



Project no. 033104

# MultiMatch

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

# **D7.4 Results of field trials**

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

Organisation Name of Lead Contractor for this Deliverable: WIND

Version: Final

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





# **Document Information**

Deliverable number:	D7.4
Deliverable title:	Results of field trials
Due date of deliverable:	September 2008
Actual date of deliverable:	January 2009
Author(s):	M. Spadoni, A. Reale, F. Tavanti, P. Savino, W. van den Heuvel, M.
	Gonzalves, J. Oomen.
Participant(s):	Alinari, B&G, BVMC, ISTI-CNR, WIND
Workpackage:	WP7
Workpackage title:	Evaluation and Field trials
Workpackage leader:	ISTI-CNR
Dissemination Level:	RE
Version:	1.0
Keywords:	Field trials, methodology

# **History of Versions**

Version	Date	Status	Author (Partner)	Description/Approval Level
0.1	04/12/2008	Early draft	Wind	
0.2	12/12/2008	Draft	Wind	
0.3	15/12/2008	Draft	IST-CNR, Wind	General revision, addition to 2.2, update of chap. 4
0.4	16/12/2008	Draft	BVMC	Improvements, re-arranging of chapters
0.5	17/12/2008	Pre final	Wind	Added users reponses
0.6	26/12/2008	Pre final	B&G	Added conclusions
0.9	29/12/2008	Pre final	B&G	All input gathered, ready for internal review
1.0	05/01/2009	Final	B&G	Final version





# Abstract

This document reports the results of the evaluation of the second MultiMatch system prototype (P2) in three different field trials. The end-user scenarios and the tasks set to test the system are described. The feedback from the users, gathered through interviews and using a questionnaire is presented and general conclusions are made. Overall, users were positive and appreciated the special information retrieval functionalities provided by the system.





# **Table of Contents**

Abstract	
Executive Summary	
<ol> <li>Introduction</li> <li>1.1. Some notes on user evaluation of information retrieval systems</li> </ol>	
<ol> <li>Defining use cases</li> <li>2.1. User groups</li> <li>2.2. Field trial execution procedure</li> </ol>	<b>10</b> 
<ol> <li>Qualitative analysis</li> <li>Field trials2<sup>nd</sup> MultiMatch prototype (B&amp;G)</li> <li>Field trials 2<sup>nd</sup> MultiMatch prototype (Alinari)</li> <li>Field trials 2<sup>nd</sup> MultiMatch prototype (UA)</li> </ol>	<b>12</b> 12 13 13
<ul> <li>Quantitative analysis</li> <li>4.1.MultiMatch Overview statistics</li> <li>4.2. MultiMatch Audio Task statistics</li> <li>4.3. MultiMatch Image Task</li> <li>4.4. MultiMatch Multilingual Task statistics</li> <li>4.5. MultiMatch Video Task statistics</li> <li>4.6. MultiMatch Web Task statistics</li> <li>4.7. MultiMatch Post Field Trial Questionnaire</li> </ul>	<b>17</b> 18 19 21 24 24 24 26 27
<ul> <li>5. MultiMatch availability on the web for general users</li> <li>5.1. Indicators</li></ul>	<b>32</b> 33 33 34 34 34 34 34 36
<ul> <li>6. Conclusions</li> <li>6.1. Cultural Heritage user group</li> <li>6.2. Tourism user group</li> <li>6.3. Education</li> <li>6.4 Final remarks</li> </ul>	<b>39</b> 39 40 40 40 41
References	
Annex 1. Presentation slides field trials	
Annex 2. Set of guided tasks	





# Abbreviations and Acronyms

B2B	Business to business
B2C	Business to consumer
CA	Consortium Agreement
СН	Cultural Heritage
DRM	Digital Rights Management
ICT	Information and Communication Technologies
IP	Intellectual Property
IPR	Intellectual Property Rights
IR	Information Retrieval
IST	Information Society Technologies
SME	Small and Medium Enterprise
STEP	Sociological, Technical, Economic and Political
SWOT	Strengths, weaknesses, opportunities, and threats





# **Executive Summary**

The aim of the MultiMatch field trial activity is to ensure that the Second Prototype is evaluated from a user-centred point of view. Time constraints have meant that evaluation has mainly focused on those areas where ground truth in some appropriate form already exists.

It must be noted that these field trials focussed on testing a system prototype of a complex system which was still being subjected to refinements and debugging both with respect to single components and the overall integration.

The deliverable is divided into six chapters. The first chapter outlines the methodological approach chosen for the task. MultiMatch Field trials are based on the execution of a certain number of use cases, which are described in chapter 2:

- Education: use of MultiMatch in the Humanities.
- Tourism: finding cultural events related to locations.
- Cultural Heritage: evaluate MultiMatch with users of Alinari, BVMC, Sound and Vision.

Each Field trial (FT) attendee performed a set of tasks, (i.e. sequence of operations) chosen according to the User Group to which the attendee belonged.

Chapter 3 describes the tasks used during the execution of the FTs. Testing was conducted in a controlled environment at the sites of the three MultiMatch CH institutions in The Netherlands, Italy and Spain. Chapter 4 reports the statistics collected during the execution of the trials using the survey tool SurveyMonkey and of the questionnaires filled by the users after their testing of the MultiMatch system.

As part of the evaluation, users of WIND Libero portal (with 28 million registered users) have been invited to use the MultiMatch Search engine. Chapter 5 reports the results of this evaluation and presents quantitative indicators, such as capture ratio, MultiMatch index coverage, response-time distribution, etc.. Chapter 6 makes some conclusions regarding the overall evaluation of the system.

Users provided comments regarding both performance problems and also comments regarding the new functionality offered and the potential of the system. In our opinion, the second set of comments bear far more weight, especially in the light of future work and experimentation

Overall, the innovative features of the system (i.e. the ones that make the MultiMatch system truly stand out) are highly appreciated, as evidenced in Paragraph 4.6 for example.

Most users fel that MultiMatch compares very well with other tools they are currently using. Users found the interface intuitive. Although participants had some problems with the use of the language tools, they did see a lot of potential for them. Improving the way multilingual search is supported by the system will be a priority in future research.







The ability to perform content-based image and video search would be highly appreciated by users. The stability of the video component should be improved.

Users also liked the idea of searching within different collections and through different types of content. Multimedia (image, speech, audio, video) and multilingual functionalities would enable them to easily perform their daily work, helping them to retrieve relevant information. If a system like MultiMatch contained more collections, it would certainly be used in education next to textbooks in education, notably in the humanities.

Overall, the results of the trials were really very positive; all users, even those that had experienced some difficulties in using certain system functionalities, declared (in one way or another that they found the system very interesting and with considerable potential.





# 1. Introduction

MultiMatch was created following a user-centered design method, widely recognized as a tried and trusted technique to achieve usable systems. The first prototype was designed after a thorough investigation of user requirements, and in designing the second prototype, the consortium made use of the results of the first evaluation cycle. For Prototype 1 extensive user testing was conducted electronically with users from a number of different countries, by requesting them to test the system in different scenarios and submitting their evaluations via on-line questionnaires.

As the second prototype is now operational, the usability of the MultiMatch system can be evaluated. Usability is usually defined as a collection of attributes or aspects. Nielsen [1993] for example, defines usability as "consisting of learnability, efficiency, memorability, errors and satisfaction". His model of the attributes of system acceptability is shown below.



The methodology adopted for the field trials of prototype 2 involved system testing in a controlled environment and post-test face-to-face interviews at the sites of the three MultiMatch cultural heritage institutions in The Netherlands, Italy and Spain.

