features. However, they can often be used instead, or additionally, as constraints on the search. For example to retrieve documents only published by a certain source or written by a named author within a specified time frame. Where the user has the requisite knowledge to impose these constraints limiting the document search space in this way can have significant benefits in terms of retrieval precision and efficiency of browsing. This can be particularly useful in multimedia environments where interactive constraints, particularly in audio browsing, mean that reducing the amount of material that must be explored in particularly useful [Brown et al, 1996].

6.5 Concluding Thoughts and Future Challenges

This chapter has demonstrated how fundamental work on English language text information retrieval has been successfully applied for multilingual and multimedia documents. For text retrieval in a new language it has been illustrated that the need is for the selection of appropriate indexing units and development of automatic indexing methods, including morphological processing, stop word lists, and suffix stripping algorithms. Research issues for CLIR relate primarily to translation methods to cross the language barrier between search requests and documents. In MultiMatch, we will advance automated translation in the CH area by using parallel corpora such as bilingual or multilingual metadata to automatically construct domain-specific dictionaries. These dictionaries will then be incorporated into a translation system with MT modules in order to translate search requests and CH documents.

For MLIR issues of translation are compounded with the need for effective merging of the document lists retrieved from different language collections. MultiMatch will investigate various techniques to find the optimal merging strategy for the CH domain and the multilingual indexes with which we will work. Speech and scanned text document retrieval have been shown to be remarkably robust to indexing errors in automatic recognition of their content. Research will be undertaken in MultiMatch to ascertain the most appropriate means of handling Cultural Heritage documents of these types. The ongoing issues of defining and recognising visual indexing features continue to be the focus of much research in visual media retrieval. However, there is already research underway exploring the use of the alternative language modelling approach to IR in visual retrieval [Westerveld & de Vries, 2004].

Solution of the problems of multilingual and multimedia information retrieval explored in this chapter does not represent the end of the story for research into information access technologies for this data. Research interest continues to evolve to embrace more challenging tasks. For example, work is currently being established in the areas of retrieval from multilingual collections of image and video archives, retrieval from multilingual web collections, and question-answering methods for multilingual and multimedia data.

References

Baeza-Yates, Ricardo & Ribeiro-Neto, Berthier. (1999). Modern Information Retrieval. Addison Wesley. 1999.

Ballesteros, L., & Croft, W. B. (1998). Resolving Ambiguity for Cross-Language Retrieval. In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 64-71, Melbourne, ACM.

Balog, K, Azzopardi, L., Kamps, J. & de Rijke, M. (2006). Overview of WebCLEF 2006. In Carol Peters, editor, Working Notes for the CLEF 2006 Workshop,2006. http://www.clef-campaign.org/

Braschler, M., & Ripplinger, B. (2004). How Effective is Stemming and Decompounding for German Text Retrieval? Information Retrieval, 7(3-4), 291-316, Kluwer.

Broder, Andrei. A Taxomony of Web Search. (2002). SIGIR Forum, 36(2) 2002

Brown, M.G.,Foote, J.T.,Jones,G.J.F., Sparck Jones,K. and Young, S.J. (1996) Open-Vocabulary Speech Indexing for Voice and Video Mail Retrieval, Proceedings of ACM International Conference on Multimedia, Boston, U.S.A., pp307-316, ACM.

Bruno E., Moënne-Loccoz N., and Marchand-Maillet, S. (2006) Asymmetric learning and dissimilarity spaces for content-based retrieval. In CIVR, pp 330-339.

Callan, J. (2000). Distributed Information Retrieval. In W. B. Croft, editor, Advances in Information Retrieval, pp. 127-150. Kluwer.

Chakrabarti, Soumen. (2003). Mining the web: Discovering knowledge from hypertext data. Morgan Kaufmann. 2003.

Garofolo, J. S., Auzanne, C. G. P., & Voorhees, E. M. (2000). The TREC Spoken Document Retrieval Track: A Success Story. In Proceedings of the RIAO 2000 Conference: Content-Based Multimedia Information Access, pp. 1-20, Paris.

Gey, F., Kando, N. & Peters, C. (2002). Cross language information retrieval: a research roadmap. SIGIR Forum, 36(2) 72-80.2002.

Glavitsch, U., & Schäuble. P. (1992). A System for Retrieving Speech Documents. In Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 168-176. ACM.

Gollins, T., & Sanderson, M. (2001). Improving Cross Language Retrieval with Triangulated Translation, In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Rretrieval, pp 90-95, New Orleans, ACM.

Grossman, David and Frieder, Ophir. Information Retrieval: Algorithms and Heuristics. Springer. 2004.

Huang, X., & Robertson, S. E. (1997). Application of Probabilistic Methods to Chinese Text Retrieval. Journal of Documentation, 53(1), 74-79.

Hull, D. A., & Grefenstette. G. (1996). Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 49-57, Zürich, ACM.

Janvier B., Bruno E., Marchand Maillet S., and Pun T. (2005). A contextual model for semantic video structuring. In 13th European Signal Processing Conference, EUSIPCO'05, Antalya, Turkey.

Johnson, S. E., Jourlin, P., Sparck Jones, K., & Woodland, P. C. (2001). Spoken Document Retrieval for TREC-9 at Cambridge University. In E. M. Voorhees and D. K. Harman, editors, Proceedings of the Ninth Text REtrieval Conference (TREC-9), pp. 117-126. NIST.

Jones, G. J. F., Foote, J. T., Sparck Jones, K., & Young, S. J. (1996). Retrieving Spoken Documents by Combining Multiple Index Sources. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 30-38, Zürich,. ACM.

Jones, G. J. F., Sakai, T., Kajiura, M., & Sumita, K. (1998). Experiments in Japanese Text Retrieval and Routing using the NEAT System. In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 197-205, Melbourne, ACM.

Jones, G. J. F., Sakai, T., Collier, N. H., Kumano, A., & Sumita, K. (1999). A Comparison of Query Translation Methods for English-Japanese Cross-Language Information Retrieval. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 269-270, San Francisco, ACM.

Jones, G. J. F., & Lam-Adesina, A. M. (2001). Exeter at CLEF 2001: Experiments with Machine Translation for bilingual retrieval. In Proceedings of the CLEF 2001: Workshop on Cross-Language Information Retrieval and Evaluation, pp. 59-77, Darmstadt, Springer Verlag.

Lam-Adesina, A. M., & Jones, G. J. F. (2003). Exeter at CLEF 2003: Experiments with Machine Translation for Monolingual and Bilingual and Multilingual Retrieval. In Proceedings of the CLEF 2003: Workshop on Cross-Language Information Retrieval and Evaluation, Trondheim, Springer.

Lam-Adesina, A.M. and Jones, G.J.F., (2006) Using String Comparison in Contact for Improved Relevance feedback in Different Text Media, In Proceedings of the 13th Symposium on String Processing and Information retrieval (SPIRE 2006), Glasgow, Scotland, pp229-241, Springer

McCarley, J. S. (1999). Should we Translate the Documents or the Queries in Cross-language Information Retrieval. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 99), pp. 208-214, University of Maryland, MD, ACL.

Melucci, M. & Hawking, D. (2006). Introduction. A perspective on Web Information Retrieval. Information Retrieval Vol 9. 119-122. 2006

Nie, J.-Y., Simard, M., Isabelle, P., & Durand, R. (1999). Cross-Language Information Retrieval Based on Parallel Texts and Automatic Mining of Parallel Texts from the Web. In Proceedings of the 22nd AAnnual International ACM SIGI Conference on Research and Development in Information Retrieval, pp. 74-81, San Francisco, ACM.

Ponte, J. M., & Croft, W. B. (1998). A Language Modelling Approach to Information Retrieval. In Proceedings of the 21st Annual International ACM SIGIR International Conference on Research and Development in Information Retrieval, pp275-281, Melbourne, ACM.

Porter, M. F. (1980). An algorithm for suffix stripping. Program, 14, 130-137.

Rath, T., Manmatha, R., & Lavrenko, V. (2004). A Search Engine for Historical Manuscript Images. In Proceedings of the 27th Annual International ACM SIGIR International Conference on Research and Development in Information Retrieval, pp369-376, Sheffield, ACM.

Robertson, S. E. (1977). The Probability Ranking Principle in IR. Journal of Documentation, 33, 294-304.

Robertson, S. E., & Sparck Jones, K. (1976). Relevance weighting of search terms. Journal of the American Society for Information Science, 27, 129-146.

Robertson, S. E., & Walker, S. (1994). Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 232-241, Dublin, ACM.

Robertson, S. E., Walker, S. & Beaulieu, M. M. (1999). Okapi at TREC-7: automatic ad hoc, filtering, vls and interactive track. In E. Voorhees and D. K. Harman, editors, Proceedings of the Seventh Text REtrieval Conference (TREC-7), pp. 253-264. NIST.

Robertson, S.E., Zaragoza, H., and Taylor, M (2004) Simple BM25 Extension to Multiple Weighted Fields, Proceedings of the 13th ACM International Conference on Information and Knowledge Management, pages 42-49, ACM.

Sakai, T., Koyama, M., Kumano, A., & Manabe, T. (2004). Toshiba BRIDJE at NTCIR-4 CLIR: Monolingual/Bilingual IR and Flexible Feedback. In Proceedings of NTCIR-4.

Salton, G, & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. Information Processing and Management, 24, 513-523, Elsiver.

Sav, S., Jones, G.J.F. Lee, H., O'Connor, N.E., and Smeaton, A.F., (2006t) Interactive Experiments in Object-Based Retrieval, In Proceedings of the 5th International Conference on Image and Video Retrieval (CIVR 2006), Tempe, AZ, U.S.A., pp,1-10, Springer.

Savoy, J. (2004). Combining Multiple Strategies for Effective Monolingual and Cross-Language Retrieval. Information Retrieval, 7(1-2), 121-148, Kluwer.

Sheridan, P. & Ballerini, J. P. (1996). Experiments in Multilingual Information Retrieval using the SPIDER system. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 58-65, Zürich, ACM.

Si, L. and Callan, J. (2005.) CLEF 2005: Multilingual retrieval by combining multiple multilingual ranked lists." In Sixth Workshop of the Cross-Language Evaluation Forum, CLEF 2005. Vienna, Austria.

Sigurbjörnsson, B., Kamps, J. & de Rijke, M. (2006). Overview of WebCLEF 2005. In Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J. F. Jones, Michael Kluck, Bernardo Magnini, Henning Müller, and Maarten de Rijke, editors, Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum (CLEF 2005), volume 4022 of Lecture Notes in Computer Science, pages 810-824. Springer Verlag, Heidelberg, 2006.

Smeaton, A. F., Kraaij, W., & Over, P. (2004). The TREC Video Retrieval Evaluation (TRECVID);' A Case Study and Status Report. In Proceedings of RIAO 2004 – Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval, pp. 25-37, Avignon.

Smeulders, A.W., Worring, M., Santini, S., Gupta, A.,and Jain, R. (2000) Content-Based Image Retrieval at the End of the Early Years, IEEE Trans. Pattern Analysis and Machine Intelligence, 22(12):1349--1380, IEEE.

Sparck Jones, K., Walker, S., & Robertson, S. E. (2000a). A probabilistic model of information retrieval: development and comparative experiments: Part 1. Information Processing and Management, 36(6), 779-808, Elisver.

Sparck Jones, K., Walker, S., & Robertson, S. E. (2000b). A probabilistic model of information retrieval: development and comparative experiments: Part 2. Information Processing and Management, 36(6), 809-840, Elisver.

van Rijsbergen, C. J. (1979). Information Retrieval. (2nd edition) Butterworths.

Wechsler, M., Sheridan, P., & Schäuble, P. (1997). Experiments in Multilingual Information Retrieval using the SPIDER System. In Proceedings of the 5th RIAO Conference, Computer-Assisted Information Searching on the Internet, Montreal.

Westerveld, T. & de Vries, A. P. (2004). Multimedia Retrieval Using Multiple Examples. In Proceeding of the Third International Conference on Imag and Video Retrieval, pp. 344-352, Spinger.

White, R. W., Oard, D. W., Jones, G. J. F., Soergel, D., and Huang,X (2006): Overview of the CLEF-2005 Cross-Language Speech Retrieval Track, Proceedings of the CLEF 2005: Workshop on Cross-Language Information Retrieval and Evaluation, Vienna, Austria, pp. 744-759, Springer.

Wilkins, P., Ferguson, P. & Smeaton, A.F. (2006) Using Score Distributions for Querytime Fusion in Multimedia Retrieval}, Proceedings of MIR 2006 - 8th ACM SIGMM International Workshop on Multimedia Information Retrieval}, Santa Barbara, CA, ACM.

8. User Interaction & Interface Design.

by Paul Clough with contributions from Jennifer Marlow and James Carmichael

"Each new piece of information [users] encounter gives them new ideas and directions to follow, and, consequently, a new conception of the query."
Bates' berry-picking model for information-seeking [Bates, 1989].

The interface acts as the intermediary between users of information retrieval (IR) systems and the search system itself. In designing an interface for an IR system, the goal is to enable users to satisfy an information need without the assistance of a human intermediary [Brajnik et al., 1996]. A well-designed interface should assist users in clarifying their information needs, and subsequently help them formulate suitable queries and understand the results [Hearst, 1999; Shneiderman, 1997]. More recently, attention has been paid to human-computer interaction in information retrieval and interface design has been driven by the needs of end users, their information-seeking behaviour and psychological aspects of the users, see, e.g. [Ingwersen & Järvelin 2005; Marchionini, 1992; Bates, 1989].

Belkin [2003] points out certain important aspects of functionality in information system design and in particular identifies two issues required to support users with information seeking tasks: (1) designing systems that support a variety of interactions and (2) personalizing the support for user interaction. The former suggests that systems should be designed with a holistic view of information seeking, e.g. adding a workspace to store items *between* individual searches and providing multiple functionalities. The latter recognises that aspects of search, such as a preferred ranking of documents, can be inferred from prior interactions between the user and the system.

Current interface design is linked strongly with research in Interactive Information Retrieval (IIR) that provides the necessary theories and frameworks for modeling user behaviour. Although a little dated in terms of describing the current state of the art, Hearst [1999] still provides an excellent general overview of user interfaces and interaction design for information retrieval systems.

8.1 Information Seeking and General Search Interfaces

With regards to search engine interfaces, it has been said that "Nearly every Web search engine offers users the identical search experience, regardless of the task they are trying to accomplish" [Rose, 2006: 797]. In order to create a more tailored and flexible search experience, users' needs and goals should be taken into consideration, in order to determine not only what users are searching for, but also why they are searching [Rose & Levinson, 2004].

People have different information needs and they make use of various information seeking strategies to solve those problems. For example, Broder [2002] analysed a large collection of queries from a search engine log and found at least three types of information need: *navigational* (find the URL of a specific web site, e.g. "BBC"), *informational* (find some information) and *transactional* (find a structured service to initiate further interaction). Rose and Levinson [2004] refined this work to create a hierarchy of users' goals. Henniger and Belkin [1996] describe analysing the process of satisfying information needs as a decision-making problem in which users learn and refine their needs as they interact with a repository.

Analysing the behaviour of users as they search for information provides informative and valuable insight into user interface design. For example, Gremett [2006] showed how an analysis of users shopping on Amazon.com revealed that in practice users would commonly mix searching and browsing while buying online products. Marchionini [1995] calls this a *mixed behaviour* strategy of information seeking in which a user searches for information by both navigational browsing and searching a site via some explicit search tool such as a search box.

In modern IR research, more emphasis is being placed on constructing models of the search process which go beyond a simplistic view of search as a one-shot matching function between the user's query and collection of documents. Bates [1989] describes search as an interactive process that evolves in response to the information found: results from a search are not just documents, but also the knowledge accumulated along the way. Bates identifies different strategies that people follow during search including following

relationships between documents (e.g. hyperlinks) or browsing over the structure of a collection. She suggests that IR interfaces would be more useful if these search strategies were supported at a higher level. Therefore, both search and browse functionalities should be present and tightly integrated, in order not to interrupt a user's exploration [Beale, 2006; Hearst et al., 2002].

Rose [2006] suggests there are three general areas in which knowledge of information seeking behaviour could inform the design of the user interface for Web search: (1) the goal of the user when conducting a search, (2) the cultural and situational relevance, and (3) the iterative nature of the search task itself. Recognising that users perform different tasks and understanding the user's goals would enable appropriate support mechanisms to be included in the interface design.

Users have different information needs, e.g. getting a specific piece of information, getting an answer to a question, getting advice and exploring a general topic. Modelling user's behaviour would enable provision of the most suitable support rather than creating a one-fits-all interface. Recognising the search context is also important as the same query may have different meanings in different cultures or sub-communities (e.g. a user searching with the query "Madonna and baby" could have in mind the pop star if a music fan, or the painting if an art historian). Different results may also be relevant to the same user at different times. Interfaces offering localisation (e.g. ranking documents with country-specific URLs higher) could help support this.

Bates [1989] suggests that search is best modelled as an iterative process and that retrieval forms part of a dialogue between the user and system to gradually refine the results. Interface support for iteration could include relevance feedback in image retrieval, or lists of related query terms for query expansion or reformulation in text searching. Rose summarises by suggesting that user interfaces should provide different interfaces or forms of interaction to meet users' search goals, allow the user to select appropriate contexts for the search (e.g. language, search options, preferences), and support the iterative nature of the search task by inviting iteration and exploration. Hearst [1999] notes that often when searching or browsing, individuals may become distracted and temporarily follow alternate paths. For this reason, it is recommended to provide ways of recording past queries and offering a means of storing intermediate results throughout the search. This also helps to reduce short-term memory load [Shneiderman et al., 1997; Hearst et al., 2002].

White et al. [2006] also advocate the development of systems to support users who are engaged in exploratory search activities (i.e. those without a pre-defined or specific search task). Henninger and Belkin [1996] review current systems in terms of the key interface and interaction techniques such as querying, browsing and relevance feedback (to support the iterative refinement of the user's information need). They also advocate the use of task modelling and interaction modelling as key strategies to improve the design of retrieval systems. Hearst et al. [2002] cite common search problems such as receiving empty results sets or disorganised result lists, and having difficulty forming special-syntax (Boolean) queries. Therefore, useful means of combating these problems can include providing suggestions for improving the query (if no results have been returned,) showing keywords in context, and giving brief search hints.

Regarding principles for future design interfaces, Rose [2006] advocates making different interfaces available to match different search goals. Another area to investigate is how to improve the browsing process, particularly because the common practice of displaying category lists takes up large amounts of space and often requires a user to guess which category heading will contain the related information of interest [Hearst, 1999].

Although related to Web search, the suggestions from Rose [2006] match existing best practices in designing interfaces to support information seeking. Resnick and Vaughn [2006] describe a set of best practices developed to assist in the design of search interfaces, these design principles are organised into five domains: the corpus, search algorithms, user and task context, the search interface and mobility. Best practices include the use of faceted metadata [Hearst et al., 2002] within a controlled corpus, the use of spell-checking during user input, hybrid navigational support through combined search and browse, the use of past queries to frame the search context, the provision of a large query box (also confirmed by Belkin et al [2000] for more expressive queries), the organisation of a large set of search results into categories, showing the keywords in context in search results and designing alternate versions of content specifically for mobile and handheld devices.

In summary, the emphasis on modern search engine interface design is on understanding and modelling the user's needs, identifying functionalities to support those needs and implementing systems which support the dynamic nature of the user's tasks and searching activities.

8.2 Multilingual Information Access (MLIA)

There are multiple sides to providing multilingual information access (MLIA) and supporting interaction with users. These can range from adapting existing information for use by local communities to providing cross-language search. Current research is focused on aspects such as the design and usability of websites [Del Galdo & Nielsen, 1996; Yunker, 2003] and the provision of multilingual search functionalities [Oard, 1997].

8.2.1 Localisation (and Multilingual Interfaces)

On the Internet, adapting websites to meet the linguistic and cultural needs of the local communities they target is referred to as *globalisation*. The different versions are known as *localised* websites and often require specific design considerations (W3C, 2003; Eurescom, 2000; Del Galdo & Nielsen, 1996; De Troyer & Casteleyn, 2004]. These might include: identifying which languages a website should be translated into, an awareness of cultural issues (e.g. the use of specific terminology or offensive references), the availability of resources (e.g. manpower, translation tools), technical and maintenance issues, how to measure success and issues surrounding design. The W3C (2003) differentiate between *international* and *multilingual* websites: the former being defined as a website which is intended for an international audience while the latter is a website which uses more than one language. According to this definition, a multilingual site is also concerned with regional and cultural differences in addition to language. International sites are often multilingual, e.g. a global company with information presented in different languages.

Multilingual versions of a website (or search engine) may also exhibit different degrees of parallelism, ranging from a collection of monolingual sites at one extreme to a completely parallel site with identical structure, navigation and content at the other (Eurescom, 2000). Typically a trade-off must be made between the cost and effort involved in creating such a site and its benefit. Further issues to consider include:

(i) The use of static versus dynamic content and whether off-line processing can be used to generate multilingual content.

(ii) Query translation, in particular the advantages/disadvantages of using automatic as opposed to manual translation techniques. For example, digital libraries traditionally provide multilingual support via the use of multilingual thesauri such as Eurovoc[65], but this prohibits the use of free-text search and thereby limits interactivity.

