



Project no. 033104

MultiMatch

Multilingual / Multimedia Access to Cultural Heritage

Technology-enhanced Learning and Access to Cultural Heritage

Instrument: Specific Targeted Research Project

FP6-2005-IST-5

Final Activity Report

Start Date of Project: 01 May 2006

Duration: 30 Months

Project Coordinator: Pasquale Savino

Organisation Name of Lead Contractor for this Report: ISTI-CNR

Version Final

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

List of Partners

Partic. role	Partic. no.	Participant name	Participant short name	Country	Date entered project	Date exit project
CO	01	Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche	ISTI-CNR	IT	month 1	month 30
CR	02	University of Sheffield	USFD	UK	month 1	month 30
CR	03	Dublin City University	DCU	IE	month 1	month 30
CR	04	University of Amsterdam	UvA	NL	month 1	month 30
CR	05	University of Geneva	UniGE	CH	month 1	month 30
CR	06	Universidad Nacional de Educación a Distancia	UNED	ES	month 1	month 30
CR	07	Alinari 24 ORE S.p.A ¹	ALINARI	IT	month 1	month 30
CR	08	Stichting Nederlands Instituut voor Beeld en Geluid (Netherlands Institute for Sound and Vision)	BandG ²	NL	month 1	month 30
CR	09	Biblioteca Virtual Miguel de Cervantes, University of Alicante	UA	ES	Entered at month 9	month 30
CR	10	OCLC Ltd ³	FDI	UK	month 1	month 30
CR	11	WIND Telecomunicazioni S.p.A.	WIND	IT	month 1	month 30

¹ Previously Fratelli Alinari Istituto Edizioni Artistiche S.p.A; Alinari was taken over by Sole 24 Ore SpA on 1 October 2007; the new name is Alinari 24 ORE S.p.A

² Although the official acronym for this institution is BandG, it is often referred to within the project as Sound & Vision.

³ Initially Fretwell-Downing Informatics (FDI); FDI was taken over by OCLC PICA at the moment of accession to the contract. The correct name of this partner is now OCLC but the acronym FDI is used in the Technical Annex (DoW).

Table of Contents

Section 1	Summary of project objectives and project results	7
1.1	The Project Structure.....	7
1.1.1	Results achieved	8
1.2	Understanding User Requirements.....	10
1.3	System Functionality.....	11
1.3.1	Search Interaction Levels in MultiMatch	13
Section 2	Relationships with the state of the art.....	16
Section 3	Impact of the project on the Industry and Research Sectors.....	19
Section 4	Dissemination of Knowledge	20
4.1	Overview Table	20
4.2	Project Presentations / Representation at Events.....	21
4.3	Participation in Focus Groups, Seminars, Research visits, Forums	26
4.4	Organisation of Workshops.....	27
4.5	Project Website	28
4.6	Posters/Brochures.....	29
4.7	Publications	29
4.8	Newsletters.....	34
4.9	Press Releases	34
4.10	Contribution to Standards.....	34
4.11	Clustering Activities.....	35
4.12	MultiMatch User Groups	36
Appendix A	List of public deliverables.....	39
References		40

Introduction

The aim of the MultiMatch project is to enable users to explore and interact with online internet-accessible CH content, across media types and language boundaries. This is achieved through the development of a search engine targeted for the access, organisation and personalized presentation of cultural heritage information.

The MultiMatch search engine was designed to (i) crawl the Internet to identify websites with CH information, locating relevant texts, images and videos, regardless of the source and target languages used to write queries and describe the results; (ii) identify relevant material via an in-depth crawling of selected CH institutions, accepting and processing any semantic web encoding of the information retrieved; (iii) automatically classify the results in a semantic-web compliant fashion, based on document content, metadata, context, and on the occurrence of relevant CH concepts in the document; (iv) automatically extract relevant information which will then be used to create cross-links between related material, such as the biography of an artist, exhibitions of his/her work, critical analyses, etc.; (v) organize and further analyse the material crawled to serve focused queries generated from user-formulated information needs; (vi) interact with the user to obtain a more specific definition of initial information requirements; and finally (vii) organize and display search results in an integrated, user-friendly manner, allowing users to access and exploit the information retrieved regardless of language barriers.

The project was completed on October 31st, 2008 and all planned objectives were achieved. In particular, a multilingual search engine, specialized for searching cultural heritage objects available on the Web and stored in digital archives, was developed and tested and three different field trials were conducted; the field trials experimented the use of the system prototype in different application settings: in the educational field, in the touristic field and for professional archives.