The usability of the MultiMatch prototype was tested in two different ways. Field trials were used to:

- 1. test the system through a series of guided tasks in order to evaluate the system from a usercentred point of view.
- 2. judge the relevance of a system like MultiMatch for a more generic user group.

Data collection and analysis methods have been based on those presented by [Hansen, 1998]. These methods have included observation and measurement of objective criteria, along with questionnaires, in order to collect a combination of quantitative and qualitative data.

In order to provide instructions and guidelines and to ensure that each group conducted the field trials in a similar manner an internal document describing a strategy for the field trials was produced, circulated and approved by all the partners involved in the Field trials . Users were presented with various scenario-based tasks to complete. Data was collected regarding the ability of users to complete the tasks and to find the requested information. Further subjective measures relating to user satisfaction (including system complexity, speed, coverage of results, etc.) were gathered through the completion of a short questionnaire following each task and a Post Field trial Questionnaire.





## 1.1. Some notes on user evaluation of information retrieval systems

Current research reports that actually trying to quantify the performance of a system on the basis of user satisfaction is very hard because people tend to discount the contribution of a system when things are going well, and then blame the system when things don't go well. [Hufnagel 1990, Bengts 2004]. MultiMatch offers some features not present in other IR (information retrieval) systems. Users participating in the evaluation only had a short time to get acquainted with these features. During the evaluation, it became clear how some features had quite a steep learning curve, suggesting users would have performed better if they had had more time to work with the system.

Not all attributes in Nielsens definition (above) can be operationalised and measured objectively. When users are at work, it can be observed whether a system is effective, learnable, memorable, safe, and efficient. However, some of these indicators are operationalized through lower level components. Satisfaction, for its part, is more difficult to observe. It relies on opinion and has to be elicited while the cognitive aspects can be measured. Attractive things work better" and understandable things are more attractive, because the affect system influences the cognitive system and the cognitive system can to some extent decide about affects. [Norman 2002] When users are at work, it can be observed whether a system is effective, learnable, memorable, safe, and efficient. However, some of these indicators are operationalized through lower level components. Satisfaction, for its part, ismore difficult to observe. It relies on opinion and has to be asked while the cognitive aspects can be measured. Developers need to understand which of the goals are important and which can be ignored depending on the nature of the system being developed. Achievement of these goals is verified by asking the users for their opinion. Utility, for its part, can be determined without test users by comparing the requirements of the work and the functionality of the system provided that the requirements are known.

Zhang notes how "the implementation of complex computer technology creates an environment that is generally characterised by users as unpredictable, uncontrollable and uncertain" [Zhang 2006]. These studies conclude that user satisfaction is not an adequate measure of assessing the effectiveness of an IR system. They found that computer-related performance levels are attributed to bad luck, system quality, task difficulty, effort, and instruction support. It needs to be noted this is a highly debatable area of research, but it does put the outcomes of the MultiMatch system in perspective. The chapters that follow offer a detailed, quantitative and qualitative account of the extended trials carried out, but it needs mentioning that MultiMatch:

- 1. offers users possibilities they have not experienced in any other IR system previously and
- 2. is a prototype and not a run-time system and, as was to be expected, some technical 'hitches' occurred during the trials.





### 2. Defining use cases

The Field trials are based on the execution of five use cases: audio, image, multilingual, video and web; the most prominent features of the system. Use cases are related to the following scenarios:

- Education: use of MultiMatch in the Humanities –BVMC was responsible for creating these use cases
- Tourism: finding cultural events related to locations Sound and Vision was responsible for creating these use cases.
- Cultural Heritage: evaluating MultiMatch with users of Alinari, BVMC, Sound and Vision -Alinari was responsible for creating these use cases.

Beside the description of the scenario, each use case contains at least three tasks to be carried out by the participants. For each task the following is described:

- the information that must be obtained by the participant;
- the operations that need to be performed to obtain th information.
- All the users executed the same FT scenario, so that the results could be comparable over all users surveyed. (See the annex for the task descriptions.)

### 2.1. User groups

User groups have been identified in accordance with the scenario related to field trial use cases. The number of participants were:

- Education: 20 users
- Tourism: 9 users
- Cultural heritage: 26 users

The spread of these user groups was about the same across the trials executed in Italy, The Netherlands and Spain. Two types of Field trials were performed: Face-to-face and Remote. The majority were conducted in a face-to-face mode in a controlled environment and post-test face-to-face interviews at the sites of the three MultiMatch cultural heritage institutions. Only a minority of field trials were made in remote mode.

### 2.2. Field trial execution procedure

The field trials followed a standard procedure:

1. A presentation of the MultiMatch system (see Annex 1) was used to provide the users with an overview of the prototype. This presentation includes a set of screen shots to show:

- Overview of system and main functionality.
- Details of the Mono- and Cross-language Retrieval Functionality (text retrieval in various languages).
- Details of the Multimedia Functionality (image, speech, video retrieval).

2. Participants were then asked to perform a set of guided tasks in order to get acquainted with the system. (these are the equivalent to very simple use cases). These consisted of two separate sets of tasks designed to enable them to get a feeling of:

- the mono- and cross-language functionality.
- the multimedia search functionality.





3. After the task driven exploration of the system, both groups were asked to test the system with a set of their own queries (e.g. 5) according to their particular interests.

4. At the end of the hands-on session, respondents were asked to complete a final online questionnaire. This questionnaire, defined in the field trial strategy document, was implemented using the Internet tool SurveyMonkey. Users were also interviewed, in order to gather general comments on the system functionalities and usability aspects. Pre-trial instructions and post-trial questionnaires were written in the native languages of the users.







# **3.** Qualitative analysis

This section describes the type of trials conducted at each site and reports the general comments gathered from users during the interviews conducted at the end of the field trial.

# **3.1.** Field trials2<sup>nd</sup> MultiMatch prototype (B&G)

### **USER GROUP**

- 5 employees from S&V from various departments (Cultural Heritage)
- 4 students from a Multimedia Design Academy (Education)
- 1 high school teacher (Education)
- 5 general users (Tourism)

### **TYPE of TRIAL**

- 7 Face-to-face trials: Face-to-face interviews at the S&V premises. The trials were performed in a controlled environment (respondents had the guidance of the interviewer). The trials were performed in two 2 pc rooms (all Microsoft).
- 8 Remote trials: Description and urls of the tasks and questionnaire (with Dutch translations) and the presentation of the MM system were previously sent to the users. Assistance was provided through mail and phone.

### **COMMENTS from users:**

### Interface

- The interface looks nice; however the translation of the interface into Dutch doesn't work properly. "Search in a specific language" was not translated.
- Languages tools next to the search bar need some time to figure out.
- Browser is not always Mozilla compatible.

### Register tool:

- Not every user registered, but most of them did.
- User found this tool very easy and fast to use but in some cases it didn't work.

### Overview screen:

- The system was slow, so users often would leave the search engine before getting any results.
- Some of the results did not seem relevant, like rss and creator.
- The typical search tools like AND/OR don't seem to work.

### Language functionalities:

- Users needed some time to figure this out.
- Translations aren't always correct.
- Searching in a single language gave more results than searching in all languages.
- Searching in Spanish was difficult for a few users, even with the translation tool.

### Faceted search:

 Users had trouble with this task, because of the Spanish and the lack of results of the Dutch translations.

### Web:

• No particular problems.





### Video:

- The word Snippet is not very common around here. It took some time for people to figure out what it meant. One user said an abstract would be helpful.
- The button "metadata" didn't contain any useful information according to some users.
- Real player can't always be installed.
- Translation of the video titles from Dutch to English isn't always correct.