8.2.2 Cross-Language Information Retrieval (CLIR)

An area of multilingual retrieval is Cross-Language Information Retrieval (CLIR) in which documents in different languages are searched by queries, also in different languages. This involves translating the query (in the *source language*) into the language of the document collection (*target language*), the documents into the query language or translating both queries and documents into a common language. Three major approaches for CLIR have emerged: (1) automatic machine translation where queries are translated into the target language, (2) the use of machine readable bilingual dictionaries, and (3) the use of corpora to train or enable cross-language retrieval [Voorhees and Harman, 2000].

It is widely recognised that the design of an effective user interface is crucial for the successful implementation of any information system, particularly a search engine [Hearst, 1999; White and Ruthven, 2006]. Understanding the users, their searching behaviour, their needs, search tasks, situational context and their interaction strategies (among other factors) are all important elements of creating effective search applications (see, e.g. Ingwersen & Järvelin, 2005; Marchionini, 1992].

Providing effective access to multilingual document collections undoubtedly involves further challenges for the designers of interactive retrieval systems. In particular, deciding how best to support interaction within the search process can involve enabling: *query formulation* (e.g. offering the user additional query terms to refine their search such as synonyms), *query translation* (e.g. enabling the user to select from multiple query

[65] http://europa.eu/eurovoc/

translations such as different word senses), *document selection* from search results (e.g. providing useable summaries for users to make informed decisions) and *document examination* (e.g. providing translated versions of documents for use by the end users) [Oard, 1997; He et al., 2006; Petrelli et al. 2006].

Practically, the interface may also enable users to indicate terms which should not be translated, identify phrases and signal out-of-vocabulary (OOV) terms (e.g. the CLARITY system [Petrelli et al., 2004; ibid. 2006]. Various studies analysing user interaction have highlighted the importance of the end user's multilingual ability. For example, Petrelli et al. [2002; ibid. 2006] consider users with competence in multiple languages (*polyglots*); whereas others such as Oard and Gonzalo [2002] and Ogden et al. [1999] consider users with no (or limited) knowledge of the target language (*monoglots*). This distinction between users alters the degree of multilingual support required in the search process (e.g. monoglot users may require the translation of retrieved documents or the back-translation of translated query terms).

The study of interactivity in CLIR ranges from studying aspects of the search process such as document selection [Oard et al., 2004; Resnik, 1997], query translation [Wang and Oard, 2001], presentation of search results [Ogden et al., 1999; Petrelli and Clough, 2005]; to the entire search process [e.g. Petrelli et al., 2002; Petrelli et al., 2005; Ogden et al., 1999; Ogden and Davies, 2000; Capstick et al., 2000; Peñas et al., 2001]. Example cross-language search systems (and interfaces) include the following: Keizai, ARCTOS, MULINEX, WTB, MIRACLE and CLARITY.

The Keizai system[66] [Ogden et al., 1999] uses a combination of automatic and user-assisted methods to build and refine cross-language queries. Queries composed of terms in multiple languages can be constructed. The user selects terms to be used in the search from a list of all possible senses of all possible translations. The result is displayed in the source language as a list of one-line summaries plus colour-coded keywords (the original word in Korean or Japanese is displayed in brackets). The Keizai system searches the Web to find documents in Japanese or Korean to answer a question in English. If the user decides to examine a document, they are able to translate the text into English using a link to an on-line MT system (Babelfish). In ARCTOS[67] [Ogden & Davis, 2000], each search term issued by the user is translated and boxed with the group of similar forms. Users can deselect translations, add new forms, or type new translations before the query is actually issued. Documents retrieved (in English, German, French and Italian) are displayed in a manner similar to Keizai.

MULINEX[68] [Capstick et al., 2000] allows users to choose the type of interface to work with: to either see all the translated query terms before proceeding with the search, or to completely hide the translation step. In Keizai and ARCTOS, when the query translation is shown, the user can edit the list and decide which terms will be included and which will not. MULINEX is multi-language (German, English, and French) and a separate column of translations is provided for each language. It also suggests a list of additional terms the user might decide to include in the query. The retrieved documents are displayed as a list; for each document a set of category words in the user language and a summary in the document language are displayed. The user can click for a summary or the full-text translation in another language.

WTB (Web site Term Browser; [Peñas, Gonzalo, & Verdejo, 2001]) shows the terms generated during the query-expansion step grouped as families of terms, e.g. synonyms, hyponyms and hypernyms. Search results are presented as a cluster of documents grouped by relevant phrases. The system makes use of phrasal information to process queries, and suggest relevant topics. By clicking on a line the user can explore the set of homogeneous documents represented by their title and an extensive set of relevant terms.

[66] http://kythera.nmsu.edu:8099
[67] http://crl.nmsu.edu/~ogden/i-clir/cltr-interactive/arctos/page1.html
[68] http://mulinex.dfki.de/demo.html

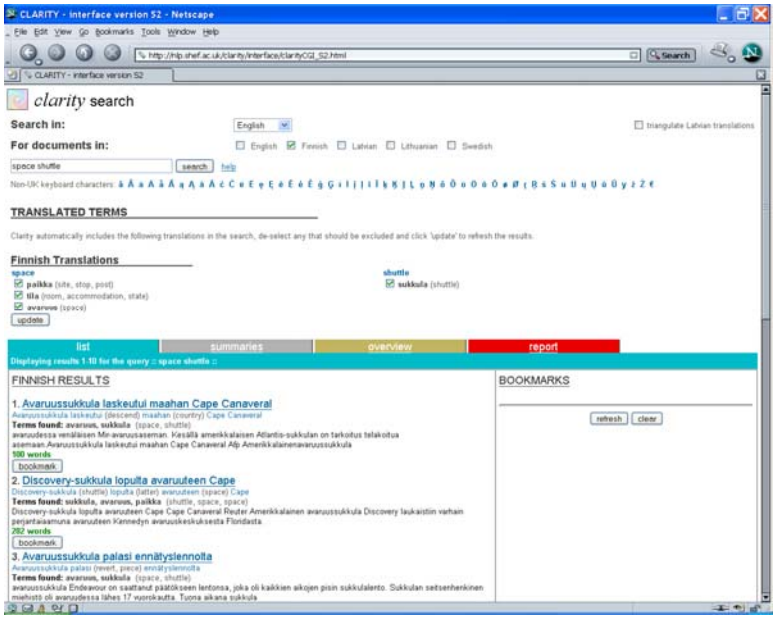


Figure 8.1: CLARITY user interface for CLIR

MIRACLE [Dorr et al., 2003; He et al., 2003] is a user-assisted CLIR system that groups translations for each query term in a tab and allows the user to view synonyms and examples of use. The list of terms actually used in the query is displayed below, followed by the list of retrieved documents for which the first two lines of machine-translated text are displayed. MIRACLE was designed with two aspects in mind: (1) a clear exposure to the user of the interaction design and (2) immediate feedback in response to user actions. Participation in the Cross Language Evaluation Forum (CLEF) interactive track (iCLEF) track has shown some interesting search behaviours from users such as adopting terms from relevant documents during query refinement (thereby confirming the need for document translations and consistency of translation resources used) and different strategies for query formulation [He et al., 2006].

CLARITY [Petrelli et al, 2005] has two interfaces: one to allow the users to modify the translation (*supervised mode*) and another interface (*delegated mode*). Using the delegated mode, the user simply enters the query, clicks the "Search" button and the results are then displayed. There is no user intervention during the query translation process. To modify the query, the user must re-enter it in the box. This system translates the queries into English, Finnish and Swedish. Figure 8.1 shows an example of the CLARITY interface (an English query searching Finnish documents).

Perhaps some of the most significant research undertaken to study the interaction with cross-language retrieval systems has been within iCLEF [Gonzalo & Oard, 2002].In 2000 iCLEF showed that users could determine the topic of retrieved documents, and could often formulate effective queries (2002 and 2003), that users could find answers to factual questions (2004), find historical images (2005), and most recently that users are able to perform multilingual searches using Flickr, the online photo management tool (2006; 2008).

More recently, Oard et al. [2008] reported on the results of several studies examining the user-assisted query translation process in the context of cross-language search. When using a system that enabled altering a translation by de- or re-selecting alternatives, participants utilized this function 23% of the time. However, it was unclear if this was done because it was seen as a helpful feature or because people were eager to experiment with the new technology. This study also reinforced the idea that people preferred to view the search results before altering any machine translations. Thus, recommendations were made for design based on this sort of progressive refinement. Marlow et al. [2008] have furthered user-orientated studies in CLIR by exploring the effects of language skills on cross-language search. Using the Google Translate service, the authors showed that users have varied language skills that are non-trivial to assess and can impact their multilingual searching experience and search effectiveness.

Table 8.1: Functionality offered by various online museums and art galleries [Marlow, 2006]

	Welcome pgs for f.l. (if more than one page available)	Multilingual. search of site (can locate material written in other languages)	CLIR (query translation)	Controlled vocabulary	Free text Search	Easy to switch languages	Easy to return to original language
Tate Online		●	○	◉	◉	○	○
British Museum	●	●	○	◉	◉	●	○
National Gallery		○	○	◉	◉	○	●
V&A Museum		●	○	○	◉	●	●
Natl. Portrait Gallery		●	○	◉	◉	○	●
Louvre	●	●	◉ Lafayette database ○ Atlas database	● - Kaleidoscope	●	●	●
Guggenheim Bilbao	●	○	○	○	○	○	○
van Gogh Museum	●	●	○	●	○	◉	●
Rijks-museum		○	◉	●	●	○	○
Centre Pompidou			●	◉	◉	●	
MoMA		●	○	○	◉		
Met New York		●	○	○	◉	●	●
Guggenheim New York		○	○	◉	◉		
24 Hr Museum	◉	◉	○	○	●	○	●
Easyart.com	●	○	●	●	●	○	

● - multilingual offering ◉ - only in main language ○ - not offered

8.2.3 Implementation of Multilingual Information Access

The Minerva survey [2006] examined the types of monolingual search functionalities provided by 671 European cultural and museum websites. Overall, it was reported that 51% of sites used no search tool at all, 24% offered free text indexing and 14% provided controlled vocabularies (some sites offered both). However, it is unclear how many of these search tools were available in more than one language. Marlow [2006] reviewed the functionality of a number of online museums and art galleries (shown in Table 8.1).

It is noteworthy that very few of the sites listed in Table 8.1 actually offer cross-language search functionality to users. This is typical of what generally obtains for most Internet search engines which tend to lack multilingual search facilities. The majority of cross-language research remains in the theoretical domain and has not often been implemented or made accessible to the end user [Peters and Sheridan, 2001]. Perhaps

this is surprising given the motivation for multilingual search in [Oard, 1997], but Evans [2006] indicates that factors such as the limited effectiveness of translation, the lack of real-world user need for this kind of functionality, the complexity in effectively providing multilingual interaction and the additional cognitive burden pressed upon the user are all limiting factors.

8.3 Multimedia Information Access

Multimedia information retrieval (MIR) systems are designed to enable the searching of data in various modalities such as text, image, video and sound. Chu [2006] defines a taxonomy of multimedia information as shown in Figure 8.2, highlighting that multimedia information can be a combination of any single media. There are multiple ways of accessing visual objects (image and video) depending upon the information associated with the object: either information *about* the object (*metadata*) or information contained *within* the object (*audiovisual features*).

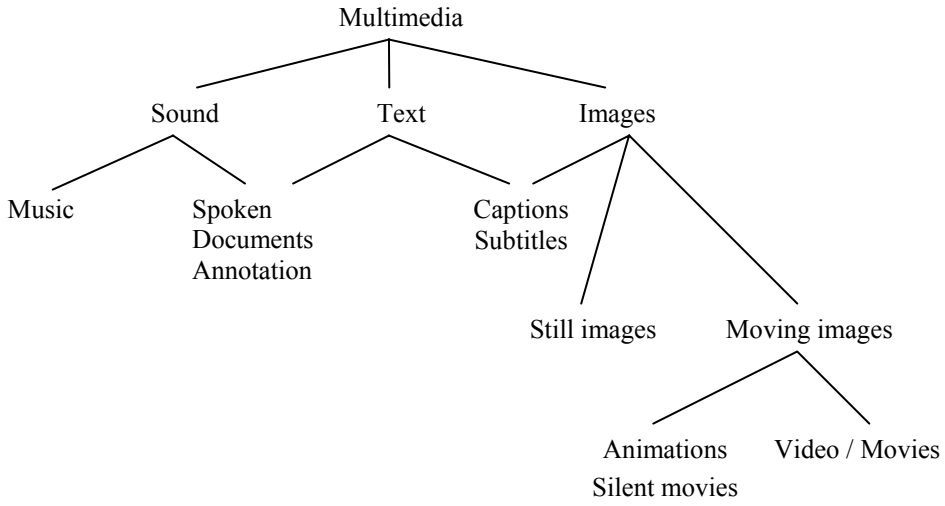


Figure 8.2: A taxonomy of multimedia information [Chu, 2006:43].

Images and video objects exhibit similar visual properties, the main difference being the additional spatio-temporal aspects of video [Gupta & Jain, 1997]. There is currently much research on combining both visual features and metadata as complementary evidence for both image and video retrieval and this is seen as one of the main research areas in current image retrieval research [Enser, 2004]. Further areas of research include both technical issues and establishing the requirements of users for multimedia information access. Ultimately the design of the interface and provision of functionality will depend on the needs of the end users, the indexing methods in use and available audiovisual data. We will start by discussing access to visual information (still images in section 8.3.1 and moving images or video in section 8.3.2). In section 8.3.3 we discuss access to audio information.

8.3.1 Still Image Retrieval

As with the design of any information system, an important part of the process is to establish what type of users will be using the system and their associated needs. For example, in describing image retrieval, Goodrum [2000] suggests that user interfaces must be influenced by considering the users' needs and their typical search tasks.

To date, most of the research and development in image retrieval has focused on providing functionality rather than giving sufficient attention to the needs of the end user [Eakins et al., 2004]. This has resulted in the design of interfaces which are inadequate (or unusable) for end users [Venters et al., 1997]. For example, a large body of research has grown up around developing algorithms to facilitate content-based retrieval [e.g. Smeulders et al., 2000]. However, studies of user needs have shown users' needs to be both linguistically and visually-orientated [Enser, 1995]. In practice, however, investigations (in particular domains) have shown that provision of text-based access is not just preferable but vital to many end users [Eakins et al., 2004; Markkula & Sormunen, 2000]. There are two main strategies for image retrieval:

(1) **description-based** (includes *text-based* and *concept-based*), which uses assigned free-text or terms from a controlled vocabulary (see, e.g. [Goodrum, 2000; Gupta & Jain, 1997; Rui et al., 1997; Smeulders et al., 2000; Veltkamp & Tanase, 2000].

(2) **content-based**, which makes use of low-level features derived from the visual content of an image Content-based retrieval [Smeulders et al., 2000] relies on indexing images by low-level attributes such as colour, shape and texture.

Since digitised images purely consist of arrays of pixel intensities with no inherent meaning, one of the key issues with CBIR and other image processing is to extract useful information from the raw data [Eakins and Graham, 1999]. By studying users' image retrieval requirements and the types of attributes images may exhibit, Eakins [1998] proposed a 3-level framework for image retrieval, classifying image queries by increasing complexity:

- **Level 1** comprises retrieval by primitive features such as colour, texture, shape or the spatial location of image elements. This level of retrieval uses features which directly extract from the images themselves, without the need to refer to any external knowledge base.
- **Level 2** comprises retrieval by derived features, involving some degree of logical inference about the identity of the objects depicted in the image. This requires reference some outside knowledge but in practice this level of query is very generally encountered (e.g. retrieval of objects of a given class such as "pictures of a passenger train on a bridge"; retrieval of individual objects or persons such as "pictures of Tony Blair" or "pictures of Nelson's Column").
- **Level 3** comprises retrieval by abstract attributes. This involves a large amount of high-level reasoning about the meaning and purpose of the objects depicted in the images. This level of query often requires some sophistication of the searcher and the reasoning judgment is often subjective. It would also require retrieval technique of level 2 to get the semantic meaning of various objects. For example, the retrieval of named events or types of activity "pictures of English folk dancing"; or retrieval of pictures with emotional or symbolic significance "pictures depicting *death*."

Description-Based Image Retrieval

Traditionally, the main approach for accessing images was based on formulating and serving text-based queries. Many of the early image retrieval systems were concept- (or text-) based utilising bespoke indexing schemes [Rasmussen, 1998] and overlapped substantially with the areas of databases and information science. Still images have unique meaning and properties that provide the basis for retrieval by users. For example, on considering the meaning of pictorial images, Panofsky [1955] categorised fine art images based on the "who, what, where and when" search paradigm and by the modes: iconography (specific requests), pre-iconography (general requests), and iconology (abstract images). Iconography describes a picture's actual subject matter (the what); iconology describes its deeper artistic or religious meaning (the why). Other authors such as Eakins and Graham [1999] have also discussed the categorisation of image attributes into various levels or strata. Pictures can therefore be described by their physical attributes (e.g. a picture of a dodo) and/or attributes of their subject (e.g. a picture of an extinct bird).

The main approach for accessing images is based on formulating and serving text-based queries which match between a user's query and image description. Rasmussen [1997] refers to descriptions of subject attributes as concept-based and Goodrum [2000] refers to descriptions based on texts associated with the images (e.g. captions, web pages) as text-based. There are many instances when images are associated with some kind of text semantically related to the image (e.g. metadata or captions); examples include collections such as historic or stock-photographic archives, medical databases, art/history collections, personal photographs (e.g. Flickr.com) and the Web (e.g. Yahoo! Images and AllTheWeb.com). Other attributes typically associated with an image which can be searched include date, time and information derived from the photographic equipment itself (e.g. the Exif[69] data provided by modern digital cameras).

Retrieval of images based on descriptions is typically through keywords (mostly derived from textual information accompanying an image) and controlled vocabularies associated with subject attributes.

[69] Exchangeable image file format is a specification for the image file format used by digital cameras: http://en.wikipedia.org/wiki/EXIF

Searching with free-text (most keyword searches enable users to perform free-text search) or controlled vocabularies has shown to be an effective method of searching image repositories [Markkula & Sormunen, 2000; Rorvig: 1988]. Often, manually assigning indexing terms is a difficult task. The main problem is that the intrinsic meaning of an image is difficult to interpret and express in written form [Jorgensen, 1998]. In addition, assigning keywords to images is a very subjective task and suffers from low index term agreement across indexers and between indexers and user queries [Enser and McGregor, 1993]. Further, the amount (and availability) of visual material is growing at an astronomical rate and manual annotation is therefore impossible and in cases such as personal image collections, people often don't bother to annotate images. This has led to the popularity of approaches based on the automatic assignment of textual attributes [Turner, 1994]

Controlled vocabularies for text-based indexing can be found in the literature which describes the concepts of using certain established thesauri to describe image, e.g. Art & Architecture Thesaurus [AAT] [Petersen & Barnett, 1994]; Thesaurus for Graphic Materials [Parker, 1987] and ICONCLASS. They have applied existing cataloguing systems like Dewey Decimal System to describe images. See [Rasmussen, 1997] for further details of controlled vocabularies. An interesting extension of a controlled vocabulary is the *visual thesauri* which uses visual surrogates to represent concepts in addition to verbal descriptions (see, e.g. [Mostafa, 1994; Rasmussen, 1997]. This offers potentially interesting ways of using a controlled vocabulary (e.g. using the visual surrogates in a query-by-visual-example paradigm and using the pictures to create a language-independent representation of the controlled vocabulary). A summary of text-based retrieval products can be found in [Eakins et al, 1999], and previous research and prototypes described in [Clough and Sanderson, 2006].

Content -Based Image Retrieval (CBIR)

In the early 1990s, because of the emergence of large-scale image collections and the aforementioned difficulties with manually indexing images, the development of content-based image retrieval (CBIR) was proposed by information researchers and scientists [Rui et al, 1999]. Content-based retrieval is implemented by automatically processing image attributes which are specified in user's queries. Typical image attributes include colour, shape, texture and spatial layout, all features which can be extracted using low-level feature extraction.

Retrieval based on colour similarity is often achieved by using a colour histogram for each image that identifies the distribution of colour pixels in an image. Image retrieval based on texture similarity is not regarded as very useful. However, the ability to match on texture similarity is often used most successfully when distinguishing between areas with similar colour in an image, e.g. between sky and sea [Eakins, 2000]. Queries by shapes are often achieved by selecting an example image provided by the system or by asking the user to sketch a rough shape. The primary mechanisms used for shape retrieval include "identification of features such as lines, boundaries, aspect ratio, circularity, and region and edge detection." [Goodrum, 2000]

Gudivada and Raghavan [1995] regard image retrieval at levels 2 and 3 of Eakin's framework as semantic image retrieval because they involve the addition of semantic information (typically by people). Most current CBIR techniques are designed for primitive levels (level 1), while some have attempted to tackle level 2 retrieval. However, this poses two non-trivial problems. The first is scene recognition: it is important to identify the type of scene presented in an image since this constitutes an important filter that can offer critical clues helping to recognise specific objects in an image. Object recognition is in itself a challenging problem in the area of computer vision. For example, Forsyth et al [1997] developed a technique for recognising naked people within images.

