

Project Activities

Identify relevant material via in-depth crawling of selected CH institutions

Crawl the Internet to identify websites with CH information

Automatically classify the results on the basis of occurrence of CH concepts

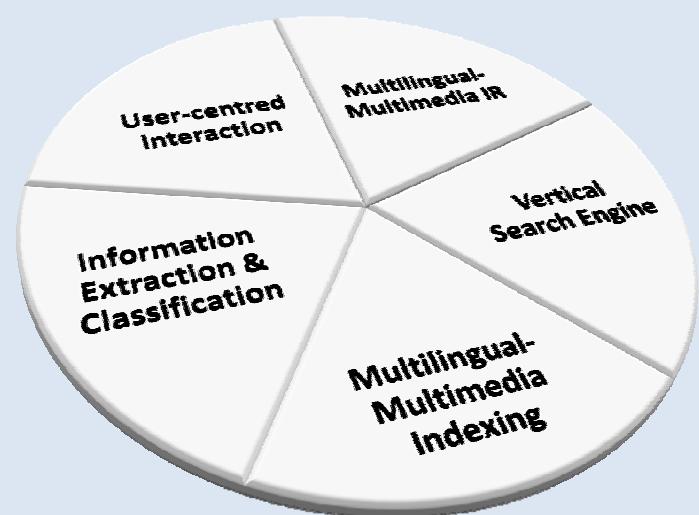
Automatically extract relevant information which is used to create cross-links between related material

Organise and further analyse the material crawled to serve focused queries generated from the users

Interact with the user to obtain a more specific definition of initial information requirements

Search results are organized and integrated allowing the users to access information regardless of language and media barriers

Research Areas Involved



Understanding User Requirements

- CH Professionals do use the Internet as part of their daily work routine but they currently depend largely on generic search engines.
- They want to query using natural language and familiar Boolean operators.
- They would like full capabilities for multimedia retrieval (i.e. Images and video as well as text) but, in most cases, are only accustomed to executing text searches.
- Their main focus appears to be on **works of art** and their **creators**, with all associated information, such as critical reviews, information on related events, etc.
- They would find information filtering, clustering and aggregation functionalities very useful.
- They demand high precision of results and need to know the source and level of authority.
- They need to be able to save both queries and results for future processing and reuse.
- They tend to restrict their searches to their own language plus English, thus missing information only available in other languages.
- If multilingual search was available, they would like to have the results associated with descriptive snippets in their own language or English at least.

System Retrieval Functionality

The screenshot shows the MultiMatch interface with the search query "van gogh painting". The results are displayed in a grid format, showing various links and images related to Vincent van Gogh's paintings. The interface includes language selection (English to Spanish), workspace management, and detailed search results.

Generic Search: English to Spanish

The system retrieves Spanish documents in all media relevant to the query "van Gogh" AND "painting".

The screenshot shows the MultiMatch interface with the search query "still life flowers". The results are displayed in a grid format, showing various links and images related to still life compositions. A detailed metadata panel is open for the first image, showing fields like Creator, Source, Name, Is Person, Identifier, Type, Title, Source, Related Creator, Location, Identifier, Digital Representation, Description Translated From, Description Language, Description, Archive Location, Source Identifier, Source, Identifier, Format, Copyright, and a preview of the image.

Specialized Search: English to Italian

The system retrieves all Italian images relevant to the query "still life" AND "flowers". Metadata for the first image are visualized.