### Audio:

- Tags aren't always correct. Some users would like to alter them freely.
- They did like the way they could browse through the transcript and the tags.

### Image:

• No particular problems.

### Generally:

Participants liked the idea of searching through different sources of content. The video and audio search tool could be very useful, especially for the Cultural Heritage group and the Education group. However, the system was low and even broke done a couple of times. This was frustrating for the users. 2 users didn't have enough time to finish the tasks.

# **3.2.** Field trials 2<sup>nd</sup> MultiMatch prototype (Alinari)

### **USER GROUP**

14 users (13 Italian users, 1 Canadian user)

- Employees from photo library department of ALINARI 24 ORE (Cultural heritage)
- Employees from educational department of ALINARI 24 ORE (Education)
- 1 external user belonging to a Canadian CH institution (expert in usability test for database interfaces)

### TYPE of TRIAL

- N° Face-to-face trials : 13
- N° Remote trials : 1
- Face-to-face interviews at Alinari company. The trials were performed in a controlled environment (respondents had the guidance of the interviewer). The trials were performed using 4 pc (3 Microsoft-1 Mac)
- Remote trial: URLs of the tasks and slides of the presentation of the MM system were previously sent to the user. Assistance during the trial was given by email.

### **COMMENTS from users:**

### Interface:

- Users liked the interface.
- Some buttons were not intuitive (i.e. the MultiMatch logo used to go back: this was misunderstood).
- Browser is not Safari compatible.
- The translation of the interface language to Italian presented some mistakes:
  - One part has not been translated to Italian: "search in a specific language".





• "Visione dinsieme" should be corrected with "Visione d'insieme" or "visione generale" (if the apostrophe is not allowed).

### Register tool:

• Almost all the users registered to the system but twice it did not work.

### Overview screen:

• One user stated: " usually there should be some visual indicator where I am when there are similar functions. For example, it says 'search in a specific' language'. Does that mean this is where I am (the current function), or if I click on it I will arrive at that search type? It is unclear. Is the default 'search all languages'? This button seems to toggle between the two possibilities, but it is not clear what option you are at. It would be nice if something provided a visual indicator of that".

### Language functionalities:

- Users liked the possibility to search in different language collections and translate the query.
- The distinction between target language and source language was not intuitive to the users
- Some users found the multilingual tasks the easiest ones to achieve.
- One user quotes:" I found out as I proceeded through the tasks that my previous queries were being carried forward, but I had not expected that once I retrieved results. It would be nice to have a clear query button if you are starting from scratch. The only way I found to do this was to go back to the Overview page"

### Faceted search:

- The instructions were not clear.
- One user stated: "I typed 'teatro Shakespeare noticias' in the search box and clicked on 'Search in a specific language' and selected English as my language (language I write in) I did not get the option to select Spanish as target language. (until the search results -- is that what was meant in the instruction? It was on the results page, but I understood from the instructions that it would be on the interface.) If this refers to choosing the language in a search result set, then, yes I found this useful".

### Web:

• In some cases it was not possible to answer the query without changing the source language.

### Video

- This was the most problematic. Users encountered problems in performing this task, and some did not get any result:
- The key word search did not work.
- One user comments "I thought perhaps the problem was that the transcript was in Dutch and I was using English terms, but I tried Dutch terms (I think I had the right terms) and still received zero results. To do the next task, I managed to open the video player, entered my username and password, but then the system got hung up and I gave up. All that is to say I can't comment about the other parts of the video task because once the system hung up I couldn't try anything".

### Audio:

- Users liked the audio transcripts and the browse.
- Some users encountered problems with the results: performing the same query as the other users they did not find the same results so they were not able to find answers to the queries.

Image:

• Generally users liked this type of search and found this task easy to achieve.





- The results were shown in a scale from 1 to 10 but if the user clicks to view the following pages the same results of page 1 are shown.
- Some users misunderstood the query: they typed "church" instead of "church tower" retrieving the results for church.

### General considerations:

Users liked the concept of MultiMatch and think it will be very powerful. One user quotes: "I love the concept of MultiMatch and am quite excited by the possibilities presented here". Users encountered some difficulties, some of which they were able to sort out: "the system, while complex, is learnable, because once you figured it out you could continue and move along more quickly". In the case of video searching users were unable to complete the task. The slowness of the system made the performance of the tasks more problematic.

# **3.3.** Field trials 2<sup>nd</sup> MultiMatch prototype (UA)

The field of work of participants at the University of Alicante was mainly cultural heritage in 55% of cases and educational in 28%. Around 16% were working in the tourism area (in some cases, as professors in Tourism Studies at the Universidad de Alicante). The trials were performed face-to-face in 10 cases and were remote in 15 cases. Partly due to their wide background, no substantial differences in the answers were detected depending on the expert's provenance. Only two users registered in the system, as this was left as an optional action. The answers to the closed questionnaire are summarized at the end of this section.

Useful feedback was obtained from the open questions, even if some users claimed that they could not address all the questions (especially those involving video or audio tasks) in the allocated time due to occasional problems with internet access through port 8090 and sometimes slow response times. In general, tests were considered a bit lengthy to be addressed by volunteers.

### **USER GROUP**

- 4 Tourism
- 7 Educational
- 14 Cultural Heritage

### TYPE of TRIAL

- N° Face-to-face trials: 10
- N° Remote trials: 15

### **COMMENTS from testers:**

Interface:

• Easy to use, but difficulties in language. Language used: Spanish/ English.

### Register tool:

• One or two participants used this tool.

Overview screen:

- Generally, language interface is not very clear since translations are not always correct. Internalization errors.
- Results are not justified at URL or at snippet.
- Cloud terms motors do not show relevant relations.





### Language functionalities:

- Not working properly. Ambiguity in source and target language. Language should be automatically identified.
- Query should be automatically translated into the other languages and show results to use tabs and segment results. Otherwise, searches are very difficult with source, target and filtered options.

### Faceted search:

No particular problems.

### Web:

• Accentuation and spelling problems.

### Video:

 Questions were not completely clear. In some cases, users were not able to complete the tasks properly.

### Audio:

Questions were not completely clear. In some cases, users were not able to complete the tasks properly.

### Image:

• Ambiguity in term "collection" in this trial. However, this exercise was easier to carry out than audio and video trials.

### General considerations:

Multimatch did not work in some search engines, like Safari. Some testers had problems with port net:8090 on MM URL. Search results are very slow, so telephone interviews took more than two hours and some users had problems completing the tasks . As stated before, tests were too long and not properly performed.





# 4. Quantitative analysis

The following paragraphs represent the data collected by:

- 1. examining the answers to the given tasks.
- 2. the post-test questionnaire.

Each paragraph reports the results obtained for a specific functionality of the MultiMatch prototype system. Paragraph 4.1. provides the overall result of the tasks. Paragraph 4.2. provides the results of the Audio Task, 4.3. of the Image Task, 4.4. of the Multilingual Task, 4.5. of the Video Task and paragraph 4.6. of the Web Task. These questionnaires were inserted in SurveyMonkey.com and the users were directed to link to this website to get guides for the tasks (these guides are reported in Annex 2) and questionnaires.

SurveyMonkey offers the possibility to create professional online surveys quickly and easily and to collect the responses given online. The questionnaires allow us to understand whether users were able to complete the tasks and to have a feedback about their evaluation of different system functionalities. Paragraph 4.7. reports the statistics of the post trials questionnaires compiled by the users and gives a quick overview of their feedbacks.