A number of general-purpose CBIR systems are commercially available on the Internet and most of these image retrieval systems support one or more of the following options: random browsing of images from the database, search by visual example, search by sketch, search by text and navigation with customised image categories [Chang et al, 1998]. Example content-based systems (both academic and commercial) include Virage's VIR Image Engine (VIR), Query By Visual Content (QBIC), VisualSEEk and Exacalibur's Image RetrievalWare. Web-based systems include WebSEEK, Informedia, Photobook and Alta Vista Photofinder. A full review of CBIR systems can be found in Veltkamp & Tanase [2000]. Most commercial and academic CBIR systems tend to offer either query-by-example functionality or support for user-input visual exemplars (e.g. colour).

One of the most cited examples of a commercial CBIR system is IBM's Query By Image Content or QBIC [Flickner et al., 1995]. It offers retrieval by combination of colour, texture or shape. Image queries can be formulated by selecting colour from a palette, sketching a rough shape of desired image, or specifying an example query image. The system extracts and stores the colour, shape and texture features from each image in its database, calculates similarity between query and stored images then displays the most similar image as thumbnails. In the cultural heritage domain, it can be used for colour and layout search in the State Hermitage Museum digital collection.70

WebSeek[71] [Smith et al, 1997], which was developed by Columbia University, is another content-based image retrieval system making keyword and colour based queries through a catalogue of images collected from the Web. The system allows the user to submit a query by choosing a subject from the available catalogue or entering a text topic. The results of the query may be used for another colour query in the whole catalogue or for sorting the results by decreasing colour similarity to the selected image. In addition, WebSeek allows the user to directly define a colour histogram's attributes in order to better refine the image search criteria.

WebSeer [Swain et al., 1996] was developed by the department of computer science at the University of Chicago as an experimental system. Besides some common characteristics such as specifying image dimensions, file size, image type and submitting keywords describing the contents of the desired images, the system was also able to detect human faces based on a neural network. If the user is looking for people, he/she must indicate the number of faces as well as the size of the portrait. Face detection is believed to meet the needs of the Level 2 user.

Most existing CBIR systems retrieve images by image appearance, using automatic extraction and a comparison of image features such as colour, texture, shape and spatial layout. This well meets Level 1 of user's image query needs. However, for Level 2 and Level 3, evidence suggests that such a facility is actually of limited use in meeting image users' real needs [Eakins et al., 2004]. First, it is impossible to start a search if no suitable query image can be found or the user has no idea about what the image should look like, e.g. searching for a rare unseen animal. Second, users may find it difficult to manipulate search parameters such as the relative importance of colour, shape or texture because such visual features are not as intuitive as text [Eakins et al., 2004].

A large number of CBIR systems take sophisticated algorithms; however, it is not clear whether they can really address user needs. As a result, to narrow this semantic gap, a powerful and user-friendly query interface is needed where users can interact with systems by providing his or her evaluation or preference of a current retrieval result to the IR system [Rui et al, 1999].

Combining approaches

Combining both description and content-based approaches is likely to be more effective than any single method alone. Eakins and Graham [1999] comment that the use of keywords and image features in combination is desirable. This coincides with best practice in designing interactive retrieval systems which suggest that a variety of interaction approaches should be offered to meet the varying needs of users and their work tasks. Chu [2001] provides examples of research from the content-based community which has combined the two approaches. The current challenge is how best to integrate functionality to provide natural access for users to both low-level primitive features and high-level semantics. Systems such as WebSEEk [Chang et al.,. 1997] have shown the benefits of combining approaches (e.g. allowing users to initiate a search based on keywords or selecting terms from a controlled vocabulary, and then using content-based approaches during refinement or to provide a "more like this" function).

User interfaces and interaction

Interaction with image retrieval systems is similar to any other retrieval system and includes: query formulation, query reformulation/modification (e.g. through relevance feedback), browsing-searching and results presentation (in context). Typically in image retrieval systems, the user interface consists of a query formulation part and results presentation part [Veltkamp & Tanase, 2000:1]. Users can select images from

[70] http://www.hermitagemuseum.org/fcgibin/db2www/qbicSearch.mac/qbic?selLang=English

[71] http://persia.ee.columbia.edu:8008

the index (or database) by browsing one-by-one, or specify an image (or set of images) through the use of keywords, by using visual properties of an image (e.g. colour, texture etc.), or providing a visual exemplar (e.g. an example image or a sketch).

Various studies have been undertaken to establish what people search for in multimedia collections, e.g. newspaper image archives, picture archives and museums [Enser 1995; Enser & McGregor 1992; Armitage & Enser 1997]. Enser and McGregor [1992] categorised queries made to a large picture archive into those which could be satisfied by a picture of a unique person, object or event (e.g. Kenilworth Castle, Sergei Prokofiev, HMS Volunteer, Alan Turing), and those which could not (e.g. classroom scenes, Clyde cruisers, shopping arcades, air raids). These categories, unique and non-unique, were also subject to query refinement in terms of time, action, event or technical specification. For example a non-unique query such as "carnival" could be modified to create "the Rio Carnival, 1996" (unique), refined by location and time.

A recent study by Eakins et al. [2004] identified user needs within a framework based on a taxonomy of image content (i.e. classifying images from a low-level representation to high-level semantics) and how professionals search for and use image data (e.g. for illustration, learning, information processing and generating ideas). Their findings reinforced previous studies (e.g. [Enser, 1995; Markkula & Sormunen, 2000]) whereby participants were primarily interested in concept-based retrieval rather than content-based. They also found the preferred method of querying was to type search terms rather than select from a hierarchy of terms or query by example. The use of text-based retrieval, however, presupposes that images are associated with textual metadata. In many scenarios this is a valid assumption, e.g. in stock photographic collections, on the Web and historical or cultural heritage archives. However, this is not always the case (e.g. for personal photographic collections).

Researchers have also considered the user's searching behaviour in image retrieval. For example, Cox et al. [1996] define at least three classes of image search: (1) target search – users find specific target images (e.g. art historian finding a specific painting), (2) category search – users seek one or more images from general categories (e.g. "sunsets" or "pictures of the Eiffel Tower"), and (3) open-ended browsing – users have a vague idea of their search needs and may change their mind repeatedly throughout the search. This last category includes exploratory tasks where users have no specific goal (e.g. browsing through a database for fun).

Two fundamental methods for accessing information include search and browse. Search consists of typing keywords; browse is more likely once an initial starting point is found Browsing support is often structured such that content is categorised into predetermined classes or a hierarchy (e.g. subject classification) into which users can further explore and navigate. However, this is typically useful only if it matches the user's expectations because it imposes a single view on a collection (alternatives are multiple alternative hierarchies, e.g. faceted metadata). Accessing information through browsing has demonstrated to be very effective in the domain of image retrieval (see, e.g. [Chang et al., 2004; Shen, 2003; Combs & Bederson, 1999]). When image browsing is combined with text searching, users are able to select their most preferred interaction mode and move between the two in a fluid way (see, e.g. [Hearst, 2002; Yee, 2003; Combs & Bederson, 1999]).

One of the biggest problems with retrieving visual information is the "semantic gap" between the low-levelled data representation (e.g. pixel light intensity values) and high-level needs/concepts that the user desires [Enser and Sandom, 2003]. As Urban and Jose [2005] state, "the images' low-level feature representation does not reflect the high-level concepts the user has in mind." The problem of the semantic gap for information retrieval is that the meaning of an image can only be defined in context. The use of relevance feedback and browsing-searching techniques can assist with formulating the user's query and narrow the semantic gap (i.e. help the user to specify the query).

Query Specification

Queries to CBIR systems are most often expressed as visual exemplars (Query-By-Visual-Example or QBVE) or specifying image attributes such as colour (e.g. picking the desired colour from a palette). QBVE can be performed by supplying an example image being sought (either from within or outside the indexed collection of images), or sketching the desired shape of an example image (e.g. QBIC offers this [Flickner et

al., 1995] and RetrieveR[72], a sketch interface to Flickr). Eakins and Graham [1999] point out those content-based approaches based on colour, texture and shape are capable of delivering useful results, but in practice some of the features are far more useful than others (e.g. colour and texture retrieval often gives better results than shape matching). The advantages of this form of querying are its simplicity for novice users and ease of expressing more "visual" queries in domains where visual attributes are important (e.g. fine-art painting [Lombardi et al., 2004]).

However, this approach has some disadvantages. For example, the success of sketched queries may depend on the user's artistic abilities. Additionally, supplying a single example image may prove quite successful when searching for a single relevant image but will probably be less successful for retrieval of groups of images related to a category. Matching variants of a supplied image can be difficult (e.g. images distorted by rotation, skew and occlusion). A further problem is the semantic gap. Gupta and Jain [1997] state that query specification for visual information should not be limited to query-by-example or the specification of visual properties of images and suggest nine further properties of a query language including: spatial arrangement, temporal arrangement and feature-space manipulation.

Most systems enable the user to evaluate or provide his or her preference of a current retrieval result to a CBIR system (*relevance feedback*) as a way of refining the query. This can be through specifying positive or negative examples, and Rui & Huang [1999] suggest that this can be used to narrow the semantic gap. Rui et al. [1998] suggest that systems involving CBIR must research into *where* in the interaction cycle users would want such support. However, CBIR systems are still not widely used by the general public after more than a decade of research effort. Urban and Jose [2005] suggest this is due to the continuing problem of the semantic gap and the fact that most current interfaces do not provide sufficient querying facilities and appropriate presentation of results.

Browsing

Many efforts have been undertaken to generate effective image indexing systems (e.g. ICONCLASS[73], the Getty Art and Architecture Thesaurus or AAT[74] and WordNet[75]) and these semantic classification systems are often used to complement search and provide browsing functionality. A study of interaction with WebSEEk found that users' preferred method of browsing was through theme-based navigation – rather than browsing through pages of image thumbnails – and preferred querying methods based on some specific subject matter rather than free-text search or advanced visual searches. The use of hierarchical structures for categorising and organising images not only facilitates browsing, but also helps to provide a context for the search results (e.g. users can browse through results in broader or narrower categories). There are several problems with using a controlled vocabulary, however, including the assignment of terms, the ambiguity of categories, and the user's unfamiliarity of subject categories used in the classification scheme (Getty photographic images).

One approach to render QBE more attractive is to use information derived from text associated with the image itself. For example, Yee et al. [2003] describe Flamenco, a text-based image retrieval system in which users are able to drill down results along conceptual dimensions provided by hierarchically faceted metadata. Categories are automatically derived from Wordnet synsets based on texts associated with the images, but assignment of those categories to the images is then manual. This interface provides effective search and browse of images and supports exploratory search tasks. A further approach is to allow the users to generate their own taxonomies in the form of *folksonomies*. The online photo management tool, Flickr, allows this form of collaborative annotation through users assigning tags (keywords) to images. These then enable users to navigate to images with the same tags and a clustering of tags helps to organise images and facilitate browsing.

[72] http://labs.systemone.at/retrievr/
[73] http://www.iconclass.nl/.
[74] http://www.getty.edu/research/conducting_research/vocabularies/aat/
[75] http://wordnet.princeton.edu/

Results presentation/visualisation

Finding appropriate results that correspond to the user's searching and browsing requirements is the first task a system must achieve; however, an equally important consideration involves determining how best to present said results in an accessible and user-friendly manner. For example, Hearst [1999] recommends providing users with information about:

- How retrieved documents are related to the query
- How the retrieved documents relate to each other, and
- How the documents relate to the collection as a whole

Currently, the widely-used standard for displaying image results is to show a two-dimensional grid of thumbnails [Karadkar et al., 2006; Rodden et al., 2001; Combs & Bederson, 1999]. However, this is not necessarily an ideal approach.

Chang & Leggett [2003] outline three main problems with current interfaces for searching and viewing image collections. First, querying by metadata is ambiguous and often does not accurately portray relations between image elements. Secondly, browsing is often time-consuming (involving a great deal of pointing and clicking) and not adaptive to users' needs. Finally, scrolling through many thumbnails is tedious, and if all results do not fit on one page, it is difficult to obtain a comprehensive view or understanding of the entire result set. Janecek & Pu [2004] note that since it is increasingly difficult to display all information in the limited space of one screen, there is often a balance that must be struck between showing a small amount of detailed information and providing a large amount of more abstract information.

Jörgensen & Jörgensen's [2005] study of image professionals revealed that 85.6% of the searches involved the browsing of results, implying that this behaviour is important in making an image selection. Therefore, developing a more effective way of enabling this to be done is the subject of much research. To combat some of the problems stated above, alternative approaches to visualising results displays have been explored.

With regards to the problem of having to scan a large set of results for relevant or related images, Liu et al [2004] developed a similarity-based results presentation that was meant to graphically depict the closeness of relationships between images, based on "regions of interest" within the images. The items were then arranged in a way so that closely related pictures were situated near and overlapped each other. To facilitate viewing, the user could control the overlapping ratio using a slider. Results of initial experimentation indicated that this approach helped to improve users' experience browsing results and sped up the search process.

Janecek & Pu [2004] advocate the use of semantic "fisheye" views to enable focusing in on relevant parts of a wide set of results. This type of visualisation helps users to examine local details while still maintaining a view of the broader context [Liu et al., 2004]. Moving the mouse over a particular element of the results display automatically brings it into greater focus. Thus, that which the user deems to be more interesting or important is emphasised, while the less important information remains in the background. The metrics used to determine "importance" are flexible and can thus be adjusted to enable a variety of search strategies.

Visualisation displays can also encourage query refinement in a variety of ways. For example, users can be given the opportunity to see a range of related items in order to decide if one of them fits their needs more closely. This can be particularly useful in the case where a query has multiple meanings (i.e. the word "Pluto" can refer to the astronomical entity or to the Disney character.) In this case, a clustering method could be helpful.

For image retrieval, clustering methods have been used to organize search results by grouping the top *n* ranked images into similar and dissimilar classes. Typically this is based on visual similarity and the cluster closest to the query or a representative image from each cluster can then be used to present the user with very different images enabling more effective user feedback. For example, Park et al. [2005] took the top 120 images and clustered these using hierarchical agglomerative clustering methods (HACM). Clusters are then ranked based on the distance of the cluster from the query. The effect is to group together visually similar images in the results. However, Rodden et al. [2003] performed usability studies to determine whether organization by visual similarity is actually useful. Interestingly, their results suggest that images organized by category/subject labels were more understandable to users that those grouped by visual features.

Other approaches have combined both visual and textual information to cluster sets of images into multiple topics. For example, Cai et al. [2004] use visual, textual and link information to cluster Web image search results into different types of semantic clusters. Barnard and Forsyth [2001] organize image collections using a statistical model which incorporates both semantic information extracted from associated text and visual data derived from image processing. During a training phase, they train a generative hierarchical model to learn semantic relationships between low-level visual features and words. The resulting hierarchical model associates segments of an image (known as *blobs*) with words and clusters these into groups which can then be used to browse the image collection.

As another form of clustering, Clough et al. [2005] propose automatically generating a set of conceptual hierarchies based on metadata, and then classifying representative images into the relevant place in the hierarchy. The result combines text and visual data and is essentially a hierarchical browsing facility with associated images displayed to illustrate and clarify the terms.

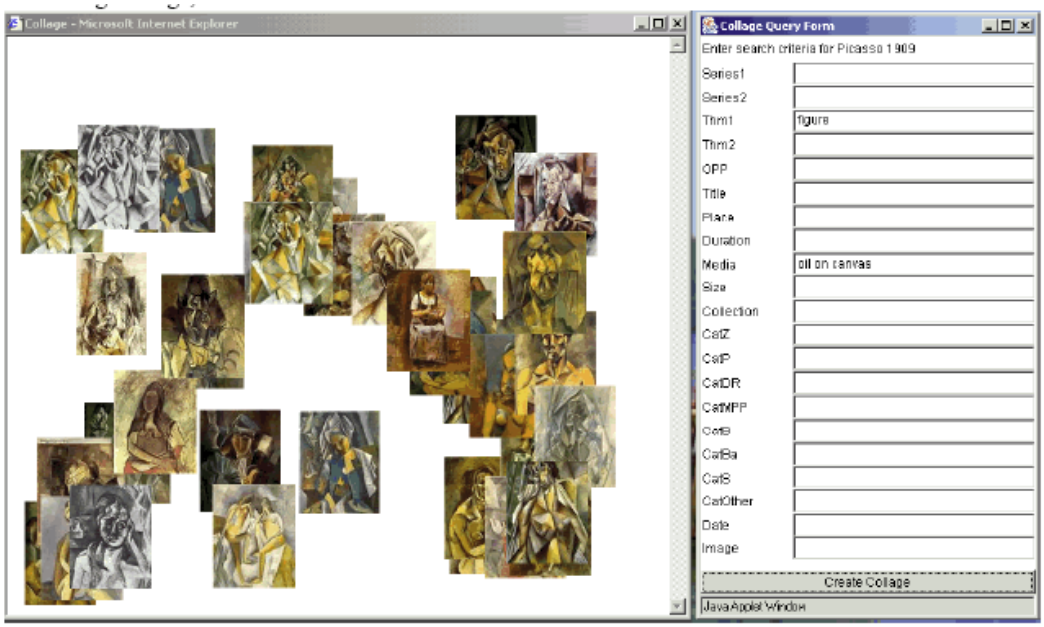


Figure 8.3: Streaming Collage interface [Chang & Leggett, 2003]

Visualising a collection overview can be slightly different from visualising results of a targeted search because rather than trying to locate a specific item, often the goal is to get a general understanding of a collection's underlying theme. To facilitate this, Chang & Leggett [2003] propose a streaming collage approach, whereby a collage of the collection's holdings is gradually and dynamically built over time, with similar items placed near one another in a way that highlights commonalities, links, and relationships. (see Figure 8.3). Another suggestion related to the browsing interface is to employ a zoomable image browser (Figure 8.4) as a way of maximising use of the screen space [Combs & Bederson, 1999].

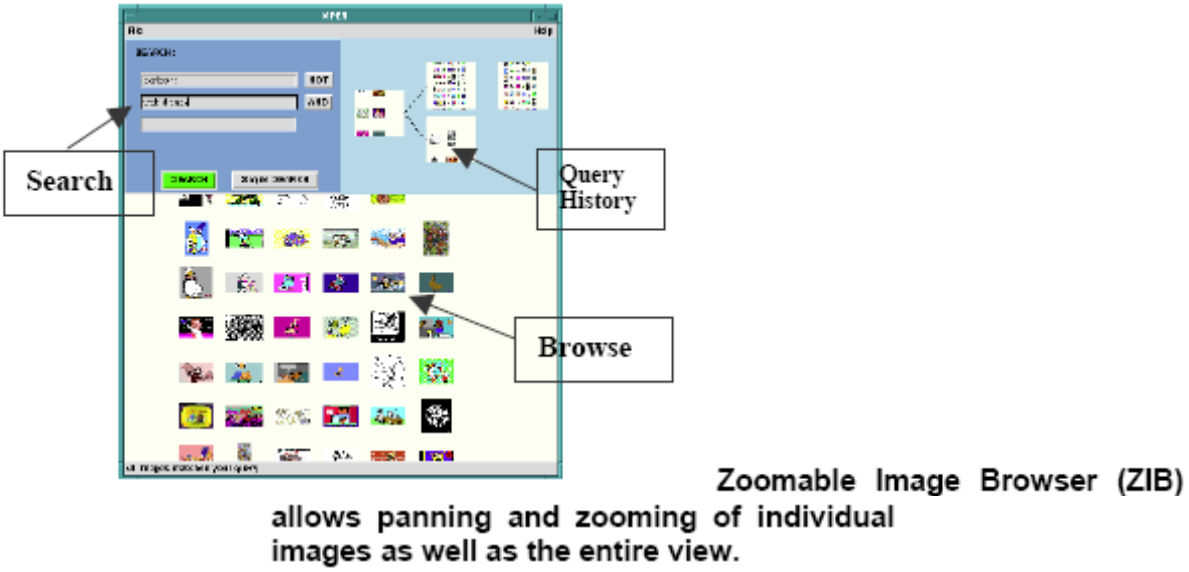


Figure 8.4: Zoomable image browser prototype [Combs & Bederson, 1999]

When retrieval is conducted across media, it is not clear how the results should be displayed. A single list of interleaved or fused heterogeneous multimedia objects to be explored in sequence may not be the best solution. Different metaphors and layouts have been proposed but limitedly to a single media (i.e. newspaper-like layout for text [Golovchinsky, 1997]; comic book [Boreczky, 2000] and storyboard [Christel, 2002] for video; or picture album for images [Kyu, 2004]. Karadkar et al. [2006] investigate various combinations of spatial and temporal layouts and their constraints on context during the design of an interface for a video and image retrieval system.

Summary

In summary, current image retrieval systems offer much functionality, some of which is not necessarily useful to users. It is important to study users, ascertain their needs, and determine their tasks to develop effective user interfaces. Rather than try and meet the needs of all users, it is important to provide functionality to meet specific user classes. For example, Jörgensen & Jörgensen's [2005] study of image professionals noted that these individuals had slightly different behaviours than more general users; these included a reliance on more descriptive and thematic queries than unique term searches.

Goodrum [2000] suggests that research is required that examines interface support for browsing, query formulation and iterative searching. Lee et al. [1994] emphasise that research must be undertaken to establish where in the interaction cycle CBIR would best be suited. Chang et al. [1997] have found with WebSEEk that users prefer to navigate through a clearly defined semantic structure organised in a hierarchical form (especially true for searching large repositories). After users have narrowed down results, the use of content-based methods can then be used to effectively organise, browse and view the content space.