The following bar charts report an estimation of users objective assessment on the different system functionalities and users subjective assessment. Objective assessments were collected through the responses provided by users to obtain specific results; for example, on the Web task the user had to search for information about W.A. Mozart, and provide biographical information (Using MultiMatch and its related pages) about Mozart sisters name, the number of pieces he composed, etc. By comparing the number of correct and wrong responses it is possible to estimate the objective usefulness of the system. In case of the overview statistics, the estimation was done by counting all provided responses.

For specific tasks, only the responses provided for that task were considered. Subjective assessment is based on users responses to questions on quality of system results. For example, for the Web task question "How easy was to find Mozart sisters name?" the user had seven possible choices to evaluate the system. For specific tasks we asked users to rate the usefulness of specific functionalities with questions like "How helpful were the following features?" We also proposed possible enhancements of system functionalities and required users to assess how much they would like to have them included.





# 4.1.MultiMatch Overview statistics



Users answered 90% of the questions during the tasks and 72.4% of the answers were correct. This is a good indication of the usability of the system and the results show that MultiMatch is a user-friendly system.



The percentage of correct answers raises to over 80% if different weights are given to different tasks, in order to simulate a normal usage of the MultiMatch system (e.g. considering users will perform a web search more frequent than audio and video searches).







Most participants found the system not very difficult to use, despite the problems that occurred during the trials. 44.5% (1-2) of the users found the system easy to use and another 26.9% (3-5) didn't find the system either simple or difficult.

# 4.2. MultiMatch Audio Task statistics



61,1% of the participants were able to complete the tasks correctly. This percentage is slightly lower than the mean percentage (72.4%). Users reported problems with the retrieval of the search results and couldn't always find the right audiofile. Also, some participants weren't able to download the player. These problems are explanation for the high percentage of participants that gave a wrong answer (22,2%) or couldn't answer the questions at all (16.7%).







The technical problems with the audio browser that were encountered during the tests made it relatively difficult for participants to complete the tasks. However, only 18.3 % of the users (categories 6 and 7) rated the system complexity as difficult or very difficult.



These results show that participants appreciate the functionalities of the audio tool. 60% (categories 1 and 2) of the users rated the tool as useful, 28.9% (3-5) had a more neutral attitude towards it and only 11.1% (6-7) of the users didn't consider the tool very useful. The interviews during the trials show that users especially liked the audio transcripts and the tag clouds.







The graph shows that almost all users would like new functionalities, although 29.7% (3-4) of them are more neutral about this. This is not necessarily negative. Users indicated during the interviews they really liked the system, but they would like to see facilities to improve and alter the tag clouds and transcripts in order to make the tool even better for other users.

# 4.3. MultiMatch Image Task



Users encountered very few problems with the execution of the image tasks. 94.4% of the users eres able to answer the questions correctly. Users made it clear during the interviews that they could perform the tasks very easily.







98.7% (1-2) of the users rated the system easy to use. Results show the image tool is very accessible and user-friendly.



The quality of the retrieved results according to users is high. 73.3% (1-2) of the users were very satisfied with the quality of the results and 26.8% (3-4) of the users had a more neutral attitude towards it. Not one of the participants rated the quality of the results as low.







The ease of the use of the system and the quality of the results was very much appreciated by the participants during the trials. 86% (1-2) of the users were very positive about the system. The rest of them rated the quality of the system as neutral (3-4).



Despite the positive experiences users had with the image tool, 23,9% of them rated the usefulness of the functionalities as low. There seems to be a contradiction with the previous results. Almost all participants rated the quality and usability of the system as high. Other users (26.7%, categories 3-5) are neutral about the usefulness. Almost half of the participants (49.6%, categories 1 and 2) consider the functionalities useful.





# 4.4. MultiMatch Multilingual Task statistics



During the interviews participants made clear they encountered problems with the multilingual tools during the execution of the tasks. Translations weren't always correct and the different languages used for the queries made it sometimes difficult for participants. However, 88.9% of the users were able to answer the questions concerning the multilingual tool correct. Only 1.4% didn't manage to provide an answer.

# 4.5. MultiMatch Video Task statistics



During the interviews users indicated they had a lot of technical problems during the video tasks. This is also apparent in the objective assessment of the responses. Only 25.9% of the respondents were able





to provide correct answers. 42.6% provided wrong answers and 31.5% of the participants weren't able to answer the questions. During interviews, users reported a few problems. The translation of the transcript and query terms didn't work properly, so users had difficulty finding the answer. Also the player for the video couldn't always be installed.



It is not very surprising that 75,5% of the users rated the video components as very difficult to use. This percentage is almost similar to the percentage of the users who provided a wrong answer or didn't provide an answer at all. This is mainly because of the problems with the translation of the transcript and query terms.







The problems encountered during the test should also influence the appreciation of the usefulness of the system. With the previous results in mind, the appreciation is expected to be very low. However, only 25.7% (6-7) of the users rated the usefulness as low. Most users (68.3%, categories 3-5) have a neutral attitude towards the video tool.



64.6% (3-5) of the participants are neutral about adding new functionalities to the video tool and 35,5% would like to see more functionalities. Together with the interviews, these results indicate that the tool itself already meets the needs of users, but also that some functionalities need improvement.

# 4.6. MultiMatch Web Task statistics



Users found the web tasks relatively easy to perform. 91.9% of the users were able to provide correct answers to the questions.







The web tool is very user-friendly according to the participants. 61.3% (1-2) of the users found the system simple to use. 18% (3-5) of the users didn't find the system particularly simple or difficult to use and 20.5% (6-7) of the participants rated the system as difficult.



# 4.7. MultiMatch Post Field Trial Questionnaire

Most users (67.7%, categories3-5) have a neutral attitude towards the system. Potentially powerful functionalities include cross-language functionalities, the combination of various forms of content and the image tool. Weak points were the performance of the video and audio tool in their current state. Tasks were often difficult to complete and there were problems with the translation tool and required players. Users did see a lot of potential in these tools and if technical problems were solved, the audio and video tool would be very easy to use.







Users were very satisfied with the multilingual functionalities. In some cases users weren't satisfied because of the amount of time it took before results were translated and because sometimes translation failed. Apparently they like the multilingual functionalities, but the performance needs to be improved.



Users were very satisfied with the specific multimedia functionalities, especially with the image and web search. They were less satisfied with the video and audio tools. This corresponds with the results





during the trials.



Many participants would use MultiMatch in their work. 34.6% (6-7) of the participants say it's very likely they would use MultiMatch and 53.3% (3-5) say they might use the system. A very small number of participants wouldn't use the system (12.2%, categories 1-2), consider the tool less valuable (15.5%, categories 1-2) than the current tools they are using or consider the tool useful (16,3%, categories 1-2).



Participants from the cultural heritage user group find the MultiMatch system very helpful. 50.1% (6-7) of the users would use MultiMatch in their work and only a very small group of 7.1% (1-2) wouldn't use it. MultiMatch is also considered as a valuable tool compared with other existing tools.





None of the users claim it's less valuable. The level of detail of the metadata isn't better or worse than other tools according to the users. Most users say in their explanation they really like the multimedia search and the focus on cultural heritage sources.



The participants of the tourism user group are less positive about the comparison of MultiMatch with other tools than the cultural heritage group. 25% (6-7) would use the system in their work and 60% (3-5) might use it. 62% (3-5) do not think MultiMatch is better or worse than the current tools they are using and 32% (1-2) say the system is less valuable. However 43% (6-7) of the participants state that MultiMatch would be useful to organize their trips. In their explanation most users say they like the combination of different forms of content.