8.3.2 Video Retrieval Interfaces

The process of searching, retrieving, and visualising videos differs from that of images due to the nature and format of video as a medium. For example, video is inherently multimodal and can contain visual, auditory, and textual elements [Snoek & Worring, 2005]. In addition, video is time-based and as a result, searching through clips to locate some information of interest can potentially be a tedious and lengthy process [van Houten et al., 2004]. Therefore, a video retrieval interface should make it easy for users to browse and/or search for relevant material in an efficient way.

Video indexing

Before videos can be searched, browsed, or manipulated, they must be indexed in some way [Snoek & Worring, 2005]. There are a variety of ways in which this can be done. One approach is to break a video clip down into its individual components and index these. The atomic unit of the video *clip* is the *frame* (the equivalent of one exposure on a celluloid film track). A video *shot* is defined as the sequence of frames captured during a single "start recording" and "stop recording" camera operation. A *scene* is a sequential collection of shots unified by a common event or locale. A video clip is normally composed of a collection of scenes. There are several scene combination possibilities, one of which is the *dialog*, defined as a series of alternating shots depicting some form of communication between two or more entities (e.g. the "cut-away" shots switching between the in-studio news anchor and the on-location news reporter). Most video document indexing techniques exploit this inherent frame→shot→scene→clip hierarchical structure to automatically segment the video document into more manageable chunks. Yeo and Yeung [1997] schematically illustrate this hierarchy as shown in Figure 8.5.

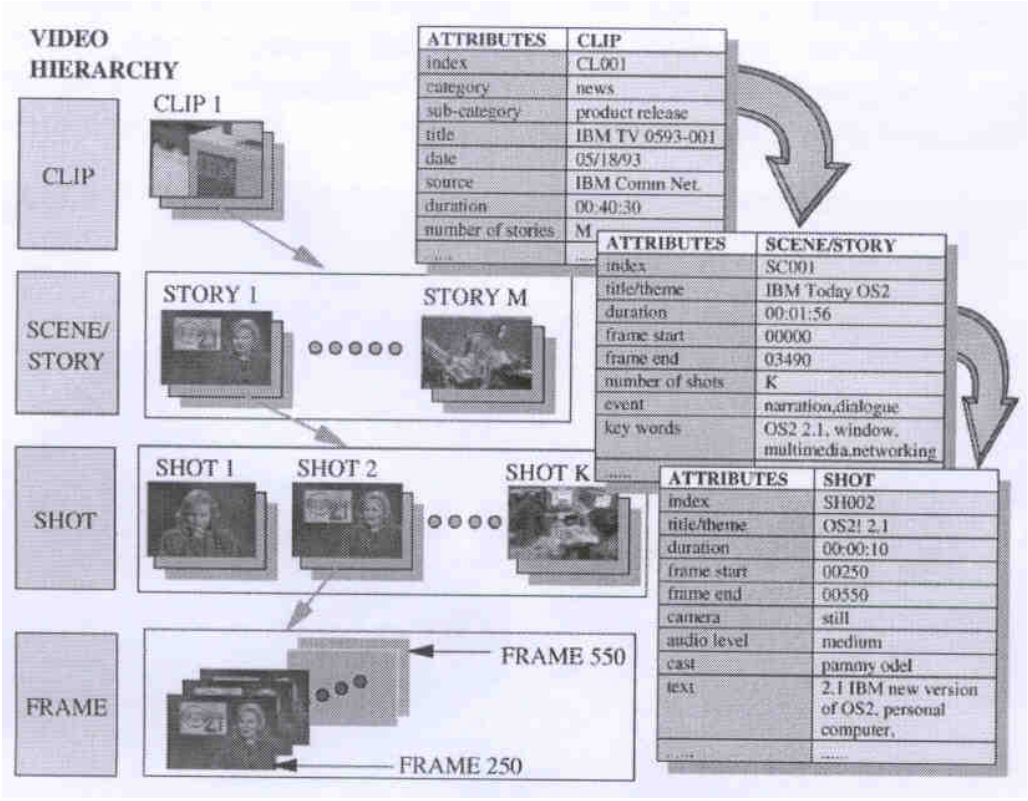


Figure 8.5: Video Decomposition Hierarchy (taken from Yeo & Young [1997])

Smeaton [2000] advises that manual annotation / mark-up of video should be kept to a minimum, with preference given to automatic techniques which yield consistent (even if occasionally incorrect or unexpected) results. Typical automatic *shot boundary detection* and *scene change detection* techniques[76] attempt video clip segmentation via the use of *scene transition graphs*, inter-frame and inter-shot colour histogram comparisons, and motion detection algorithms. Benini et al. [2008] propose a thematic segmentation based on *logical story units (LSU)* where hidden Markov model (HMM) analysis is used to group shot sequences which – based on low-level audio-visual feature extraction – appear to exhibit features/properties which share "a common semantic thread" [Benini et al., 2008]. HMM video segmentation analysis has also been successfully used to detect dialogue scenes [Wang et al, 2000].

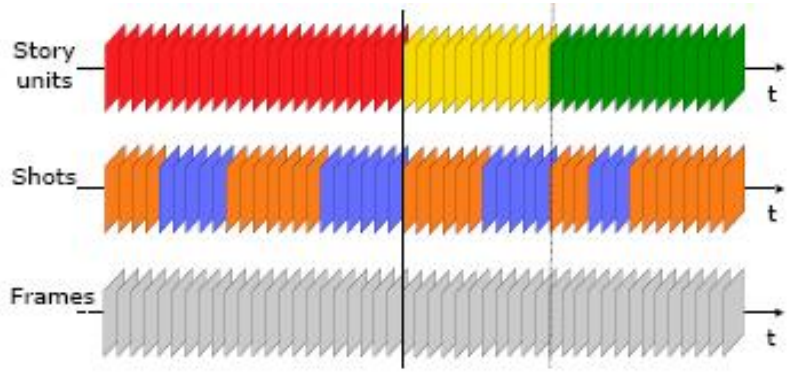


Figure 8.6: Hierarchical decomposition of a video into shots and story units (taken from Benini et al [2008])

User Actions

Once video content is indexed, it is important to consider the ways in which users may wish to interact with it. Lee & Smeaton [2002] define the following potential user actions that a video library interface should support:

- browsing and selecting video programmes from a collection
- content querying of a video programme
- browsing the content of a video programme
- watching a video programme (all or part of one)
- re-querying the video digital library or within a programme

With regards to browsing or searching, Lee & Smeaton [2002] mention that searching is often done based on querying video metadata (i.e. the title, date, or description of a clip.) Smeaton [2002] explains that this can take the form of matching a query against some unit of information which can be as broad as a whole video or limited to some subset therein. However, van Houten et al. [2004] assert that browsing is a more natural behaviour in the context of videos, because it can sometimes be difficult to articulate or find what one is looking for when using a keyword search. Yang & Marchionini [2005:1] agree that browsing is easier and faster for users, stating that "video information needs are sometimes hard to express in words, but are easily clarified when the picture/video clips are seen." Additionally, it is often the case that initial browsing often leads to the formulation of more specific search criteria.

Once an individual has located a video of interest, content browsing can occur in the form of allowing him/her to fast-forward and rewind through the clip, although alternative approaches do exist such as implementing video playback functionality via the manipulation of some static graphical component, such as a thumbnail image or even hyperlinked text. This intra-video "click-and-play" browsing technique has been implemented for the MultiMatch video retrieval interface [Carmichael et al., 2008]; this MultiMatch video retrieval interface also incorporates automatic speech-to-text transcription of the video's soundtrack (which is synchronised with the shot sequences) so that a specific video segment can be located based on its verbal

[76] Shot boundary and scene boundary detection techniques are quite similar, the principal difference being that the latter boundary detection technique works at a higher hierarchical level.

rather than visual content (see Figure 8.7). Actual playback is often the final step and many interfaces support this by providing video player software (such as RealPlayer) to display the content. However, re-querying is also important to consider, as often a user will need to continue to interact with the system as his/her goals and information needs evolve [Lee & Smeaton, 2002].

Surrogates

After segmentation, the information extracted from the video clip must be displayed in a manner which is readily accessible and easily interpreted by the viewer. There are several approaches that can be taken when displaying the results of a video search. However, in general, some sort of surrogate must be presented. Yang et al. [2003: 3] define a video surrogate as "a compact representation of the original video that shares major attributes with the object it represents." They go on to mention that the goal of a surrogate is to act as a summary and to enable the user to get the gist of the video's content. A successful surrogate allows the user to make accurate judgements about the relevance of a video without having to watch the entire clip. There are a variety of surrogates that can be used, according to Yang et al [2003]:

- text surrogates (bibliographic information/metadata)
- still image surrogates (keyframes)
- moving image surrogates (sped-up versions of the video)
- audio surrogates (extracted audio information from the video)
- multimodal surrogates (a combination of video, audio, and text)

What many researchers seem to agree upon is that since humans process visual images more quickly than text and have an accurate recognition memory for pictures, providing easy visual access to video information is desirable [van Houten et al., 2004; Yang et al., 2003]. Christel et al. [2002] add that the combination of both textual captions and visual summaries is better than using textual summaries alone. Text can be extracted from associated closed-caption information (if available) or obtained using speech recognition programs.

In terms of the layout and presentation of video surrogates, Lee & Smeaton [2002: 11] propose that "keyframe-based browsing is similar to the now de-facto standard feature of 'thumbnail browsing' in image retrieval interfaces…" Keyframes are selected frames from a video displayed either as an individual image or as a temporally-ordered sequence of images. The idea behind choosing which keyframes to display is that they are the most representative of the overall video content. Christel et al. [2002] used synchronization metadata and inverse document frequency metrics to find the highest-scoring shot for an individual query. However, this is not always an easy task. Keyframes may be chosen either manually or automatically (if automatically, they can be selected at regular time intervals in the video, or they can be taken from a certain place in each scene with the help of boundary detection methods, such as the gradual transition method proposed by Tsamoura et al [2008]). Regardless, there is also the question of how many keyframes to display. Lee & Smeaton [2002: 14] mention that there is no easy answer to this question: "it will not be possible to say which level of granularity is best for every situation as one user in one situation will have different needs from another user."

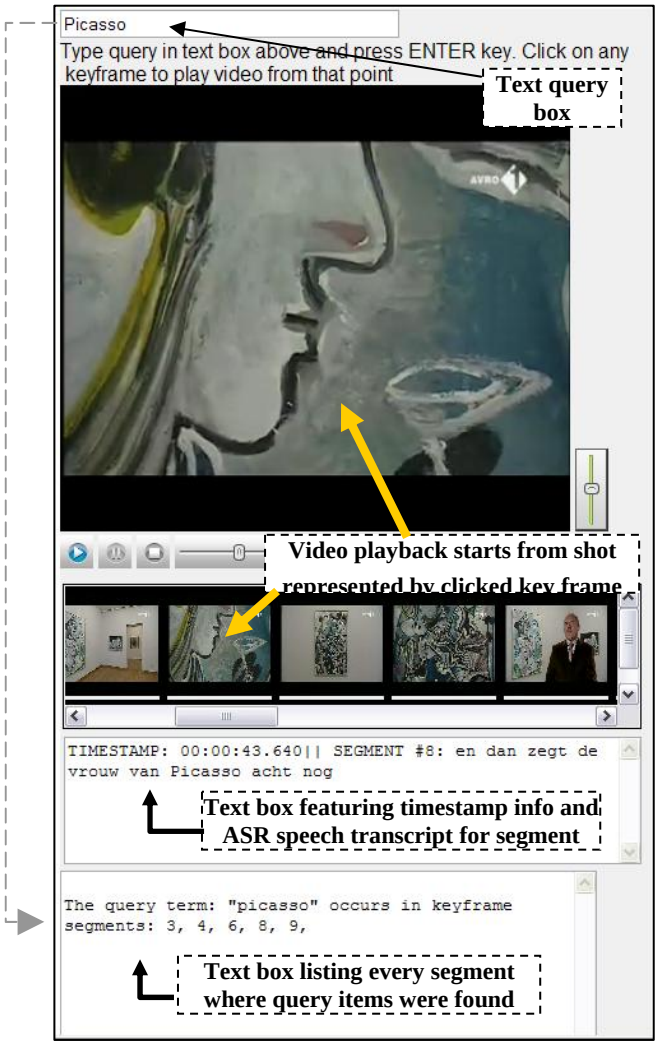


Figure 8.7: Screenshot of the MultiMatch Video Interface (with selected video document already loaded).

Visualisation layouts

There are several ways in which the surrogates can be laid out within an interface. Many approaches have made reference to Shneiderman's [1998] mantra of "Overview first, zoom and details on demand" as a guiding principle. Some common approaches are [Lee & Smeaton, 2002]:

- Storyboards (a series of small keyframes displayed spatially on the screen in chronological order)
- Slideshows (the keyframes are displayed one at a time in a slideshow. The transition from one keyframe to the next can either occur automatically or can be controlled by the user.)
- Hierarchically arranged browsers (in which keyframes can be viewed by drilling down—best for structured programmes such as the news.)

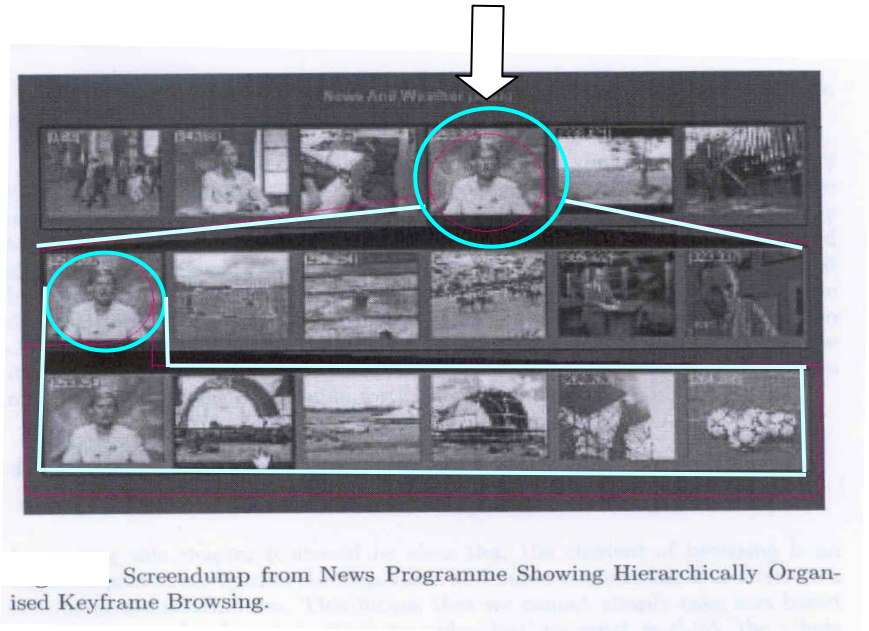
However, these are not the only options. Other approaches to visualisation design will now be described. As previously mentioned, the most common display paradigm is the 2D story-board style grid layout. Since it is usually not feasible to display every frame in a shot, most video information visualisation techniques attempt to identify the frame within a shot- or scene- sequence which typifies the content of said sequence. This most typical frame is then displayed on the grid and configured to support some form of interactive playback – clicking on this representative frame will result in the playing of some or all of the frames within the same shot or sequence.

These representative interactive frames are usually arranged on a 2D grid in a sequential and/or hierarchical fashion. Figure 8.8 shows a typical frame sequence displayed in a strictly hierarchy-flattened sequential fashion, while Figure 8.9 depicts a similar 2D grid but with a frame→shot→scene hierarchy. It is to be note that Figure 8.9's three-tier display reflects this 3-level frame→shot→scene hierarchy with the top level corresponding to the scene and the 2 lower levels corresponding to the shot and frame collections respectively. Selecting a typical frame from the uppermost level (i.e. scene level) for playback will have a "drill-down" effect, i.e. the displays in the two lower levels will be updated to show:

- the most-typical frames from all the shots at level 2
- representative frames amongst all the single frames at level 3[77]

Screendump from Fischlár Showing Keyframes Taken from the Middle of a 30 Minute News and Weather Broadcast.

Figure 8.8: Hierarchy-flattened Frame Sequence Display [Yeo & Young, 1997]



Screendump from News Programme Showing Hierarchically Organised Keyframe Browsing.

Figure 8.9: 3-Tier Hierarchical Frame Sequence Display (ibid).

[77] Of course, this hierarchy could extend one level higher with the uppermost level displaying a series of separate video clips, the second level would then display a series of scenes from any video clip which has been selected at the top level, the third level would then display a series of shots.

Boreczky et al. [2000] have implemented a refinement of the 2D grid layout, where representative frames considered to be of greatest relevance are presented on bigger panels, in a fashion similar to that employed in comic books where climatic scenes in the narrative are given more space on the page. The frames are then slotted into position using a near-optimal "row block" packing algorithm, an example of which appears in Figure 8.10. Note that some panels (as is the case with panel 5 in this example) may be resized to better fit available space.

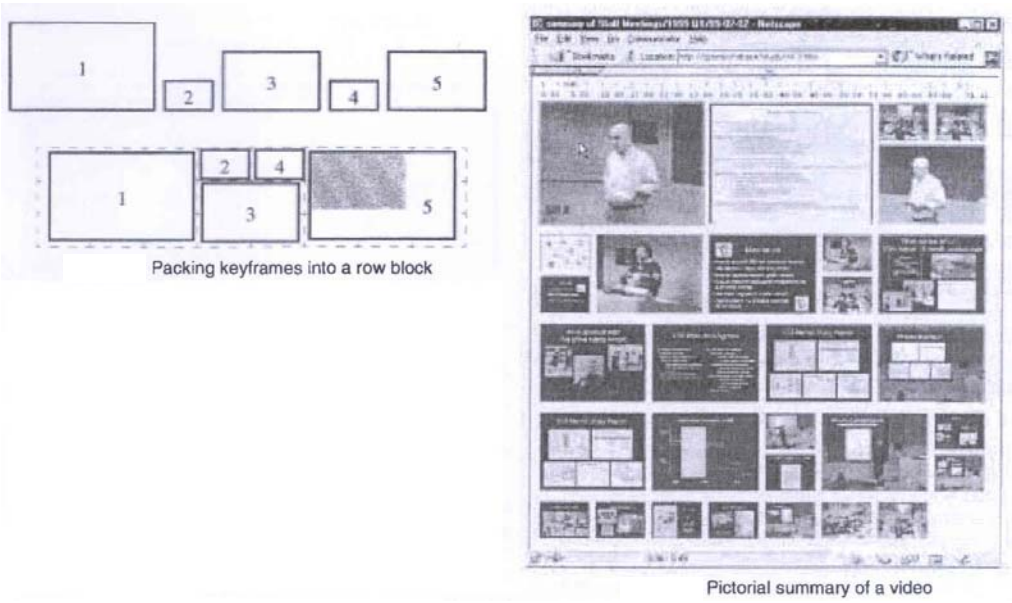


Figure 8.10: Boreczky's [2000] Comic Book Style Layout

Yeo and Yeung [1997] also implement a (less sophisticated) variation of the comic book layout (Figure 8.11), but theirs does not incorporate the level of user interaction evidenced in the Boreczky model.

Figure 8.11: Yeo and Yeung Implementation of Comic Book Layout

Smeaton [2002] reminds us that, in the case of video clip retrieval and indexing, it is important to use a variety of IR techniques which will be capable of processing all possible data types which may be embedded in the video object, these include:

- Using OCR to decipher any text captions or titles (as is often the case with clips originating from news programs, documentaries, etc).
- Automatic speech recognition (ASR) and musical instrument recognition to process sound tracks. If full-blown ASR recognition returns low accuracy rates, Smeaton [2000] advocates a phone recognition approach, where the user's text-based request is decomposed into a string of phones and this phone string is compared to the phone sequences extracted from automatic phone recognition processing of the video clip's sound track. Sound tracks (and their associated video clips) with a high hit rate are deemed to be a good match and included in the list of returned documents.

Christel et al. [2002: 561] present the idea of visual collages as "new interactive tools facilitating efficient, intelligent browsing of video information by users as they follow their shifting information needs." A collage is a dynamic overview of video results where users can "drill down" or zoom in on areas that are of particular interest. More targeted browsing of videos by location can be done via a map collage interface (Figure 8.12), or by time via a timeline interface (Figure 8.13). The collages also contain text that refers to the most frequently-occurring phrases in the videos (which are all news reports.) Overall, these collages incorporate automatically-generated data and the user's query context to create a dynamic and interactive way of exploring a large quantity of results.

Figure 8.12: Map collage interface [Christel et al., 2002].