28.6% (6-7) of the participants say it's likely they would use MultiMatch in their work and 57.1% might use MultiMatch. A small number of 14.3% (1-2) wouldn't use it. Results show MultiMatch is well suited as an educational tool: 16.7% (6-7) say it's likely they would use it to collect information for their courses and 66.6% (3-5) might use it. In comparison with other tools, the system is also valuable: only 14.3% (1-2) of the participants claim MultiMatch is less valuable than the tools they are using. Strong points according to this group are the multilingual search and the different types of content.





# 5. MultiMatch availability on the web for general users

The field trials were conducted in a controlled environment and with pre-defined user groups. The results from these trials provide insight into the usability of the system and how users compare it to other services they use, but do not give insight in the accessibility of the system for general users. To give insight into the potential impact of a more generic user group, the components of the second prototype have also been evaluated in an uncontrolled environment.

Users of WIND Libero portal Search section have been invited to use the MultiMatch search engine by exposing a "recommendation" to MultiMatch just on top of the natural result-set (by the term "natural result set" we indicate results obtained from the general Web index, to distinguish them from the ones obtained by Advertising-related links, a.k.a. "Sponsored links"). In the picture below you can see an example of such an exposition:

🗿 Libero Ricerca: leonardo da vinci - Microsoft Internet Explorer	
File Modifica Visualizza Preferiti Strumenti ?	A
😮 Indietro 🔹 🕑 🔹 😰 🏠 🔎 Cerca 👷 Preferiti 🤣 🎯 🗣 🃓 🖬 ど 📑 😭 🖏 📽 🖏	
Indrizzo 🕘 http://arianna.libero.it/search/abin/integrata.cgi?query=leonardo+da+vinci&regione=16&tr=mondo	🔽 🄁 Vai 🛛 Collegamenti 🎽
Coogle 🕞 - Cerca 4 🧔 🛩 🦉 - 😭 Segnalibri - 👰 86 bloccati 🛛 Controllo - 🍙 Invia a - 🖉	🔘 Impostazioni+ <b>Tenovo</b>
Netbook Samsung NC10.	IUNG
LIBERO Community Blog Video News Mail Search	ADSL & Internet
vveb immagini video Pagine Gialle Annunci Aitre kicerche ♥ Trova: leonardo da vinci O nel Web ● Pagine in italiano   <u>Ricerca avan:</u>	TROVA CONTRACTOR
Web Risultato ricerca: 1 - 10 su 860.000 per leonardo da vinci (0.08 s	BC) Filtro Famiglia OFF
Leonardo Da Vinci La Consegna è Gratis e ti Regaliarno un DVD, un Libro od un CD a 0,01 €l www.bol.it Leonardo Da Vinci Confronta prezzi, cerca hotel, Offerte viaggi su TripAdvisorl www.TripAdvisor.it	CCHE CORRELATE
Multimatch         sereor           Image: Sereor         Leonardo da vinci nel motore di ricerca europeo per i beni culturali         botel le firenze	aorto leonardo da vinci eonardo da vinci 2
Leonardo da Vinci - Vikipedia Leonardo da Vinci (Vinci, 15 aprile 1452 – Amboise, 2 maggio 1519) è stato un artista e scienziato italiano. Uorno d'ingegno e talento universale del http://it.wikipedia.org/wiki/Leonardo_da_Vinci - <u>Copia cache - Parine simil</u> Leonardo da Vinci - museoscienza Pronone la comologia della vista dell'artista le invenzioni i manoscritti le	eonardo da vinci eonardo da vinci milano científico leonardo da
macchine, le opere e alcune risorse. scuola cache - Pagine simili museo http://www.museoscienza.org/leonardo/ - <u>Copia cache - Pagine simili</u> museo	<u>leonardo da vinci</u>
Il Museo Nazionale della Scienza e della Tecnologia <b>'Leonardo da Vinci'</b> di <b>IMMA</b> Milano è il più grande museo tecnico-scientifico in Italia. http://www.museoscienza.org/english/leonardo/ - <u>Copia cache</u> - <u>Pagine simii</u> ANNU	GINI → NCI
Museo di Vinci La città di Vinci Vinci - History, The Leonardo Museum, Biographical Notes - The Leonardo Library - The Castle, Visitor N. Italiano - Visit the Museum http://www.leonet.ti/comuni/wici/ - <u>Copia cache</u> - <u>Pagine simi</u>	> <del>*</del> 5 *
<u>life long learning programme</u> Per avere informazioni sul Programma di Apprendimento Permanente (2007 - 2013) Programma settoriale <b>Leonardo da Vinci</b> http://www.programmaleonardo.net/- <u>Copia cache</u> - <u>Pagine simili</u>	atto su Libero Search
	Internet

The recommendation was exposed when queries met the following demands:

1. The end-user query has to be recognized as "Cultural Heritage related". This is accomplished by a stage built on WIND Natural Language Processing software called Cognito<sup>1</sup>;

<sup>&</sup>lt;sup>1</sup> <u>www.expertsystem.net</u>





- 2. The end-user query has not to return an empty result-set when issued inside MultiMatch. This is accomplished by searching in an index built with data extracted from MultiMatch repository;. Even when the end-user query is recognized as CH-related, there could be the possibility that that query, when issued in MultiMatch, would not return results. Thus the end-user would see the recommendation, but after clicking on it she/he could see a message stating "there are no results for your query". So the end-user would say "Why are you recommending MultiMatch if there are no results?". To avoid this, we tried to set-up a quick way to predict if the query, even CH-related, would return results when issued in MM or not: in the latter case we do not expose the recommendation. This has been implemented by building an index for our search engine using MM data.
- 3. A frequency-limit value has to be passed. This has been made to avoid too many concurrent users visiting MultiMatch, and is accomplished simply by exploiting a random-number generator; This does not influence data on index coverage, since the steps above are always executed sequentially, thus even if we do not pass step 3, we retrieved and logged the results of steps 1 and 2.

Each filter is implemented via a software stage which is executed sequentially: if a check fails, the following ones are skipped, and the recommendation is not shown. If an end-user clicks on the recommendation, the original query is issued towards MultiMatch, having programmatically set both the language in which the query is written and the language of the results to Italian (since most of Libero portal end-users are from Italy). The query is issued toward the MultiMatch Overview interface. Notice that the recommendation can be exposed only on the first page of a query result-set, i.e. together with results whose ordinal value ranges from 1 to (no more than) 10.

# **5.1. Indicators**

By analysing data from log-files collected both on the WIND and the Multimatch sites, the following indicators can be measured:

- MultiMatch index coverage with respect to end-user queries. Even if this measurement is
  mediated by the fact that the query is issued toward a copy of the MultiMatch index and
  executed by the WIND proprietary text-retrieval engine (which has been set-up to implement
  the filter 1 above), we believe that it is a good indicator of the effectiveness of the MM index
  itself with respect to general-web users activity. Notice that this can be measured both respect
  to CH-related queries only or respect to the overall queries.
- The capture-ratio, i.e. the number of times the end-user accepted to enter MultiMatch with respect to the number of times the recommendation was exposed;
- The mean number of queries per entered user;
- The distribution of the returned number of results and the response-time of MultiMatch as experienced by end-users.

Data presented in the following sections are related to two timeframes in which the recommendation was enabled to be exposed; they are:

from 2008-11-27-17:16 to about 2008-11-28-19:00

from 2008-12-02-16:48 to about 2008-12-08-23:59

The first run was interrupted due to problems on WIND site, not by problems on the MultiMatch installation.

# 5.2. MultiMatch Index coverage

The table below reports the number of events related to the different possible combinations of the output of the filters pipeline described above:





Query recognized as CH-related	Match in WIND MultiMatch index	MultiMatch recommendation shown	Number of events
Y	Y	Υ	15090
Y	Y	N (due to frequency limit)	135510
Ν	Ν	Ν	6338465
Y	Ν	Ν	415325
Ν	Y	Ν	1458707

Thus the MM index coverage is about 19.2% (with respect to the overall queries) which rises to about 26.6% if we take into account only the queries recognized as CH-related by the WIND query classificator. This is considered to be quite a good result, since at this level of MultiMatch implementation; it was not supposed to index all content pertaining to CH area, especially for a specific language (Italian in this case).

# 5.3. Capture ratio

As can be seen from the table above, the recommendation has been exposed about 15000 times. By analysing the Tomcat access logs, we can count the number of accesses from WIND, since these operations are marked with a special parameter in the request URL: this number is 592. The capture ratio is thus about 3.9%. This can be considered a good result, since, as an example, the CTR (click-through) on the widget that can be seen on the right of the result-page (see picture in paragraph 1 of this chapter) is about 2.9%.

### 5.4. Mean number of queries per landed-user

This can be calculated by examining the MM application debug-log (namely the catalina.out file in the Tomcat logs directory): the number of logged queries is about 657. It turns out that the mean number of queries per landed user is about 1.1, which can be summarised by saying that, in mean, 10% of the users tried another query in the MM engine.

### 5.5. Results-number and response-time distributions

For the logged queries it has been possible to plot the total number of results and the overall responsetime. These two distributions can be seen in the following pictures:





#### Results-number distribution











The values of the two indicators can be summarized using the distributions mean and variance. They are:

Indicator	Mean	Standard deviation
Total number of results	26.4	13.1
Response-time (sec)	5.37	4.24

The mean value of 5.4 seconds for the response-time can be considered a good value for a prototype, since a typical response time for a multi-source result-set commercial search-engine has been measured to be about 1.5 seconds.

# 5.6. A snapshot of end-user queries

This section provides an insight for specific queries issued by end-users which visited MultiMatch: these are kept from the set of the 657 queries In the following table the top-most queries are listed, together with their score and a (hand-made) classification in term of the top entities classes MultiMatch search for:

Top Queries			
Score	Query	Class	
20	Picasso	Creators	
11	www.istruzione.it/studenti	Organization	
9	van gogh	Creators	
7	testi canzoni	Creations	
7	leonardo	Creators	
6	traduttore		
6	Degas	Creators	
5	Musica	Creations	
5	museo madrid	Organization	
5	Clown		
4	www.istruzione.it	Organization	
4	museo Roma exhibition	Events	
4	dal futurismo al fascismo		
3	www.unifi.it	Organization	
3	traduzione dal latino all'italiano		
3	traduttore lingue		
3	traduttore inglese		
3	santa lucia immagini	Places	
3	Palla		
3	opere di Paul klee	Creations	
3	musica da scaricare	Creations	





3	monet le ninfee	Creations
3	mercatini di natale	Events
3	marco dal re	Creators
3	canzone nuove 2008	Creations
3	Liturgia del giorno	
3	Come svolgere un buon tema?	

In the following table, instead, a sample of queries grouped by the classes are shown:

Queries samples grouped by class		
Class	Queries	
Creators	picasso	
	van gogh	
	leonardo	
	degas	
	marco dal re	
	shakespeare	
	The Cure	
	Salvatore De Francesco artista	
	pittori in piemonte	
Creations	testi canzoni	
	musica	
	opere di Paul klee	
	musica da scaricare	
	canzone nuove 2008	
	il ballo nel 1930	
	arte figurativa	
	tecniche pitture murali	
	quadri di napoli	
	primavera del botticelli	
	poesia san francesco	
	amore e psiche canova	
Places	santa lucia immagini	
	duomo monreale	
	vicenza palladio	
	antica scozia	
	FIRENZE	
	chiese gotiche	
Events	museo Roma exhbition	
	mercatini di natale	
	venezia mostre	
	museo esposizione	
	prima guerra mondiale	
	la storia di Savoia	
	IL NEOREALISMO	
Organizations	www.istruzione.it	
	museo madrid	
	biblioteca comunale como	
	MUSEI DI GENOVA NERVI	
	terzo reich	





	www.unifi.it
Other	traduttore
	contributi di fondi europei per la cultura
	tutte le donne di picasso
	ARALDICA
	clown





# 6. Conclusions

# 6.1. Cultural Heritage user group

The Cultural Heritage user group was the most substantial group performing these Field Trials. The following items synthesize the results of the Field trials conducted by the Cultural Heritage users and their comments, feedback and suggestions for the MultiMatch engine.

### Results from the field trials in the Cultural Heritage user group

Performing the Field trial tasks the Cultural Heritage users were generally satisfied with the system: they needed some time to figure out some buttons and related functionalities, but while complex, they found the system learnable, because once they figured it out they could continue and move along more quickly.

Users were generally satisfied with the results of the system: they found them relevant and from reliable sources, but it was not always easy for them to find the information requested. In general, they found the Metadata provided by the system detailed enough to satisfy their Cultural Heritage related needs.

Concerning specific functionalities, this user group found both the multimedia itself and the multimedia functionalities very useful; even if video and audio were the most difficult to perform, users found them particularly relevant to the Cultural Heritage field.

Due to the slowness of the system, in some cases it took a long time for users to perform and complete all the tasks. This problem was emphasized in the remote mode field trials where, in some cases, it caused the suspension of tests.

### Positive and strong points of MultiMatch for the Cultural Heritage user group

Generally Cultural Heritage users liked the concept of MultiMatch and found it very powerful. The possibilities presented by the system could provide advantages in the CH field. They also liked the idea of searching within different language collections and through different sources of content. Multimedia (image, speech, audio, video) and multilingual functionalities (mono and cross-language) would enable them to easily perform their daily work in the Cultural Heritage field, helping them to retrieve relevant information.

### Added value and possibilities of MultiMatch for the Cultural Heritage user group

The MultiMatch engine is particularly designed for the Cultural Heritage domain and it brings high added value to this user group. Statistics on the post field trial questionnaire show that a dominant part of the users conducting the field trials would use MultiMatch in their Cultural Heritage related daily work. The system would enable them to accomplish tasks more quickly and, compared with tools they use now, MultiMatch could provide advantages in terms of relevance of the results collected.

### Further research/improvements

For Cultural Heritage users, improvements in the system should concern the performance of system, in terms of the time to produce results, functionalities performance and relevance of the results with respect to the query. Specific functionalities could be improved, in particular the video functionalities (videos and videos transcripts should be in all languages) and the audio functionalities (audio transcripts should be in all languages, and tags should be checked). MultiMatch should be made compatible with other internet browsers like Safari or Mozilla.





# 6.2. Tourism user group

The Tourism user group provides good insight in the value of MultiMatch for this type of user. The following items synthesize the results and the comments, feedback and suggestions for the MultiMatch engine.

### Results from the field trials in the Tourism user group

Users did like the system, but also encountered difficulties during the execution of the tasks. Web and image tasks were relatively easy, but the video and audio tasks were harder to complete, due to technical problems. Users needed some time to figure out the multilingual tools and translations weren't always correct. The use of different languages during the tasks made it sometimes hard for users to find the right answers, but the translation tool made it sometimes easier.

The speed of the system was a problem for users. Most participants of the Tourism user group had little experience with the testing of demo's, so expectations were maybe a bit too high. Sometimes the system collapsed and it took a long time for some users to complete the tasks.