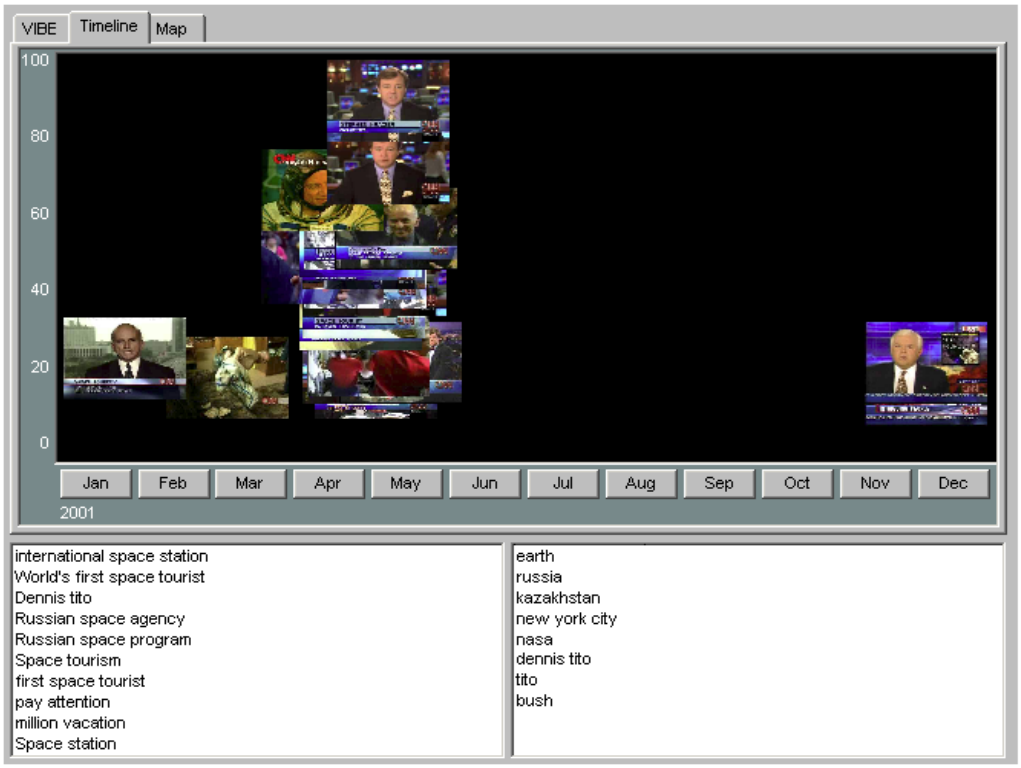


Figure 8.13: Timeline collage interface [Christel et al., 2002].

Zhang et al. [2007] conducted an evaluation of a system for searching multilingual videos employing ASR and MT to convey some notion of the content. They found that while users were sometimes able to extrapolate the meaning of mistranslated terms, in other cases, errors in the machine translated ASR output were problematic and required creative strategies to overcome (e.g. consulting non-text aspects of the video such as the visual feature). Rautiainen et al. [2006] constructed a video browser based on two orthogonal facets: temporal (e.g. a video timeline) as well as content-based similarity clusters. When using this system to retrieve video segments, improvements were found in retrieval effectiveness, although these were more beneficial for novice as opposed to expert users.

Overall, general good practice to follow when designing a video retrieval interface is to support as many types of tasks and behaviours as possible, while making it easy to switch between different features [Lee & Smeaton, 2002]. Similarly, Smeaton [2002:222] recommends providing "video navigation which seamlessly combines searching for objects, shots, or scenes, browsing and following hyperlinks between related video elements, and summarisation based on generated summaries or sets of keyframes" as the most efficient and useful way of enabling navigation through video libraries. It must also be noted, however, that if such multimodal searches incorporates automatic speech recognition (ASR) technology that may feature word error rates (WER) exceeding 20%, it may be necessary to alert the user to the possibility that the speech transcripts of the video soundtrack may be – to some degree – incorrect and thus hinder the information retrieval process. Carmichael et al [2008b] have proposed a "more like this" matching algorithm which searches for word sequences of similar phonemic structure to the user-defined query term in order to suggest that such words/phrases may actually be misrecognised instances of the query term. Moreover, the professional video archivists participating in the evaluation study by Carmichael et al [2008b] indicated a strong preference for an information-rich but graphically minimalist interface, keeping to a minimum embedded textual components such as combo/text boxes. This preference for a minimalist interface design – already successfully adopted by the well-known Google search engine – represents a significant challenge in

the face of user demand for even more multimedia and multimodal information to be presented as part of the online IR process.

8.3.3 Audio Retrieval Interfaces

Indexing and retrieval of audio documents is, in principle, quite similar to its video counterpart with one significant exception: important progress has been made in developing methods for extracting semantic meaning from acoustic signals containing music and speech. The three principal *music information retrieval* (MIR) techniques are *automatic musical instrument recognition, automatic music score transcription* and *automatic genre classification*. West and Cox [2005] have reported considerable success in classifying music recordings according to *genre* (e.g., musical style such as jazz, rock or classical, etc.). In terms of instrument recognition, Eggink and Brown [2004] achieved a recognition rate accuracy averaging 80% for certain types of instrument and given certain conditions[78].

Accuracy rates for *automatic speech recognition* (ASR) vary significantly depending on the constraints and scope of the task, ranging from in excess of 90% if the recogniser is small vocabulary and *speaker dependent* (i.e. trained on speech samples from the target speaker) to around 70% if the recogniser is *speaker independent*[79] and the number of word items to be recognised is quite large (e.g. in excess of 5,000). In the context of the speech indexing tasks to be attempted by the MultiMatch project, the most appropriate ASR system configuration would be speaker independent and large vocabulary. Furthermore, it would be necessary to devise some method of segmenting an audio clip or video clip sound track into thematically distinct units representing, for example, individual news stories or musical performances. These segmentation techniques are discussed in the following section.

Thematically indexing audio data

Thematic segmentation of speech and music has a well-established tradition with associated technologies being sufficiently mature as to permit commercial exploitation. A notable example of such technology is the THISL speech recognition and indexing system implemented by Renals et al. [2000] for the indexing of radio broadcasts from the United Kingdom's BBC news network. The segmentation methods employed by THISL are typical of most state of the art applications and consist of the following pattern recognition techniques:

- Detection of significant non-speech events: it is usually the case that individual news items will be separated by some type of non-speech event, usually in the form of a period of silence and/or a *station ident* – an ident being a short musical jingle or other distinctive audio event which is recognised as the acoustic equivalent of a company logo.
- Detection of a shift in term frequency: given that a news item normally has some unifying theme, it is quite likely that there will be some specific word or phrase which will be mentioned repeatedly for the duration of that news item but which will be mentioned less frequently – if at all – in subsequent or preceding news items.
- Detection of change in ambient noise quality: a sudden change in the loudness and quality of background noise is often an indicator of a change in physical location. This may in itself not indicate a boundary between two items, but when used in conjunction with the two techniques listed above, it can offer useful clues to facilitate segmentation.

Therefore, audio files can be indexed in a variety of ways, based on extracted metadata, acoustic indexing (i.e. using automatic speech recognition), or semantic indexing (based on topic or theme, as described above.)

Visualisation of audio search results

[78] The instrument recognition software application devised by Eggink and Brown proved more capable at recognising certain types of instrument (namely the wind instruments such as the flute). Furthermore, performance rates dropped if there were more than six other instruments being played simultaneously.

[79] In speaker independent ASR systems, the recogniser is trained on speech samples from a variety of individuals who typify the speaking style of the target population. Such training procedures will normally produce an ASR system which will work reasonably well for most but with an accuracy rate below that of a speaker dependent system customised for a specific individual.

After audio files have been properly indexed to facilitate searching, the next issue involves determining how best to display the results of a search. Logan et al. [2004] describe two common ways for users to find an audio file: either by searching for keywords contained in the files' metadata or associated transcripts, or by conducting a "similarity search" for items that are related to a given file. In general, most searchable audio archives present results in a simple list of links to files, often ranked by supposed relevance [Van Thong et al., 2001; Foote, 1999].

A related consideration involves enabling a user to find the relevant content within a given media file [in case he or she only is interested in one section of a longer recording.) Foote [1999] mentions that the typical interface for audio playback and browsing is based on the tape recorder metaphor. In this presentation, the audio file is presented as a continuous stream which the user can navigate using play, stop, fast-forward, and rewind buttons. However, this approach is fairly unsophisticated and current research has focused on optimising ways of letting users search for and browse audio content. Such research can take two different approaches, either focusing on improving the presentation and navigation of search results, or concentrating on novel ways of enabling navigation within a given file.

With regards to the first kind of approach, much of recent thinking focuses on presenting results "in a way that allows users to quickly identify the files that are really important for their particular information needs" [Hürst & Venkata, 2003]. Sometimes a brief amount of metadata relating to audio files (such as title, author, and file name) is displayed in search summaries, but this does not necessarily help a user to judge the file's relevance (or lack thereof.)

The SpeechBot project [Van Thong et al., 2001] attempted to address this problem by using speech recognition technology to automatically generate transcripts of audio files. Once the contents of an audio file have been transcribed, the retrieval task becomes essentially a text retrieval task: users enter search keywords and then can view the transcript to get an idea of the relevance of the result. Logan et al. [2004] mention that such an approach is advantageous because it is able to show the precise position of a word's occurrence within the file as a whole. However, the automatic nature of the transcription means that misrecognition of words or out-of-vocabulary terms can pose problems.

Although the SpeechBot transcriptions did contain such recognition errors, they were still deemed helpful in providing users with a general gist of the audio files' contents. They could then determine which files were worthy of further investigation based on these brief textual summaries. Overall, in the case of SpeechBot, it was determined that highlighting search query words in the transcription "was essential, and gave the user strong feedback on the relevance of the document even if the speech recognition output was sometimes hard to read and understand" [Van Thong et al., 2001: 12]. Whilst SpeechBot is no longer publicly accessible on the Web, newer audio search sites devoted to podcast searching operate using a similar approach of automatic transcript generation. One such example is the PodZinger site (www.podzinger.com). This site uses a similar approach of presenting automatically generated transcripts that show the keywords in context.

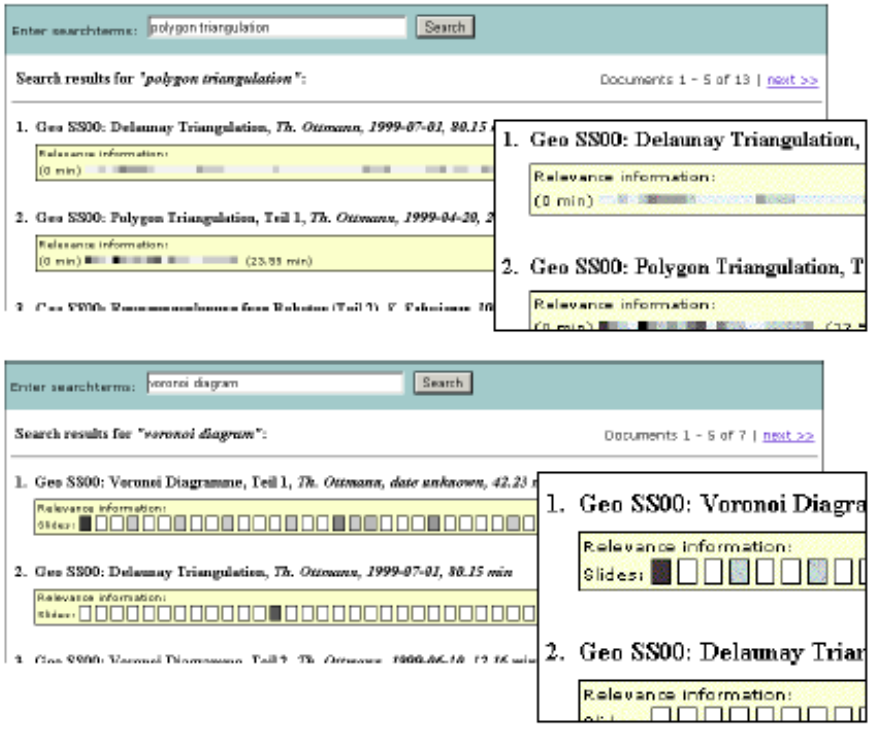
The second type of approach to interaction with results involves navigating within a specific audio file. As discussed before, the "tape recorder" method is commonly used for this purpose but it often has drawbacks. First, it is time consuming to listen to a long audio file when only a small subsection contained somewhere within is of interest. Although many playback features offer some indication of a timeline (i.e. how much time has elapsed at a given point in the recording,) it can sometimes be difficult to go back and re-locate the exact position of a point of interest. Again, new approaches have been explored in this area. Foote [1999] mentions a technology called SpeechSkimmer, which can compress audio recordings so that they can be played back at an accelerated but still comprehensible rate. Tucker & Whittaker [2006] tested different compression techniques in order to reduce the amount of time needed to listen to a file. Both excision (the removal of insignificant information) and compression (speeding-up) techniques were evaluated, and it was found that excision was generally more effective and better-liked by users than compression.

Even more useful than either of these methods, however, is the ability to skip directly to relevant portions of an audio recording. Hürst & Venkata [2003] explored ways of enabling this in the interface for a collection of archived lectures and presentations. They explored the idea of search using automatically generated transcripts but found in their case that these were not of high enough quality to be used even for gist or overall topic identification. As an alternative way of aiding visualisation, they designed a graphical timeline display with icons representing the subdivisions of the recording (in this case, each icon stood for one slide

in a lecture). The icons were then colour coded to show relevance to a search keyword: a darker coloured icon indicated higher relevance, suggesting that the keyword occurred most frequently in this section. Figure 8.14 displays two such timeline displays that were tested.

The overall advantages of this design include giving easy access to the audio file at several intermediary points (often linked to a change in topic,) and visually displaying some indication of relevance. PodZinger also employs a means of quickly and easily locating the occurrence of keywords. Clicking on a term in the displayed transcription will automatically begin playing the audio file at the point where the word was mentioned (See Figure 8.15).

In summary, it is not always easy to search audio files and display the results in a clear, informative format, but enabling users to get an overview of a file's content and its likely relevance is important. If they find a file that could be of interest, providing ways of quickly and efficiently browsing the content is also useful, particularly if the file is longer than a few minutes (or the length of the searcher's limits of patience.)



Interface 1: The audio file is represented through a symbolic timeline. Each square represents an equally sized set of words of the audio transcript. Relevance of the corresponding audio clip is indicated through different colors.

Interface 2: Icons for the slides from a lecture are used to represent the corresponding audio parts. Relevance is indicated through different colors.

Figure 8.14: Graphical displays showing location of relevant words [Hürst & Venkata, 2003]

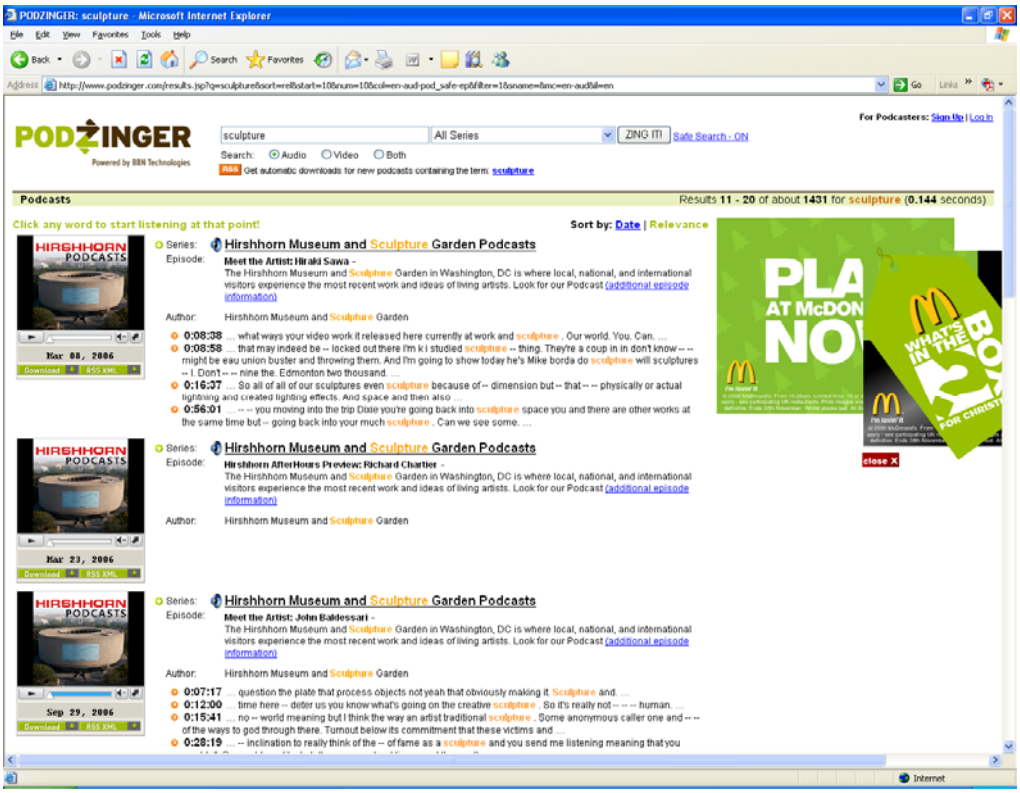


Figure 8.15: Sample results screen from PodZinger enabling the playing of the file at points where the keyword is mentioned.

8.3.4 Example Multimedia Search Interfaces

Multimedia search engines can offer a variety of possible media formats to be searched. Based on a sample of 16 online multimedia search systems, Table 8.2 shows a breakdown of the number of combinations for each type (image, audio and video).

Table 8.2: Search Category Combinations Supported by Popular Internet IR Sites

Media types	Number of sites	Examples
Images Only	9	www.live.com www.clusty.com http://www.google.co.uk/imghp?hl=en&tab=wi&q=
Images, Audio, Video	4	www.alltheweb.com
Video Only	2	www.youtube.com
Audio & Video	1	www.singingfish.com

Table 8.2 summarises the main functionalities exhibited by the sample selected. Of the six sites that had content in more than one medium, only one of them (www.Singingfish.com) offered the possibility of searching several media types at once. For the rest, search had to be limited to a specific type (i.e. image OR audio OR video, but not a combination.) The results of the Singingfish search, however, are not separated by type. Free text was the predominant means of searching. Only one site (www.YouTube.com) had the possibility of browsing by category. Most of the sites followed a similar layout and respected similar conventions. They were simple and based on the Google interface model. In terms of results presentation,

again, clear conventions prevailed, with image results displayed in a grid and Audio/Video results shown as a list, often with a thumbnail and a brief description.

Table 8.3: Example online multimedia retrieval systems

Collection holdings	Percentage	Example
Images	86 %	See above
Audio	36 %	
Video	50 %	
Tabs for different media	60 % (3 of 5)	www.altavista.com
Searching functionalities		
Free text search	100 %	
Advanced search	53 %	
Search all types of media at once	20 % (1 of 5)	www.singingfish.com
Browsing functionalities		
Category list	14 %	www.youtube.com
Hierarchical browsing	0 %	
Tag cloud	7 %	www.youtube.com
Results		
Displayed in grid / rows	100 %	
Other display	7 %	www.live.com (infinite scroll bar)
Ability to refine search / change result layout	57 %	www.creative.gettyimages.com http://www.google.co.uk/imghp?hl=en&tab=wi&q=
Multimedia results segregated by type	40 % (2 of 5)	www.altavista.com
Recommendations / "more like this"	14 %	www.youtube.com
Clustering of results	6 %	www.clusty.com

8.4 Semantic Web Interfaces

According to Fluit et al. [2005], "the Semantic Web is an extension of the current World Wide Web, based on the idea of exchanging information with explicit, formal and machine-accessible descriptions of meaning." As Maedche & Staab [2002] explain, these descriptions can be utilised to facilitate finding, integrating and connecting information in a way above and beyond that which can be done with a simple keyword search. Benjamins et al. [2004: 434] highlight the value of semantics in the humanities domain, stating that most information-seeking in this area involves "events, persons, and movements in a historical or cultural context."

Similarly, Hyvönen [2007] asserts that the cultural heritage domain is well suited to the creation of semantic portals. These can, among other things, (1) give an aggregated, global overview of heterogeneous content and (2) provide a more "intelligent" way of examining content through semantic linkages. There are several ways in which said intelligent services can utilize semantic information. These include semantic search, semantic auto-completion, faceted semantic search, semantic browsing and recommendation links, relational search, and visualizations on maps and timelines.

Hildebrand et al. [2007] outline the various elements of the semantic search process, which include construction of the query, execution of the search algorithm, and presentation of results. With regards to interface design matters, they mention both typical and more experimental visualization techniques ranging from ranked lists, clustered result displays, tag clouds, cluster maps, and data-specific designs such as timelines. Some examples of these will be subsequently illustrated; however, for an extensive list of example systems and their features, please refer to Hildebrand et al. [2007] and associated survey[80].

Semantic web information needs to be contained in some sort of structure in order to be useful. Ontologies are one such structure: they help to make semantics explicit and machine-readable using metadata [Fluit et al., 2005]. Overall, an ontology is a formal conceptualization of a shared domain [Benjamins et al., 2004; Maedche & Staab, 2002] that can be communicated across people and computers.

Table 8.4: Various functionalities that could be employed to help visualize and represent semantic relationships [Albertoni et al., 2005].

Functionality	**Explanation**
Graphical selection	To select different information sources such as URI, PDF or DOC documents
Visualisation manipulation	To re-organise, move and add graphical elements
Co-occurring terms visualisation	To visualise a statistical thesaurus to expand user queries with other highly frequent terms
Highlighting	To visualise a selected element and all its related sources
String search	To search for a co-occurring word and to navigate the ontology hierarchy
Hierarchical visualisation	To browse content at different levels of granularity
Clustering visualisation	To group content by similarity criteria
Ontology instances	To visualise the instances of a selected class separately or directly in the ontology graph
Venn diagram representation	To describe and compare elements and characteristics
Ontology graph navigation	To easily navigate the ontology graph structure
Map based visualisation	Organising content by theme (i.e., on a geographical map)

[80] http://swuiwiki.webscience.org/index.php/Semantic_Search_Overview [Accessed 3 June 2008.]