Users found the retrieved results very reliable and they liked the multimedia search. Especially the image and web functions were found very useful. Despite the technical problems, users liked the video and audio functionalities, but they considered them less relevant for the retrieval of information.

### Positive and strong points of MultiMatch for the Tourism user group

Test results show that a lot of the participants from the Tourism user group could use MultiMatch to organize trips. Most users equally value the MultiMatch engine with respect to other tools they are currently using. Although participants had some problems with the use of the language tools, they did see a lot of potential for them.

### Added value and possibilities of MultiMatch for the Tourism user group

Although not particularly designed for the Tourism user group, MultiMatch can have an added value for this user group. The tools will help them to search through content in different languages more easily. The Tourism user group would especially benefit from the image and web search. The video and audio search could also be helpful.

### Further research/improvements

Improvements should concern the multilingual functions and the quality of the translations. The speed of the system has to be improved. At this point, retrieval of the results takes too long for users. The performance of the video and audio tool could be improved by providing correct translations of the queries within the transcript.

### 6.3. Education

MultiMatch can potentially play an important role in education. The sources indexed by the system can for example be used in the humanities for studies. Having these sources available online and accessible in multiple languages would greatly influence the way cultural heritage can be enjoyed online.

### Results from the field trials in the educational user group

Users were very impressed by the system, notably the sources that have been indexed. Users in the educational domain see a lot of potential to search through these resources. Some of the interviewees knew about the Europeana and The European Library (TEL) projects and mentioned how the features of MultiMatch, notably the multilingual search, should also be made available on the Europeana and TEL websites, as they obviously contain more content. Users like how you can propose new terms to be translated and appreciated the use of tag-clouds.





### Positive and strong points of MultiMatch for the Educationl user group

The system provides access to added value resources that cannot be found with general search engines, especially in the cultural heritage domain. The system provides useful translations of textual content. Users in this user group found the interface intuitive.

<u>Added value and possibilities of MultiMatch for the Educational user group.</u> If a system like MultiMatch would contain more collections, it would certainly be used in education next to textbooks, notably in the humanities. Students and teachers in higher education would use many of the advanced search options and also look at the metadata for its provenance. Teachers said they appreciated MultiMatch more than a general search engine or Wikipedia, because it is more focussed and also offers content from trusted sources.

Further research/improvements

- The search engine did not work with some web browsers (reported with Safari).
- Metadata access and exploitation can be improved..
- Boolean search should be implemented.
- Too many options and languages involved: user language, search language, interface language, translation language, filter language. The interface can be improved so that makes this more intuitive.
- The content is now rather limited and, when expanded, could be noisy.

The interface is easy to use, but it is difficult to understand language options (reported when using Spanish and English). Speed of the system needs to be improved. Usability: results should appear higher in the page and there is too much information in one page. Term clouds do not show relevant relations.

User's language should be automatically identified and the query automatically translated into the other languages. Then, the results should be segmented. Now, searches involve source, target and filtered options and the query language must match the resource language, e.g., tour Eiffel. The term "collection" in the context of image retrieval is ambiguous. Image and text retrieval were easier to perform than video and audio tasks.

### **6.4 Final remarks**

It must not be forgotten that in these field trials we were testing a system prototype of a complex system that was still being subjected to refinements and debugging both with respect to single components (e.g. the video component) and the overall integration.

The testing environment was research rather than industry oriented and this of course impacted on performance. From this perspective the feedback from the testers can be divided into two sets:

- comments regarding performance problems
- comments regarding the new functionality offered and the potential of the system.

In our opinion, the second set of comments bear far more weight, especially in the light of future work and experimentation.

Overall, the innovative features of the system (i.e. the ones that make the MultiMatch system truly stand out) are highly appreciated, as evidenced in Paragraph 4.6 for example.

Most users felt that MultiMatch compares very well with other tools they are currently using.. Overall, users found the interface intuitive. Although participants had some problems with the use of the language tools, they did see a lot of potential for them. Improving the way multilingual search is supported by the system will be a priority in future research.





Users also liked the idea of searching within different collections and through different types of content. Multimedia (image, speech, audio, video) and multilingual functionalities would enable them to easily perform their daily work, helping them to retrieve relevant information. If a system like MultiMatch contained more collections, it would certainly be used in education next to textbooks in education, notably in the humanities. For these and other reasons, it would be advisable to include the technology developed by MultiMatch in the Europeana portal. This was suggested by some partners involved in the projects and networks that develop the technology for Europeana.

The ability to perform content-based image and video search would be highly appreciated by users. The stability of the video component should be improved.

Building the ideal search engine will encompass many academic disciplines, including many information channels created by cataloguers, algorithms, systems and users themselves. MultiMatch made a special effort to have users in the loop continuously, and by doing this work towards search systems that return relevant results, for any type of query imaginable...

A known needle in a known haystack A known needle in an unknown haystack An unknown needle in an unknown haystack Any needle in a haystack The sharpest needle in a haystack All the needles in a haystack Things like needles in any haystack Affirmation there are no needles in the haystack Most of the sharpest needles in a haystack Let me know whenever a new needle turns up Where are the haystacks? Needles, haystacks, whatever





### References

D7.1 Evaluation methodology
D7.2 Evaluation of first prototype
D7.3 Evaluation of second prototype
Hansen P. (1998). Evaluation of IR User Interface – Implications for User Interface Design.
Hufnagel, E. M. (1990). User satisfaction - are we really measuring system effectiveness? Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences, 4, 437-446.
Nielsen J., 1993. Usability engineering. Boston: Academic Press
Norman D. A., 2002b. Emotion and design: attractive things work better.
Interactions 9(4), 36-42
www.SurveyMonkey.com
Zhang, Ping and Dennis Galletta. Human-computer Interaction and Management Information
Systems: Foundations Published by M.E. Sharpe, 2006





# Annex 1. Presentation slides field trials





	<ul> <li>Illustrate the home page of MM PT2</li> <li>Start with the overview mode</li> <li>Show access to other specialized search modes</li> <li>Show search for all languages</li> <li>Show that user can register or login in order to use personalized functionalities</li> <li>Show that it is possible to select the interface language</li> </ul>	





Simple	e query
	<ul> <li>Try a simple query, such as flowers</li> </ul>















Image specialized view	
Image: constraint of the second of the se	<ul> <li>In this view, the user has two new functionalities</li> <li>He can search images similar to one of those already retrieved</li> <li>Or he can search other images similar to the one selected and that specify his query</li> </ul>
Video speci	alized view 💙 🥙
	<ul> <li>In this view, the can</li> <li>Search the transcripts of all video files retrieved</li> <li>Open the video payer and view the video</li> </ul>
	Land Schry
Video specia	
	<ul> <li>If he searches the transcripts of all video files retrieved</li> <li>He can search for a term and display the number of occurrences of that term in all videos</li> <li>He can also display the transcript of a specific video</li> </ul>
	Internation Society





Video specialized view	
	<ul> <li>Only registered users can play the video.</li> <li>Users can have an overview of video content, by looking at key frames, associated to each video shot and to the corresponding audio transcript</li> <li>Looking at the transcripts and key frames, the user can play the video from a specific point</li> <li>The user can search all video shots containing a certain term. As a result of the query, the user gets the shot numbers</li> </ul>
	Line Science





# Annex 2. Set of guided tasks

### **TESTS:** These were prepared in Duch, Italian and Spanish)

### **General Overview**

You are using the Multimatch search tool. The URL is: http://homer.multimatch.hostedbyfdi.net:8090/overview/

• MultiMatch offers the option of choosing the language interface in six languages. Choose now Spanish. Click on "interface language" at the right upper part of the screen.