Benjamins et al. [2004] add that ontologies can portray relations between domain concepts in a way that thesauri cannot. The issue with ontologies, however, is that the information they contain is typically unsuitable to be published as-is: for example, navigation may become tedious if there are too many concepts. Therefore, it is important to consider how best to design the visualization of an ontology, to support an understanding of its structure and to facilitate navigation. Various approaches to this challenge have been attempted and will now be discussed.

Albertoni et al. [2005] outline various functionalities that could be employed to help visualize and represent semantic relationships. These are used in various combinations by current systems with regards to graphical visualisation and/or interaction with information. The most frequently used functionalities, based on a survey of 9 systems, included those listed in Table 8.3 (in order of prevalence).

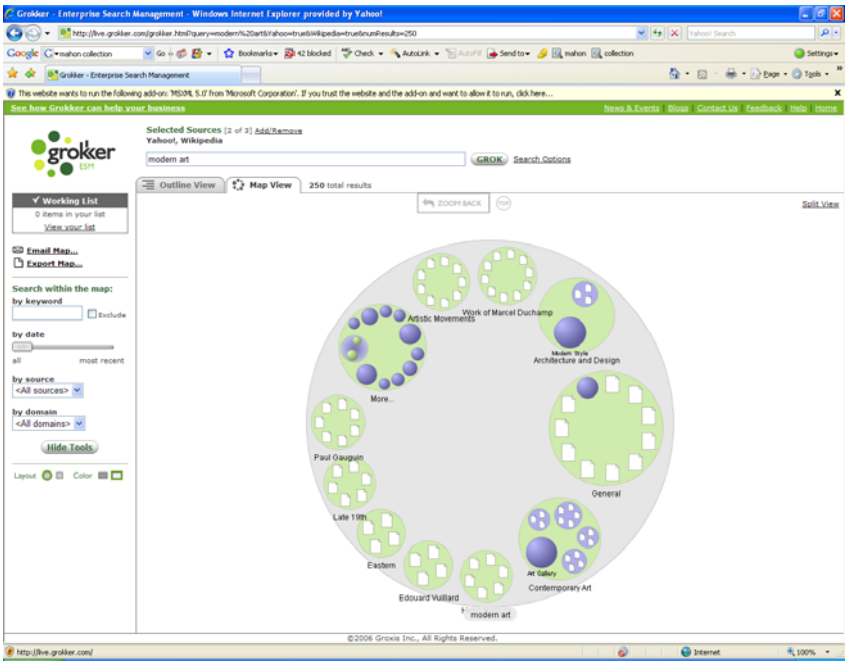
Katifori et al. [2007] provide a thorough survey of the present state of the art in ontology visualization methods. For purposes of their survey, they group visualization types into six categories:

- **Indented list**: a Windows Explorer-like tree view.
- **Node-link and Tree**: displaying an ontology as a set of interconnected nodes, which can be expanded and retracted to increase or decrease the level of detail.
- **Zoomable**: present "child" nodes nested within their parents; the user can zoom into the child nodes to enlarge and view them in greater detail.
- **Space-filling**: use the entire amount of screen space by subdividing each node into the appropriate number of children
- **Focus + context or Distortion**: similar to a fish-eye view, the main area of focus is centrally displayed, with the other related nodes presented around it and gradually decreasing in size.
- **3D Information Landscapes**: documents appear as 3D objects arranged on a plane; colour and size are used to help depict relationships.

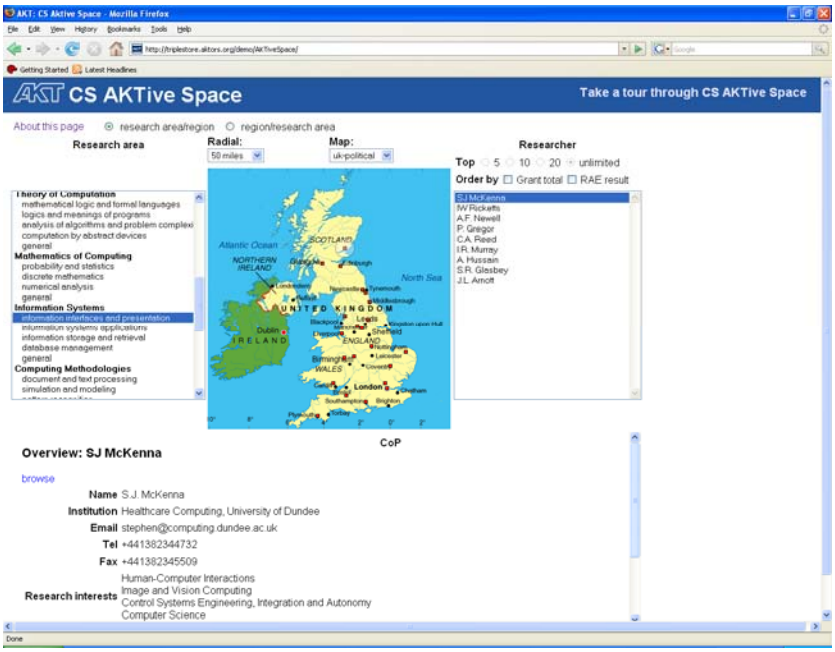
Each of these approaches has advantages and disadvantages with respect to navigation and interaction issues. The survey's conclusion was that there is not one specific method that is appropriate for all applications; the best method to use may depend on the characteristics of a certain ontology (e.g., how large or complex it is). Alternatively, another approach could be to provide the users with several visualization options, so that they may choose the one best suited to their needs.

Various commercial and experimental systems have been created to provide a means of navigating and understanding relationships between different categories in an ontology or other large information set. A comprehensive listing of many systems can be found in Katifori et al. [2007]; however, a selection of some currently-existing tools will now be described.

- Grokker (http://www.grokker.com)
  Grokker is a research and information management tool that searches multiple sources at once and displays results either in a clustered, filterable outline view, or as a topically organised visual map.

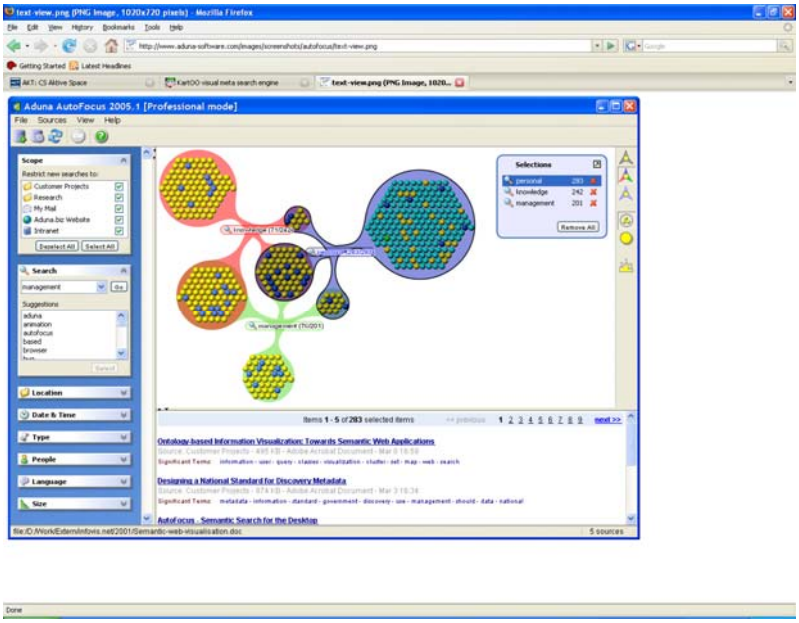


- CS AKTive Space (http://triplestore.aktors.org/SemanticWebChallenge/CSAKTiveSpace/)
  According to the website, "CS AKTive Space is a smart browser interface for the semantic Web which combines an ontologically motivated view of the application domain, namely the UK computer science research community, and simple geographic information to provide information on the leading researchers and research hotspots in the UK." It enables information retrieval by metabrowsing; one can find lists of researchers using categories of research areas filtered by a location on a map, or vice-versa.



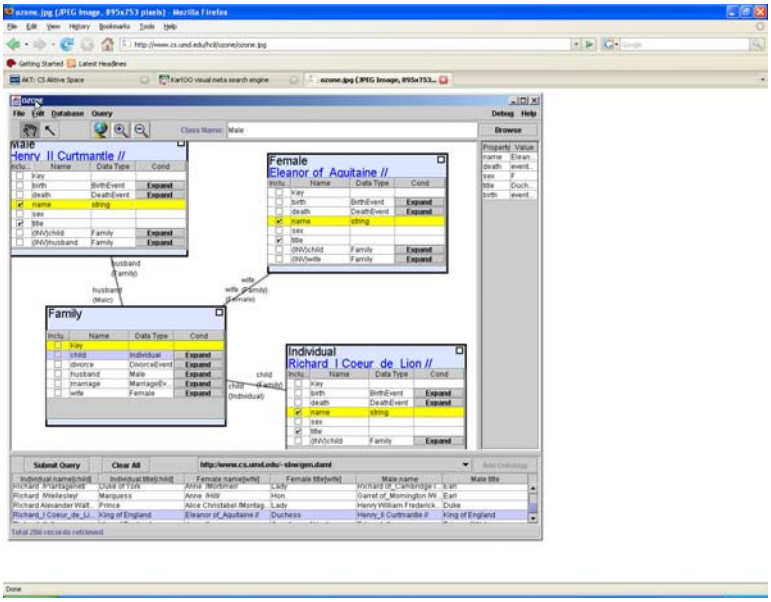
- Kartoo (http://www.kartoo.com)
  This "is a free metasearch engine which presents its results in the form of an interactive map, exactly like a roadmap on which the cities are replaced by websites and the roads by thematics."

- Aduna AutoFocus (http://aduna-software.com/products/autofocus/overview.view)
  AutoFocus is a program that enables the searching of information on a PC or on websites, using facets and a visualization technique called Cluster Maps.

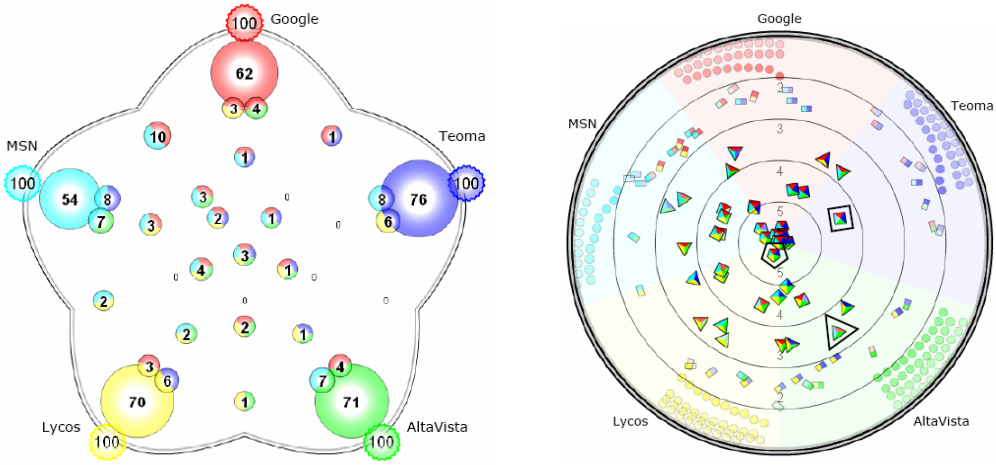


As described in Fluit et al. [2005], Cluster Maps are used to represent relationships between classes in an ontology. Instances belonging to the same class are grouped into clusters. The physical proximity of clusters in the map corresponds to semantic closeness. It is also possible, through visual inspection, to see which classes overlap and which items (if any) belong to multiple classes.

- OZONE (http://www.cs.umd.edu/hcil/ozone)
  OZONE is a query interface that gathers ontological information from the semantic web and lets users search through this using a relational database-like model. It is designed to facilitate interactive and incremental query formulation and graphical representations of ontology artefacts.



- MetaCrystal (described in [Spoerri, 2004])
  This system consists of several tools designed to provide a visual overview of the overlap between results gathered by several different search engines. The relationships are signalled using cues such as colour, size, proximity and orientation. Two different views (category and cluster bulls-eye view? can be used. These provide an overview of the top documents retrieved by combinations of different search engines, as well as by individual engines.



- WebTheme [Whiting & Cramer, 2002]
  WebTheme is designed to provide a visual overview and understanding of large collections of Web pages. The theme view visualization shows concepts on a sort of topographical map with more dominant themes being represented as taller peaks. This helps to convey the main themes in a collection and gives a sense of how they relate. The Galaxy view consists of a series of dots (each one representing a web page,) with dots clustered near each other relating to thematic similarity. Theme clouds are overlaid on top of dot clusters to indicate concept labels. From these broad visualizations, it is then possible to zoom in and discover more information about individual documents or groups thereof.

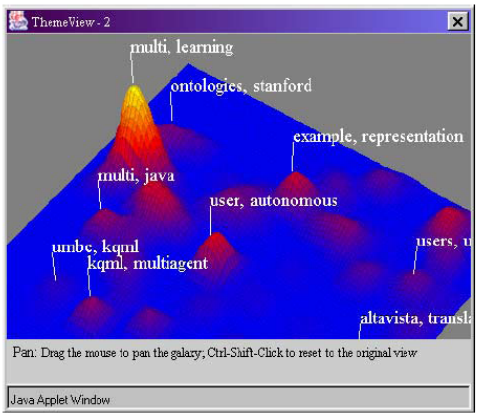


Fig. 1. ThemeView

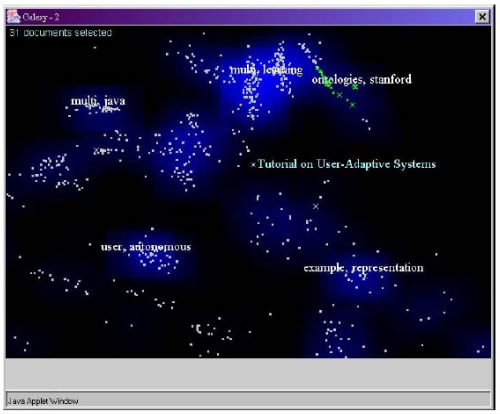
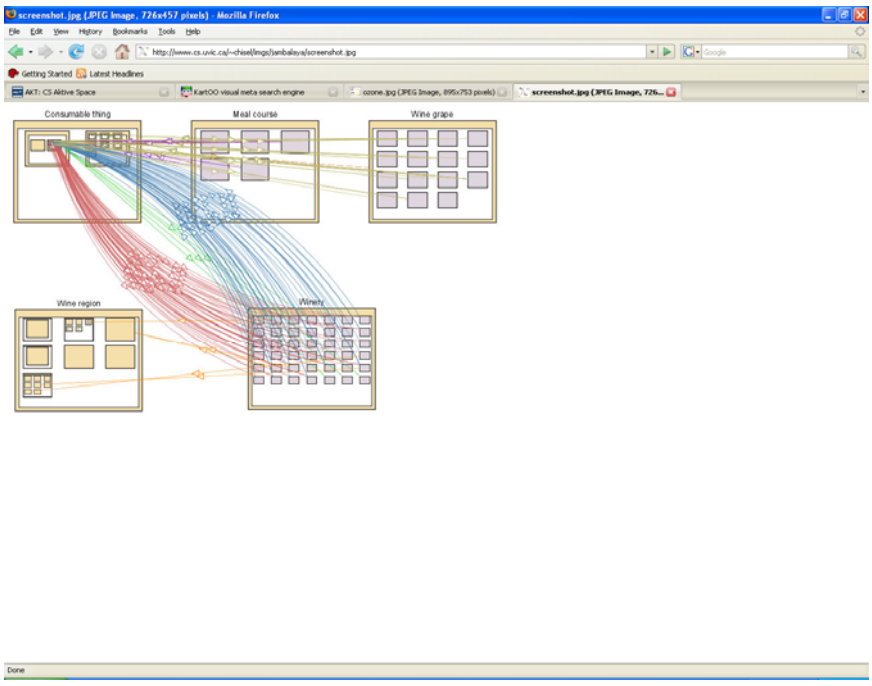
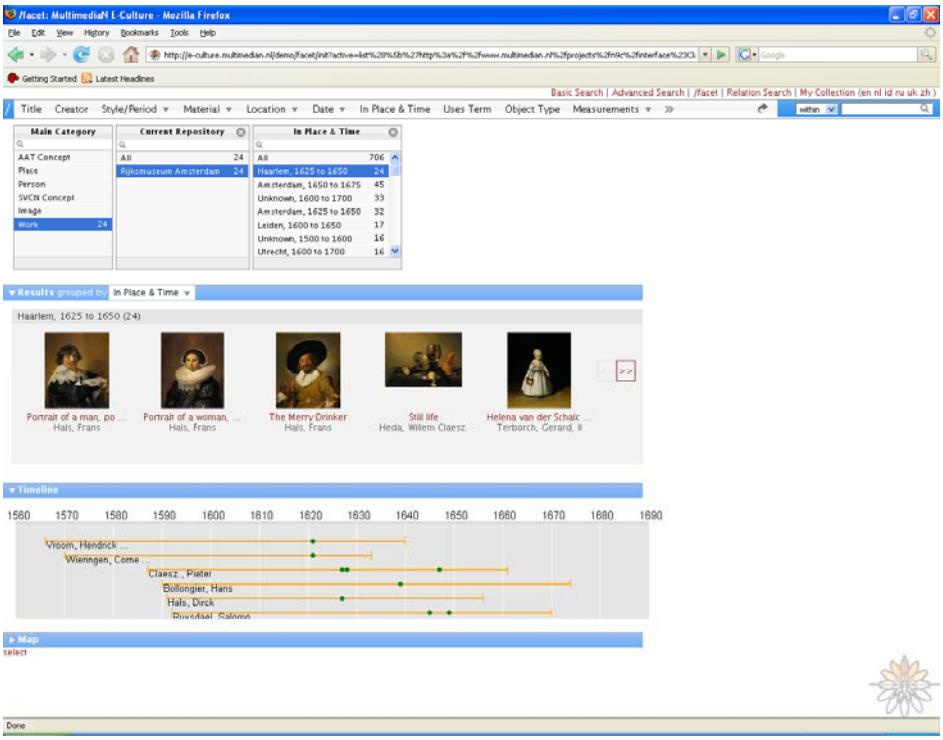


Fig. 2. Galaxy

- Jambalaya [Storey et al., 2002; http://www.thechiselgroup.org/jambalaya)
  Jambalaya is an ontology visualization environment that uses classes as nodes in a graph with subclasses (or other nested nodes within them). Slots connect the classes to reveal relationships. It is possible to zoom in and out for detail and context, respectively.



- /facet [Hildebrand et al., 2006]
  This project aimed to create a system that goes beyond the traditional faceted browsing model by navigating heterogeneous collections of data. The system integrates both browsing and keyword search options for navigating within the facets, and also provides images to help users unfamiliar with the terms to determine if the results are what they expected. Additionally, timeline and map views are included. This system was integrated into the MultimediaN project (http://e-culture.multimedian.nl/demo/facet).



Although semantic search and browsing are growing areas of interest, resulting in the creation of myriad prototype systems, few user evaluations of these have been conducted. The challenge of evaluating such systems is real, either because finding a baseline for fair comparison is difficult [Hildebrand et al., 2007] or because it is not trivial to obtain quantifiable, objective measures when dealing with exploratory search tasks [Kules & Shneiderman, 2008]. In the future, such evaluations should be attempted in tandem with the design of semantic web interfaces, in order to yield more information about how they are likely to be used and appreciated.

8.5 Cultural Heritage Interfaces

Currently, most cultural heritage institutions have some sort of online presence in the form of a website. Museums and art galleries have homepages and sometimes specific archives or collections that are part of a larger body have web portals of their own. These websites often provide some degree of access to the associated institution's collection in a digitised format. The degree of material that is available and the sophistication of exploration of this content vary from site to site, depending on the resources available to the cultural heritage institution in question. However, overall, a majority of these sites do have common features which include both search and browse functionalities at the very minimum. A summary of the relative proportions of functionalities taken from a sample of 56 cultural heritage sites is presented in Table 8.5.

Table 8.5: A summary of the functionality of selected multimedia search engines

Functionality	Percent	Example
Free text search	91 %	
Browse by category	71 %	www.archinform.net
Advanced search	70 %	
News/Calendar	61 %	www.tate.org.uk
Registration/login	45 %	
Multilingual	34 %	www.louvre.fr
Geographical search / Map	29 %	http://whc.unesco.org/en/map
Shopping	29 %	
Search within results / See "more like this"	29 %	www.fotolia.com
Ability to segregate multimedia results by type (if applicable)	29 %	www.archive.org
Feedback section	23 %	
Timeline / Search by time (12 sites total; 25% of these offer search by time only, 75% have a timeline (2 of the 8 were interactive)	21 %	www.birth-of-tv.org
View results in popup window	21 %	
Change results layout (order by..)	21 %	www.artandarchitecture.co.uk
Hierarchical browse	20 %	http://www.staffspasttrack.org.uk/
Sitemap	20 %	
Controlled vocabulary	9 %	www.tate.org.uk
Colour/layout search	7 %	www.hermitagemuseum.org
Query translation	5 %	www.fotolia.com
Multimedia results arranged by type	5 %	http://ec.europa.eu/avservices/home/index_en.cfm
Faceted browse	3%	http://orange.sims.berkeley.edu/cgi-bin/flamenco.cgi/famuseum/Flamenco
Allow user annotation	2%	BRICKS workspace

Overall, most of the sites surveyed offered basic, expected, useful ways of searching and browsing their collections but were not very interactive or advanced. As technological capabilities have improved, there has been an increasing realisation that the current functionalities for accessing cultural heritage information online can be enhanced and upgraded. For example, it has been argued that in the area of humanities, a keyword-based search "is not sufficient because one is above all interested in *relations* e.g. between artists, their works, the friends, their studies, who they inspired, etc." [Benjamins et al., 2004: 433.]

Kravchyna [2004] surveyed five categories of users to assess their information needs when using museum websites. The categories included were (i) museum professionals, (ii) scholars/art historians, (iii) the general public, (iv) university students, and (v) high school teachers. Across all groups, primary purposes for using museum sites were to determine the main exhibits and activities of interest, to gain knowledge about museum collections, and to learn of any upcoming activities by consulting any available event calendars. Additional priorities that were unique to the scholar group were related to gathering information for research (i.e. looking for specific images or looking for textual information on a museum object). Therefore, while some needs crossed group boundaries, there were also group-specific requirements.

Current research and projects are focusing on new ways to aggregate, search and display multimedia cultural heritage material originating from several different sources. A selection of these projects will now be discussed briefly.

8.5.1 Cultural Heritage Projects

There are a variety of projects, both past and present, focusing on some degree to the electronic cultural heritage of Europe. These include:

- The European Library project (www.theeuropeanlibrary.org)
  - focusing on searching the content of European national libraries
- MICHAELplus (www.michael-culture.org)
  - creating a multilingual, open source platform with a search engine able to retrieve objects from cultural heritage collections across Europe
- BRICKS (www.brickscommunity.org)
  - integrating existing digital resources into a shared and common digital library
- ECHO
  - making a web-based digital library service for the historical film collections of various European national audiovisual archives
- Birth of TV project (www.birth-of-tv.org)
  - (internet archive of films from the early days of European television)

These projects have used or plan to use a variety of methods to implement their creations; however, most of them rely on exploiting metadata, thesauri, and controlled vocabularies in one way or another. Another set of related projects (SCULPTEUR and its successor, eCHASE) adopt a more advanced, ontology-based system in order to describe complex relationships and enrich the searching or browsing process.

SCULPTEUR Project

The objective of the SCULPTEUR project was "to create a distributed multimedia digital library for storing, searching and retrieving of more diverse multimedia types, with significant support for 3D objects" (www.sculpteurweb.org). It particularly focuses on "new ways to create, search, navigate, access, repurpose and use multimedia content from multiple sources over the Web" [Addis et al., 2005:1]. The project's main goal is finding new ways of searching and navigating online museum collections.

The SCULPTEUR functionalities include basic, common features such as free text search and controlled vocabulary. However, it also incorporates novel ways of searching by concept and content. The concept search is based around the use of a common ontology (CIDOC CRM), which encourages interoperability. It is meant to serve as a unifying query interface for heterogeneous databases. The CIDOC-based structure enables one to visualise the ontology itself. In addition, the interface also incorporates mSpace technology (for a sample, see http://beta.mspace.fm). MSpace facilitates the navigation of multidimensional spaces such as those provided by a given ontology; thus, it is essentially a form of faceted browsing.

With regards to searching by content, functionality provided allows users to find or compare objects based on colour, pattern, and shape. This can potentially simplify the search in various situations, depending on the searcher's objectives. Overall, it must be noted that these more advanced search features were not developed for use by the general public but rather for the interface's target audience (i.e. museum professionals or similar "power users") [Addis et al., 2005]. Other features of the SCULPTEUR interface include:

- A lightbox for storing search results
- Attribute map (graphical representation of metadata attributes)
- Results overview
- Query history

eCHASE Project

The eCHASE project draws on the past experiences of SCULPTEUR. Its objective is to create "a single, on-line site that provides a contextualized access point for the multimedia cultural content currently distributed across the museums, galleries, photo libraries and audiovisual archives of Europe" (www.echase.org).

Therefore, its mission is to link related content items from a variety of sources into a coherent whole, using aggregation and contextualization [Sinclair et al., 2005]. The eCHASE portal will focus in particular on content related to the cultural heritage of Central and Eastern European countries. Functionalities to be offered include:

- Searching and browsing of content (via text and context-based queries)
- A facility to collect and annotate objects (a lightbox)

Like SCULPTEUR, the eCHASE architecture will employ CIDOC CRM as a common metadata schema which is capable of describing complex relationships between the objects in the database. Once again, the mSpace system will be used for browsing, and as a result users will be able to navigate multi-dimensional spaces through interaction with the interface. Other functionalities the project will provide include thesaurus navigation in the form of thesaurus trees or concept hierarchies, and a geographical gazetteer for visualizing place information. The former will present the structure of the data in a way that allows users to focus queries on a specific place in a specific country. The latter will utilize Google Map technology along with latitudinal and longitudinal data to present a zoomable map of the place in which a given object was created.

8.5.2 Typical Functionality

Timelines and Maps

Both SCULPTEUR and eCHASE are similar to MultiMatch in terms of their overall scope and goals. They share characteristics such as the use of the CIDOC ontology and have similar features (i.e. a lightbox, search and browse, etc). However, some features proposed by MultiMatch go beyond the offerings of these similar projects. Differentiating features proposed by MultiMatch could include increased interactivity in browsing functionalities: for example, with the use of timelines and/or maps.

Bates, Wilde & Siegfried [1993] analysed humanities scholars' search strategies and noted that most online searches were based around subjects, as opposed to specific works or authors. Other popular search terms were related to geographical names, dates and historical periods.

According to Allen [2005: 260], while event-oriented timelines are commonly-used graphical devices, "surprisingly, only a few systems have employed *interactive* event-oriented timelines as a framework to support information access." The use of interactive timelines can be useful in the cultural heritage domain for several reasons. First, investigations in this area often incorporate elements relating to place, time, topic, and creator, with a particular interest in change over time and relationships in context [Buckland & Lancaster, 2004]. Timelines can inform, show context, encapsulate ideas, and provide contextual links [Allen, 1995]. Secondly, a visual presentation is often easier to understand than a purely textual display [Shneiderman, 1998]. Examples of dynamic timelines (some of which are linked to maps) can be seen here. Some are related to cultural heritage and others are more history-oriented.

- www.ina.fr/fresque
  (Interactive, multimedia timeline of French radio and television history)
- http://digitalhistory.uh.edu/timeline/timelineO.cfm (Integrated map and timeline relating to American history)
- www.birth-of-tv.org/birth/timeline2.do (see Figure 8.16)
  (Birth of TV project's timeline of television history)
- http://ecai.org/Area/AreaTeamExamples/Korea/tm_korea.html
  (TimeMAP visualization of Korean history: integrated map and timeline)

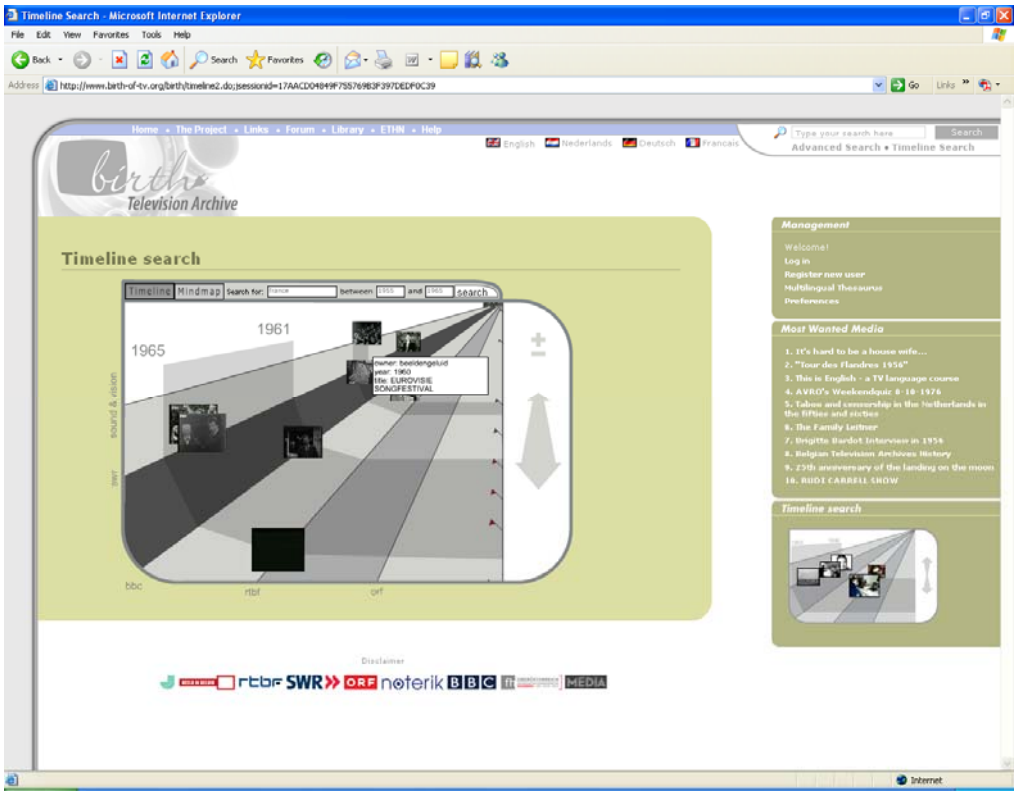


Figure 8.16: Sample interactive timeline interface (Birth of TV project)

8.6 Concluding Discussion

Current challenges in the area of online information retrieval include determining how best to classify, organise, and present objects of diverse origins and media types in a way that is intuitive and easy for the user to navigate. For example, although research indicates it may be beneficial, the use of a faceted browsing feature has not been widely adopted by most websites. Additionally, relevance feedback is often unimodal and does not always help users to specify exactly what facets they would like to use to search for similar items (i.e. colour, subject, etc). It can also be difficult to appropriately cluster results in the case where a query can have multiple meanings. With regards to cross language functionality in the form of query translation, again, this feature is not prevalent and when it is employed, it often does not function perfectly. Finally, there is the issue of the semantic gap query resolution as discussed previously. These areas have all represented potential opportunities for MultiMatch to experiment with new and potentially different means of improving the information-seeking experience. MultiMatch has exploited existing interfaces and incorporate ideas including the following:

- Faceted search and browse
- Multimodal search and reformulation (multimodal relevance feedback)
- Interactivity and exploration (variety of interaction methods)
    - Multiple ways to access the collection, i.e. multiple views (search/browse based on facets and time etc.)
    - Providing multimodal prompts, such as audiovisual surrogates (e.g. collection overviews), to assist the user in initiating searches and refining search parameters.
    - Use of workspaces, potentially to provide relevance feedback (dragging items into the workspace tells the system to “find more like this”)
- Interaction and relevance feedback (through browsing)
    - Implementation of functionality to support the formulation of multimedia queries for complex needs, an example of this would be allowing the user to input both text (e.g. “van

Gogh") along with an image selected from some pre-existing 'visual hints' gallery to assist in a query about the famous Dutch artist.
    - Implementation of some type of on-line storage facility (i.e. a "lightbox" analogous to the shopping cart on an e-commerce site) for the collection of relevant items along the way [Bates, 1989]. Items stored in such a shopping cart object could be retained or discarded as appropriate.
- Previews and overviews (dynamic queries)
    - Creating a more interactive search experience for the user
- Use of visual thesaurus to help bridge the semantic gap
    - also provides prototypical images for multimodal query expansion
- Use of multilingual thesaurus and facets
    - e.g. as implemented in the Birth of TV project
- Providing an adaptive, personalised interface
    - e.g. for images, relevance depends on work context, therefore rather than ranking images, we can create other displays and allow browsing

Overall, there are currently several related sites and projects with similar aims and functions to those of MultiMatch. However, in one sense, MultiMatch is unique in that it provides a set of characteristics (multilinguality and multimediality) that may exist elsewhere, but usually are not found together in this combination.

In the cultural heritage domain, people often use "creative and exploratory thought processes involved in translating conceptual ideas to visual instantiations" [Jörgensen & Jörgensen, 2002: 1357]. Given this, there are a number of areas where MultiMatch has endeavoured to improve upon current practices in terms of information seeking, retrieval, and presentation. For example, most multimedia or cultural heritage sites follow fairly standard ways of presenting browse and search results, even though these may not be the most effective methods of doing so. Inspired by research on alternative means of visualizing search results, MultiMatch has considered different and more interactive methods, including but not limited to clustered concept hierarchies, visual collages, fisheye views, or other methods beyond the standard thumbnail grid display.

Additionally, interactivity has been a main emphasis of the MultiMatch interface, since searching or browsing is often a fluid and evolving process in which users' needs and strategies may constantly change. How best to support these needs has been a major focus of MultiMatch which will draw on a user-centred approach to interface design that takes into account user input and requirements. Ways discussed for facilitating interaction have included the development of features for storing items and searches, refining queries, giving relevance feedback, navigating between results, and exploring relationships between items on a variety of planes.

Given that the cultural heritage field is heavily based on themes and relationships between people, places, time periods, and media, it has been necessary to consider ways of describing and navigating said relationships, be this through a more advanced type of faceted browsing, using concept maps, or including interactive means of visualizing interactions or connections over time and geographical location (e.g., seeing when, where, and by whom artworks related to Shakespeare's "A Midsummer Night's Dream" were produced.)

The present state-of-the-art research provides a variety of technological or design concepts that enable new and innovative ways of interacting with virtual objects; however, many of these have yet to be implemented in practice on a wide scale. In theory, new ideas and concepts are meant to improve upon the weaknesses of current practice, but it is not always the case that these methods are appreciated by users. Therefore, by examining and testing a variety of approaches with potential user groups, MultiMatch has endeavoured to build an interactive, innovative interface that is first and foremost successful at meeting its users' needs.

8.7 MultiMatch and the State of the Art

According to a recent survey of search engine offerings, "despite the rapid growth of multimedia data that are available from the World Wide Web, current search engines have yet to provide an exciting, intuitive and user-centred set of the functionalities that support and sustain this phenomenon" [Tjondronegoro & Spink,

2008: 356]. MultiMatch has been working towards the aim of doing just this, following a user-centred approach to design access to multimedia material (with the unique addition of cross-language search as well.) Another main focus of MultiMatch has been on content aggregation and providing a global view to heterogeneous, distributed contents enhanced by semantic links. These features match those which Hyvönen [2007] mentions as ways in which the cultural heritage domain is well suited to the construction of semantic portals.

Intermediate studies conducted throughout the design process have led to publications advancing knowledge in areas such as the effects of language skills on cross-language search [Marlow et al., 2008,] user requirements in the cultural heritage domain, video retrieval system design [Carmichael et al., 2008] and the evaluation of a faceted browser [Clough et al., 2008] The research themes, including an image collection overview, clustering of results, and dynamic summarization have explored new ways of presenting, organizing, and previewing material. Future work with these may involve user testing and evaluation to see how they are utilized in a naturalistic setting.

While MultiMatch has made headway into the exploration of a variety of topics relating to multimedia and multilingual information access and retrieval, unfortunately the scope and timescale of the project mean that all areas were not able to be extensively investigated. Future work inspired by the project could include, but is not limited to, developing tools for automatic language identification, annotation, translation, and correction of ASR output for multilingual videos; continued work with exploring the ways in which both experts and naïve users search for cultural heritage material (and ways of facilitating this); furthering knowledge of use cases of cross-language search in the cultural heritage domain; and further work with developing innovative image search and result interfaces (including multimodal search).

References

Addis et al., New Ways to Search, Navigate, and Use Multimedia Museum Collections over the Web. In J. Trant and D. Bearman (eds.). *Museums and the Web 2005:Proceedings*, Toronto: Archives & Museum Informatics, published March 31, 2005 at http://www.archimuse.com/mw2005/papers/addis/addis.html

Albertoni, R., Bertgone, A., & De Martino, M. (2005). Information Search: The challenge of integrating information visualization and semantic web. Proceedings of 16th International Workshop on Database and Expert Systems Applications, 529-533.

Allen, R.B. (1995). "Interactive Timelines as Information System Interfaces." Symposium on Digital Libraries, Japan, 175-180

ibid. (2005). "A focus-context browser for multiple timelines." Proceedings of the 5th ACM-IEEE-CS joint conference on digital libraries, 260-61.

Armitage, L. & Enser, P. (1997). "Analysis of user need in image archives." Journal of Information Science, 23(4), 287-299.

Bates, M.J. (1999). "The design of browsing and berrypicking techniques for the online search interface." Online Review, 13, 407-424.

Bates, M.J., Wilde, D.N., & Siegfried, S. (1993). "An analysis of search terminology used by humanities scholars: The Getty online searching project, report no. 1." The Library Quarterly, 63(1), 1-39.

Beale, R. (2006). "Improving Internet interaction: From theory to practice." JASIST 57(6): 829-33.

Belkin, N. J. (2003). Interface techniques for making searching for information more effective. Retrieved October 4, 2004 from http://home.earthlink.net/~searchworkshop/docs/belkin-final.pdf

Benjamins, V.R., Contreras, J., Blázquez, M., Dodero, J.M., Garcia, A., Navas, E., Hernandez, F., & Wert, C. (2004)"Cultural Heritage and the Semantic Web." In: Proc. of First European Semantic Web Symposium, 433-44.

Benini, S., Migliorati, P., Leonardi, R., (2008) "Retrieval Of Video Story Units By Markov Entropy Rate", Proc. of Sixth International Conference on Content Based Multimedia Indexing (CBMI '08), Queen Mary Univ., London.

Bernard, K. and Forsyth, D. (2001) "Learning the Semantics of Words and Pictures." In: Proceedings of the Intentional Conference on Computer Vision, 2, pp. 408-415.

Boreczky, J., Girgensohn, A., Golovchinsky, G., Uchihashi, S. (2000). "An Interactive Comic Book Presentation for Exploring Video." CHI Letters Vol. 2, No. 1

Brajnik, G., Mizzaro, S., & Tasso, C. (1996). "Evaluating user interfaces to information retrieval systems: A case study on user support." Proceedings of 19th annual SIGIR Conference on research and development in information retrieval, 128-136.

Broder, A. (2002). "A taxonomy of web search." ACM SIGIR Forum, 36(2), 3-10.

Buckland, M., & Lancaster, D.L. (2004). "Combining Place, Time, and Topic: The Electronic Cultural Atlas Initiative." Digital Library Forum (D-Lib) Magazine, 10(5), 4.

Cai, D., He, Xiaofei., Li, Zhiwei., Ma, W-Y., and Wei, J-R. (2004) "Hierarchical clustering of WWW image search results using visual, textual and link information". In: Proceedings of the 12th annual ACM international conference on Multimedia, 952-959.

Carmichael, J., Larson, M., Marlow, J., Newman, E., Clough, P., Oomen, O., and Sav, S. (2008), Multimodal Indexing of Digital Audio-visual Documents: A Case Study for Cultural Heritage Data, In Proceedings of the Sixth International Workshop on Content-Based Multimedia Indexing (CBMI2008), London, UK, 18-20th June, pp. 93-100.

Carmichael, J., Clough, P., Jones, G., Newman, E., (2008) "Multimedia Retrieval in MultiMatch: The Impact of Speech Transcript Errors on Search Behaviour", Paper accepted for the Information Access to Cultural Heritage workshop (IACH '08), Aarhus, Denmark, 18th September 2008.

Capstick, J., Diagne, A.K., Erbach, G. Uszkoreit, H., Leisenberg, A., & Leisenberg, M. (2000). "A system for supporting cross-lingual information retrieval." Information Processing and Management , 36(2), 275-289.

Chang S.F., Eleftheriadis, A., & McClintock, R. (1998). "Next-generation content representation, creation, and searching for new-media applications in education." Proceedings of the IEEE, Vol.86(5), 884-904.

Chang, M. & Leggett, J. (2003). "Collection understanding through streaming collage." In Proceedings of the Information Visualization Interfaces for Retrieval and Analysis (IVARA) Workshop, associated with the Joint Conference on Digital Libraries, Houston, Texas.

Chang, M., Leggett, J. J., Furuta, R., Kerne, A., Williams, J. P., Burns, S. A., and Bias, R. G. (2004). "Collection understanding." In Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (Tucson, AZ, USA, June 07 - 11, 2004). JCDL '04. ACM Press, New York, NY, 334-342.

Chang, S., J.R. Smith, M. Beigi & A. Benitez. (1997). "Visual Information Retrieval from Large Distributed Online Repositories." Communications of the ACM, 63-71.

Christel, M.G., Hauptmann, A.G., Wactlar, H.D., Ng, T.D. (2002). "Collages as dynamic summaries for news video." Proceedings of the tenth ACM international conference on Multimedia, 561-569.

Chu, H. (2001). "Research in image indexing and retrieval as reflected in the literature." J. Am. Soc. Inf. Sci. Technol. 52 (12), 1011-1018.

Chu, H. (2006) Information Representation and Retrieval in the Digital Age, Information Today Inc (August 2003), ISBN: 1573871729.