• If you wish, you can register at start of session in case you wish to personalise functionalities. However, this is not necessary.

• MM also offers advanced search, like audio, archives, images, news, webs, authors, and video.

• To the right of the search box, you will find the option search all languages, translate the query and advanced search.

Let's try to make a simple query:

• Type in the search box "teatro", click on search box.

• Now Multimatch translates "teatro" into the other languages and performs different searches of these expressions and words among indexed contents.

### SEARCH IN ONE LANGUAGE OR MORE

• Click in "Search all languages". You will see on the screen two options among the six languages used in MM. The first language option is the source language and the other one is the target language.

• Keep the query term "teatro". Select español in "escribo en:". For the second option language, select inglés, and click the option "Traducir consulta (Translate my query)"

• Next you will get the search results sorted in different modalities.

• Following the steps given, search the query "literatura infantil". Select Spanish language as the target language.

• You will be able to obtain results sorted in web and image.

### **ADVANCED SEARCH**

Advanced search has four different fields. These are: "All the fields", "Keyword", "Title" and "Location".

• Type in "Cervantes" in all the fields and in title "El Quijote" You can observe that the results returned let you search within results: cloud terms, metadata, close terms or piece fragment.

• Following this example, type in "El Quijote" only in the field "Title". The results returned show up in different documents, formats...

• Now, following the advanced search, type in "keyword" "Quevedo" and in title "La dama boba".

Follow with this example and select Spanish as source language and Italian as target language

Image task

1. You are looking for an image of a church tower

Please find an appropriate image in the Spanish collection.





What is the title of the image (in Spanish?)

2. How easy was it to:

### Locate the image

1 very easy-7 very hard

3. How helpful were the following features?

### "Translate my query" link next to the main search box Tabs for different language collection results Automatic query translation type Interactive query translation type

4. Now imagine you would like to find an image of a landscape.

You want to see what is available in the different collections. Please search for "landscape" and then inspect the results in the following sections:

English results (93~images) Italian results (19~images) German results (61~images)

How relevant are the images displayed to the query for each section? **1. English2. Italian 3. German** 

5. How could the cross-language search features be improved?

Web task

You are searching information about W.A. Mozart. Please provide the following biographical information (Using MultiMatch and its related pages):

- 1. What was his sister's name?
- 2. How many pieces did he compose?
- 3. What were the dates of his life and death?
- 4. What is the URL of the page where you found this information?

### 2. How easy was it to: Find the sisters name Find the number of compositions Find the birth and death dates

1 very easy- 7 very hard

3. Which language collection(s) did you use for this task? English Spanish





Italian Dutch Polish

4. Did Mozart meet any of the following people? YES/NO
J.S. Bach
J. Haydn
A. Vivaldi
G.F. Händel
A. Salieri

5. How easy was it to:Answer this question1 very easy - 7 not at all easy

Audio task

Use the audio browser to find the program Episode for Families. Use the transcript browse to find the episode with a story about a fisherman.

1. What is the name of this episode?

Use the transcript to find the segment where the Metropolitan Museum is mentioned.

2. Where does that segment start?

Click on the bar to listen to the fragment.

- 3. What can you visit in the museum?
- 4. How was it to use the transcript browse tool?

1 very easy- 7 very hard

Close this audio file and use the browse tool to find the story about frogs. Click on the tag. The clip should start playing from that point.

5. To who did the frogs asked for a king?

Use the tags to find the part about a stork and a fox and listen to it.

7. Was it easy to find this part?

very easy -7 very hard

8. How useful were the following features: (1 very – 7 not at all)Browse transcripts





Browse by tags

9. How helpful would the following features be?(1 very- 7 not at all)Being able to browse all transcripts at once.Being able to manually fix incorrect parts of the audio transcript for future reference.Being able to manually fix incorrect tagsBeing able to add tags.

Video task

Use the following URL:

http://homer.multimatch.hostedbyfdi.net:8090/video/

To watch videos: Username: multimatch2 Password: f64ty1

You are searching for a program about World War I by Midas Dekkers. Use the button "search transcripts of all video files shown on this page" to search within the results.

Click on open video player

1. How many Dutch citizens died during the mobilization? Use the "search for keyword" function.

Use the keyframe bar to search the part in the video about archives.

2. What's the first keyframe about archives?

3. What's the last keyframe about archives?

Click on the first keyframe about archives. The video will start playing from that point.

4. What was the name of the archive system?

5. This episode is partly based on a book. Use the different search methods to find the:

6. How easy was it to:

(1 very easy- 7 very hard)

Find the video Find the place in the video about the archives Answer the questions





7. How useful were the following features:
(1 very – 7 not at all)
Search transcripts of all video files shown on this page
Search for key words within the videoplayer
Keyframes
Transcript of the keyframes

8. How helpful would the following features be?(1 very- 7 not at all)Having tag clouds (key terms) instead of pure audio transcripts to represent the spoken content.Being able to manually fix incorrect parts of the audio transcript for future reference.Being able to add tags to the keyframes.

Post Field trial Questionnaire

This questionnaire is intended to test the user satisfaction on the system after the field trial activity.
General satisfaction with system and results

Rating:1=Not at all, 7=Very

Question	
How easy was the system to use?	
How satisfied are you with the results?	
How relevant were the results?	
Did you feel that the results came from	
reliable sources?	

Open Question	
What did you like about the system?	
What difficulties did you encounter?	

Satisfaction with specific system functionalities Rating: 1=Not at all, 7=Very

### 2.1 MULTILINGUAL FUNCTIONALITIES How useful did you find the following functionalities:

Functionality	
Select the interface language	
Search in all languages	
Search in a specific language	
Translate the query	
Advanced search	
View results in different formats	
Search within returned results in term clouds,	
metadata, related terms, snippets	

# 2.2 MULTIMEDIA FUNCTIONALITIES

### How did you find the following functionalities:





# 3 Overall impressions

Rate the following statements Rating: 1=Not at all, 7=Very

Statement	
I would find the system useful in my job	
Using the system would enable me to	
accomplish tasks more quickly	
It was easy to learn to use the system	
It was easy to find the information I needed	
This system has all the functionalities and	
capabilities I expect it to have	
Overall, I am satisfied with the ease of	
completing tasks	
Overall, I am satisfied with the amount of	
time it took to complete the tasks	
Overall, I am satisfied with the level of	
support given by the system when completing	
the tasks	

## 4 Satisfaction of the target user group

Question	
To which user group do you belong?	□ cultural heritage
	🗆 tourism
	□ educational

Rating: 1=Not at all, 7=Very

### 4.1 Cultural Heritage

Question	
Would you use MultiMatch in your CH-	
related daily work?	
Compared with the tools you use now, do you	
think that MultiMatch could provide added	
value?	
Do you find the metadata provided by the	
System detailed enough to satisfy your CH	





needs?	
Open question	
From your perspective, what are the strong points of MultiMatch?	

### 4.2 Tourism

Question	
Would you use MultiMatch to satisfy your	
Cultural tourism needs?	
Compared with the tools you use now, do you	
think that MultiMatch could provide added	
value?	
Would you find MultiMatch useful to	
organize your cultural trips.	
Open question	
From your perspective, what are the strong	
points of MultiMatch?	





### 4.3 Educational

Question	
Would you use MultiMatch in your	
educational-related daily work?	
Compared with the tools you use now, do you	
think that MultiMatch could provide added	
value?	
Would you find MultiMatch useful to collect	
information for your courses/research	
activities	
Open question	
From your perspective, what are the strong	
points of MultiMatch?	