Clough, P., Joho, H. & Sanderson, M. (2005). "Automatically Organising Images using Concept Hierarchies." Workshop held at the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Workshop: Multimedia Information Retrieval, Salvador, Brazil.

Clough, P. and Sanderson, M. (2006). "User Experiments with the Eurovision Cross Language Image Retrieval System." In Journal of the American Society for Information Science and Technology (JASIST) Special Topic Section on Multilingual Information Systems, 57(5), 697 - 708.

Clough, P., Marlow, J., & Ireson, N. (2008). Enabling Semantic Access to Cultural Heritage: A Case Study of Tate Online, In the Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2008), Aarhus, Denmark, (In Press).

Combs, T. T. A., & Bederson, B. B. (1999). "Does Zooming Improve Image Browsing?" In: Proceedings of Digital Library (DL 99) New York: ACM, 130-137.

Cox, I. J., Miller, M. L., Omohundro, M., & Yianilos, P. N. (1996). "Target Testing and the PicHunter Bayesian Multimedia Retrieval System." Proceedings of the Third Forum on Research and Technology Advances in Digital Library (ADL'96), pp. 66-75, Washington, D.C. USA. IEEE Computer Society Press.

Del Galdo, E.M., & Nielsen, J. (1996). International User Interfaces. New York: John Wiley & Sons.

De Troyer, O., & Casteleyn, S. (2004) "Designing Localized Web Sites." In Proceedings of the 5th International Conference on Web Information Systems Engineering (WISE2004), 547–558.

Dorr, B., He, D., Luo, J., & Oard, D. (2003). "iCLEF at Maryland: Translation selection and document selection." In C. Peters (Ed.), Working Notes for the CLEF 2003 Workshop.

Eakins, J. P. (1996). "Automatic image content retrieval – are we getting anywhere?" Proceedings of Third International Conference on Electronic Library and Visual Information Research (ELVIRA3), De Montfort University, Milton Keynes, pp 123-135.

Eakins, J.P. (1998). "Techniques for image retrieval." Library and information briefings, 85, 1-15.

Eakins, J.P. (2000). "Retrieval of Still Images by Content." Lectures on Information Retrieval, Springer-Berlin/Heidelberg, pp. 111-138.

Eakins, J. & Graham, M. (1999). "Content-based image retrieval: A report to the JISC Technology Applications Programme." Technical report, Institute for Image Data Research, University of Northumbria at Newcastle.

Eakins, J. Briggs, P. and Burford, B. (2004), "Image Retrieval Interfaces: A User Perspective." Lecture Notes in Computer Science, Vol3115, pp. 628-637.

Eggink, J., Brown, G. (2004). "Instrument Recognition in Accompanied Sonatas and Concertos." ASP, 217-220.

Enser, P. (1995). "Pictorial Information Retrieval." Journal of Documentation, 51(2), 126 170.

Enser, Peter, Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science* 26(4), 2000, 199-210.

Enser, P. & McGregor, C. (1993), Analysis of visual information retrieval queries, British Library Research and Development Report, 6104.

Enser, Peter. & Sandom, Christine, Towards a comprehensive survey of the semantic gap in visual image retrieval. In: Bakker, E.M. et al. (eds.) *Image and video retrieval; Second International Conference, CIVR 2003*, Urbana-Champaign, IL, USA, July 24-25, 2003 Proceedings (Lecture Notes in Computer Science, Vol. 2728. Berlin: Springer-Verlag, 2003, 291-299.

Eurescom (2000). "Multi-Lingual Web Sites: Best Practice Guidelines and Architecture (P923)." Eurescom Project report, Available online: http://www.eurescom.de/Public/projectresults/P900-series/923d1.asp

Evans, D. (2006). "From R&D to practice – challenges to multilingual information access in the real world." Presented at SIGIR Workshop on New Directions in Multilingual Information Access (MLIA), Seattle, Washington.

Flickner M, Sawhney H, Niblack W (1995) Query by image and video content: the QBIC system. IEEE Computer 28(9), pp. 23-32.

Fluit, C., Sabou, M., & Van Harmelen, F. (2005). Ontology-based information visualization: Towards Semantic Web applications. In V. Geroimenko (Ed.), Visualizing the Semantic Web (2nd ed.): Springer Verlag.

Fluit, C., Sabout, M., & van Harmelen, F. (2003). Supporting user tasks through visualization of light-weight ontologies. In Ontology Handbook: Springer-Verlag.

Foote, J. (1999). "An overview of audio information retrieval." Multimedia Systems, 7: 2-10.

Forsyth, D. A. et al (1996) Finding pictures of objects in large collections of images, Proceedings International Workshop on Object Recognition, Cambridge, 1996.

Golovchinsky, G. (1997). "Queries? Links? Is there a difference?" Proceedings of CHI 1997, 407-414.

Goodrum, A. (2000). "Image information retrieval: An overview of current research." Informing Science, 3(2), 63-66.

Gremett, P. (2006). "Utilizing a user's context to improve search results." JASIST 57(6), 808-812.

Gudivada, V.N. & Raghavan, V.V. (1995). "Content based image retrieval systems." Computer 28(9): 18-22.

Gupta, A. & Jain, R. (1997). "Visual information retrieval." Communications of the ACM 40(5), 70-79.

He, D., Wang, J., Oard, D., & Nossal, M. (2003). "Comparing user-assisted and automatic query translation." In LNCS 2785, C. Peters, M. Braschler, J. Gonzalo, & M. Kluck (Eds.), Advances in cross-language information retrieval, 400-415.

He, D., & Oard, D. (2006). "Studying the Use of Interactive Multilingual Information Retrieval." In New Directions of Multilingual Information Access, A workshop of Annual Conference of SIGIR 2006.

Hearst, M. (1999). "User Interfaces and Visualization". In: Baeza-Yates, R. & Ribeiro-Neto, B. (eds.), Modern Information Retrieval, 257-323. New York: ACM Press.

Hearst, M., et al. (2002). "Finding the flow in web site search." Communications of the ACM, 45(9).

Henninger, S., & Belkin, N. (1996). "Interface issues and interaction strategies for information retrieval systems." Conference companion on human factors in computing systems: Common ground, 352-353.

Hildebrand, M., van Ossenbruggen, J., & Hardman, L. (2006). /facet: A browser for heterogeneous semantic web repositories. Proceedings of 5th international semantic web conference, 272-285.

Hildebrand, M., van Ossenbruggen, J., & Hardman, L. (2007). An analysis of search-based user interaction on the semantic web: Centrum voor Wiskunde en Informatica.

Hürst, W., & Venkata, L. (2003). "Interface issues for accessing and skimming speech documents in context with recorded lectures and presentations." Proceedings of HCI International 2003, pp. 656-660.

Hyvönen, E. (2007). Semantic portals for cultural heritage, http://www.seco.tkk.fi/publications/2007/hyvonen-portals-2007.pdf (Accessed 2 June 2008).

Ingwersen, P. and Järvelin, K. (2005). The turn: integration of information seeking and retrieval in context. Dordrecht, The Netherlands: Springer.

Janecek, P., & Pu, P. (2004). "Opportunistic search with semantic fisheye views." EFPL Technical Report: IC/2004/42.

Jörgensen, C., & Jörgensen, J. (2002). "Image querying by image professionals." JASIST 56(12): 1346-59.

Karadkar, U., Nordt, M., Furuta, R., Lee, C., Quick, C. (2006). "An exploration of space-time constraints on contextual information in image-based testing interfaces." ECDL 2006, LNCS 4172, 391-402.

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E. (2007). Ontology visualization methods - A survey. To appear in ACM Computing Surveys, http://sdbs.cst.uop.gr/files/onto-vis-survey-final.pdf.

Kravchyna, V. (2004). "Information needs of museum visitors: Real and Virtual." PhD dissertation, University of North Texas.

Kules, B., & Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. Information Processing and Management, 44(2), 463-484.

Lee, H., & Smeaton, A. (2002). "Designing the User Interface for the Físchlár Digital Video Library." Journal of Digital Information, 2(4).

Liu, H., Xie, X., Tang, X., Li, Z., & Ma, W. (2004). "Effective browsing of web image search results." Proceedings of MIR '04, New York, 84-90.

Logan, B., Moreno, P., Van Thong, J.M., Marston, J., & MacCarthy, G.(2004). "NewsTuner: A simple interface for searching and browsing radio archives." IEEE International Conference on Multimedia and Expo (ICME).

Lombardi, T., Cha, S., & Tappert, C. (2004). "A graphical user interface for a fine-art painting image retrieval system." Proceedings of the 6th ACM SIGMM international workshop on multimedia information retrieval, 107-112.

Maedche, A., & Staab, S. (2002). Applying semantic web technologies for tourism information systems. 9th International Conference for information and communication technologies in tourism (Enter-2002), 311-319.

Marchionini, G. (1992). "Interfaces for end-user information seeking." Journal of the American Society for Information Science, 43(2):156-163.

Marchionini, G. (1995) Information Seeking in Electronic Environments, Cambridge University Press (May 26 1995), ISBN: 0521443725.

Markkula, M., & Sormunen, E. (2000). "End-user searching challenges: Indexing practices in the Digital Newspaper photo archive." Information Retrieval, 1(4), 259-285.

Marlow, J. (2006). "Designing a localisation strategy for Tate Online: requirements and recommendations." MSc dissertation for Master of Arts in Multilingual Information Management, University of Sheffield.

Marlow, J., Clough, P., Cigarrán Recuero, J. and Artiles, J. (2008). Exploring the Effects of Language Skills on Multilingual Web Search, In Proceedings of the 30th European Conference on IR Research (ECIR'08), Glasgow, UK, April 2008, LNCS 4956, pp. 126-137.

Minerva Project (2006). "Multilingual Access to the digital European cultural heritage." http://www.mek.oszk.hu/minerva/survey/delir20060130. [Accessed 16 August 2006]

Mostafa, J. (1994). "Digital image representation and access." In M.E. Williams (Ed.), Annual Review of Information Science and Technology, vol. 29.

Oard, D. W. (1997). "Serving Users in Many Languages: Cross-Language Information Retrieval for Digital Libraries." D-Lib Magazine, December 1997.

Oard, D., & Gonzalo, J. (2002). "The CLEF 2001 Interactive Track." Evaluation of Cross-Language Information Retrieval Systems, Springer-Verlag LNCS 2406.

Oard, D., Gonzalo, J., Sanderson, M., López-Ostenero, F., & Wang, J. (2004). "Interactive Cross-Language Document Selection." Information Retrieval, Vol. 7 (1-2), 205-228.

Oard, D., He, D., & Wang, J. (2008). User-assisted query translation for interactive CLIR. Information Processing and Management, 44(1), 181-211.

Ogden, W., Cowie, J., Davis, M., Ludovik, E., Nirenburg, S., Molina-Salgado, H., et al. (1999). "Keizai: An interactive cross-language text retrieval system." Paper presented at the Machine Translation Summit VII, Workshop on Machine Translation for Cross-Language Information Retrieval, Singapore, PRC.

Ogden, W.C., & Davis, M.W. (2000). "Improving cross-language text retrieval with human interactions." Proceedings of the Hawaii International Conference on System Scuence (HICSS-33), Vol. 3.

Panofsky, E. (1955). Meaning in the Visual Arts: Papers in and on art history. Garden City, NY: Doubleday Anchor Press.

Park, G., Baek, Y., and Lee, H-K. (2005) "Re-ranking algorithm using post-retrieval clustering for content-based image retrieval." Information Processing and Management, 41(2), 177-194.

Parker, E. B. (1987), LC Thesaurus for Graphic Materials: Topical Terms for Subject Access. Washington, D. C., Library of Congress.

Peñas, A., Gonzalo, J., & Verdejo, F. (2001). "Cross-language information access through phrase browsing." Paper presented at the 6th International Conference of Natural Language for Information Systems (NLDB'01), Madrid, Spain.

Peters, C. & Sheridan, P. (2001) "Multilingual information access." In Lectures on information Retrieval, M. Agosti, F. Crestani, and G. Pasi, Eds. Springer Lecture Notes In Computer Science Series, vol. 1980. Springer-Verlag New York, New York, NY, 51-80.

Petersen P. and Barnett P. (1994). Guide to indexing and cataloguing with the Art & Architecture Thesaurus, The Getty Art History Information Program. OUP.

Petrelli, D., Hansen, P., Beaulieu, M., Sanderson, M. (2002). "User requirement elicitation for Cross-Language Information Retrieval." New Review of Information Behaviour Research, 3, 17-35.

Petrelli, D., P. Hansen , M. Beaulieu, M. Sanderson, G. Demetriou, P. Herring. (2004). "Observing Users - Designing Clarity: A Case study on the user-centred design of a cross-language retrieval system." JASIST, 55(10), 923-934.

Petrelli, D., Levin, S., Beaulieu, M., & Sanderson, M. (2006). "Which User Interaction for Cross-Language Information Retrieval? Design Issues and Reflections." JASIST - special issue on "Multilingual Information Systems". 57(5), 709-722.

Petrelli, D. and Clough, P. (2005) Concept Hierarchy across Languages in Text-Based Image Retrieval: A User Evaluation, In the working notes of the CLEF workshop, Vienna, Austria, 21-23 September 2005, online.

Rasmussen, E.M. (1997) Indexing Images, Annual Review of Information Science and Technology (ARIST), Vol. 32, pp. 169-196.

Rautiainen, M., Seppänen, T., & Ojala, T. (2006). In Advancing content-based retrieval effectiveness with cluster-temporal browsing in multilingual databases (pp. 377-380). Paper presented at the 2006 IEEE International Conference on Multimedia & Expo, Toronto.

Renals, S., Abberley, D., Kirby, D., & Robinson, T. (2000). "Indexing and Retrieving Broadcast News." Speech Communication, 32, 5-20.

Resnick, P. (1997). "Evaluating Multilingual Gisting of Web Pages in Cross-Language Text and Speech Retrieval." AAAI Technical Report SS-97-05.

Resnick, M., & Vaughn, M. (2006). "Best practices and future visions for search user interfaces." JASIST 57(6): 781-787.

Rodden, K., Basalaj, W., Sinclair, D., & Wood, K. (2001). "Does organisation by similarity assist image browsing?" Proceedings of SIGCHI Conference on Human Factors in Computing, 190-197.

Rorvig, M. (1988). "Psychometric measurement and information retrieval." In M.E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST), 23, 157-189.

Rose, D.E. (2006). "Reconciling information-seeking behaviour with search user interfaces for the Web." JASIST 57(6): 797-799.

Rose, D. and Levinson, D. (2004). "Understanding User Goals in Web Search. In Proceedings of WWW 2004," New York, USA. ACM.

Rui, Y, Huang, T., Ortega, M., and Mehrota, S. (1998). "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval." IEEE Trans. on Circuits and Systems for Video Technology , Special Issue on Segmentation, Description, and Retrieval of Video Content, 8 (5), 644-655.

Rui, Y., & Huang, T. (1999). "A novel relevance feedback technique in information retrieval." Proceedings of the 7th ACM international conference on Multimedia, 67-70.

Rui, Y., Huang, T., & Mehrotra, S. (1997). "Content based image retrieval with relevance feedback in MARS." Proceedings of the International Conference on Image Processing, 815-818.

Shneiderman, M. (1998). Designing the User Interface: Strategies for Effective Human- Computer Interaction. Reading, MA: Addison Wesley.

Sinclair, P., et al. (2005). "eCHASE: Exploiting Cultural Heritage using the Semantic Web." In Proceedings of the 4th International Semantic Web Conference, ISWC 2005, Galway.

Smeaton, A. (2000). "Indexing, Browsing and Searching of Digital Video and Digital Audio Information." ESSIR 2000, LNCS 1980, 93-110.

Smeaton, A. (2002). "Challenges for content-based navigation of digital video in the Físchlár Digital Library." LCNS 2383, 215-224.

Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). "Content-based image retrieval at the end of the early years." Pattern analysis and Machine Intelligence, 22(12), 1349-1380.

Smith J. and Chang, S (1997). "An Image and Video Search Engine for the World Wide Web." Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases V, 84-95.

Snoek, C., & Worring, M. (2005). "Multimodal video indexing: A review of the state-of-the-art." Multimedia Tools and Applications, 25, 5-35.

Spoerri, A. (2004). MetaCrystal: Visualizing the degree of overlap between different search engines. 13th annual World Wide Web Conference on Alternative track papers and posters.

Storey, M., Noy, N., Musen, M., Best, C., Ferguson, R., & Ernst, N. (2002). Jambalaya: An interactive environment for exploring ontologies. Proceedings of 7th international conference on intelligent user interfaces, 239.

Swain, M., Frankel, C. and Athitsos, V. (1996). "WebSeer: An image search engine for the World Wide Web." Technical Report TR-96-14, Department of Computer Science, University of Chicago.

Tsamoura, E., Mezaris, V., Kompatsiaris, I., (2008) "Video Shot Meta-Segmentation Based On Multiple Criteria For Gradual Transition Detection", Proceedings of the Sixth International Conference on Content Based Multimedia Indexing (CBMI '08), Queen Mary University, London.

Tjondronegoro, D., and Spink, A. (2008). Web search engine multimedia functionality, Information Processing and Management, 44(1), 340-357.

Tucker, S., & Whittaker, S. (2006). "Time is of the essence: An evaluation of temporal compression algorithms." Conference on Human Factors in Computing Systems (CHI), Montreal, Canada.

Turner, J. (1994). "Determining the subject content of still and moving image documents for storage and retrieval: an experimental investigation." PhD thesis, University of Toronto.

Urban, J. and Jose, J.M. (2005). "Exploring results organisation for image searching." In Proceedings of INTERACT 2005: human-computer interaction, LNCS1973, v.3585, 958-961.

Van Houten, Y., Schuurman, J., & Verhagen, P. (2004). "Video content foraging." CIVR Proceedings, pp. 15-23.

Van Thong, J., Moreno, P., Logan, B., Fidler, B., Maffey, K., & Moores, M. (2001) "SPEECHBOT: An experimental speech-based search engine for multimedia content in the web." Cambridge Research Laboratory Technical Report Series CRL 2001/06.

Veltkamp, R., & Tanase, M. (2000). "Content-based image retrieval systems: A survey." Technical report UU-CS-2000-34, Department of Computer Science, Utrecht University.

Venters, C., Eakins, J. and Hartley, R. (1997). The user interface and content-based image retrieval systems, 19th Annual BCS-IRSG Colloquium on IR.

Voorhees, E.H. and Harman, D (2000). Overview of the sixth text retrieval conference (TREC-6). Information Processing and Management. 36. 1, pp. 3-35.

Wang, Y., Liu, Z, Huang, J. -C., (2000) "Multimedia content analysis using both audio and visual clues", IEEE Signal Processing Magazine, vol. 17, no. 11, pp. 12–36, Nov. 2000.

W3C (2003) W3C FAQ: International and Multilingual websites, http://www.w3.org/International/questions/qa-international-multilingual

West, K. and Cox, S.J., (2005). "Finding an Optimal Segmentation for Audio Genre Classification." In Proc. 6th International Conference on Music Information Retrieval (ISMIR 2005), London.

White, R. W. and Ruthven, I. (2006). "A study of interface support mechanisms for interactive information retrieval." JASIST 57(7), 933-948.

Whiting, M.A., & Cramer, N. (2002). WebTheme: Understanding Web information through visual analytics. LNCS, 2342, 460-468.

Yang, M., Wildemuth, B.M., Marchionini, G., Wilkens, T., Geisler, G., Hughes, A., Gruss, R., & Webster, C. (2003). "Measuring user performance during interactions with digital video collections." ASIST 2003 Contributed Paper, 3-11.

Yang, M., & Marchionini, G. (2005). "Deciphering visual gist and its implications for video retrieval and interface design." CHI '05 extended abstracts on Human factors in computing systems, 1877-1880.

Yeo, B., Yeung, M. (1997). "Retrieving and Visualizing Video." Communications of the ACM, 40 (12), 43-52.

Yunker, J. (2003). Beyond borders - Web globalization strategies. Indianapolis, IN: New Riders Publishing.

Zhang, P., Plettenberg, L., Klavans, J., Oard, D., & Sorgel, D. (2007). Task-based interaction with an integrated, multilingual, multimedia search engine: A formative evaluation. Proceedings of 2007 ACM/IEEE joint conference on digital libraries, 117-126.

Acknowledgments

The authors would like to thank all their colleagues in the MultiMatch project for many useful discussions and much input.

In addition, we must express our immense gratitude to the following external reviewers for having accepted to read through and comment earlier versions of this report:

- Antonella Fresa, Technical Coordinator of MICHAEL project, for the chapter on Cultural Heritage Technologies
- Nicholas Kushmerick, QL2 Software Inc., Seattle, USA (previously Dublin City University, Ireland), for the chapter on Information Extraction and Classification
- Arjen de Vries, Centrum voor Wiskunde en Informatica (CWI), The Netherlands, for the chapter on Multilingual/Multimedia Indexing.
- Douglas W. Oard, University of Maryland, for his suggestions for the chapter on Multilingual/Multimedia Indexing, some of which will be added in a future revision of this report.
- Jussi Karlgren, Swedish Institute of Computer Science (SICS), for the chapter on User Interaction and Interfaces

And finally Costantino Thanos, ISTI-CNR, for having the patience to read and provide valuable feedback with respect to the entire report.