The MultiMatch project involved 11 partners: Istituto di Scienza e Tecnologie dell'Informazione – Consiglio Nazionale delle Ricerche (ISTI-CNR), University of Sheffield, Dublin City University, University of Amsterdam, University of Geneva, Universidad Nacional de Educacion a Distancia, Fratelli Alinari Istituto Edizioni Artistiche SpA, Netherland Institute for Sound and Vision, Biblioteca Virtual Miguel de Cervantes, OCLC PICA, WIND Telecomunicazioni SpA.

The project website is active (<http://www.multimatch.org>) and an online demo of the MultiMatch prototype can be accessed by registered users (free registration available).

In this introduction we briefly summarize the results achieved in the project, while the remaining sections are devoted to providing more details on project results (Section 1), a comparison with the state of the art in the field (Section 2), the impact on industry and research (Section 3) and the dissemination activities (Section 4). Appendix A contains a complete list of public deliverables available on the project web site.

Results achieved

- The MultiMatch Multilingual/Multimedia Search Engine for Cultural Heritage supports retrieval of cultural objects through different modalities:
 - The simplest one is a traditional *free text* search. This search mode is similar to that provided by general purpose search engines, such as Google, with the difference that MultiMatch is expected to provide more precise results – since information is acquired from selected sources containing Cultural Heritage data – and with support for multilingual searches. 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. Multilingual 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.

- Multimedia searches, based on similarity matching and on automatic information extraction techniques.
- Metadata based searches. 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.
- A browsing capability allows users to navigate the MultiMatch collection using a web directory-like structure based on the MultiMatch ontology.
- Content selected from different sources

The main source of information stored in the MultiMatch prototype system is composed of cultural heritage objects obtained through crawling and indexing of material obtained from cultural heritage sites (e.g. museums), web encyclopedias (e.g. Wikipedia), digital libraries of specific cultural heritage organizations, OAI compliant digital resources, and RSS feeds from cultural web sites.
- The project dedicated a significant research effort in the general field of multilingual/multimedia information access and retrieval, and in particular in the following areas:
 - Focussed search engine

MultiMatch aimed to take a significant leap forward from today's vertical search engines, by offering "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, presenting both certified and non-certified information.

The MultiMatch search engine: i) is the first search engine combining automatic classification and extraction techniques with semantic web compliant encoding standards; ii) considers complex user profiles and search scenarios; iii) can search across language boundaries and across different media.
 - Multilingual/multimedia indexing

Instead of returning documents in isolation, MultiMatch provides complex search results that puts documents of various media types into context. For the indexing-end of MultiMatch, complex object retrieval generates special challenges. First, documents of various media types (text, audio, image, video, or mixed-content) and accompanying metadata are indexed. The indexing strategies used also recognise and cater for multilingual content
 - Information extraction and classification

MultiMatch allows users to interpret the wealth of CH information by presenting objects not as isolated individual items, but as situated, richly connected entities. A range of classifications, as well as various links to reviews, experience reports, and general background knowledge, are provided. Documents are classified on the basis of diverse dimensions, such as topical, geographical, and temporal. MultiMatch uses large scale information extraction from documents to identify entities and their relations in large Web corpora.
 - Multilingual/multimedia information retrieval

There has been a significant increase in research on information retrieval techniques for multimedia (i.e. constituted of audio, visual and textual streams) and multilingual document collections, recently. Unfortunately, there has been little transfer of research advances to real world applications. MultiMatch aimed at bridging this gap.

Multilingual information retrieval (MLIR) focuses on the problem of using a request in one language to retrieve documents from a collection in multiple different languages. MultiMatch developed components for both document and query translation and procedures for matching one against the other. Much effort was dedicated to the building of domain-specific multilingual resources catering for the terminology adopted in the CH domain.

- User-centred interaction
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. Users look mainly for concepts (e.g. individuals, facts, places) and far less for features (e.g. mountains, sunset, clouds). A "semantic gap" exists. The MultiMatch user interface integrates automatic techniques for low level feature extraction and automatic concept classification. Structures for browsing were created, allowing users to explore content or search results following multiple facets.

Section 1 Summary of project objectives and project results

Europe's vast collections of unique and exciting cultural content are an important asset of our society. On the web, cultural heritage (CH) content is everywhere, in traditional environments such as libraries, museums, galleries and audiovisual archives, but also reviews in popular magazines and newspapers, in multiple languages and multiple media. Moreover, many consumers have now become contributors to information about CH objects through weblogs ("blogs") and discussion boards. CH objects on the web are no longer isolated objects, but situated, richly connected entities, equipped with very heterogeneous metadata, and with information from a broad spectrum of sources, some with authoritative views and some with highly personal opinions.

What means do users have to access these complex CH objects? How can they explore and interact with CH content in ways that do justice to the richness of the objects without being overwhelmed? Currently, users interested in accessing CH content - be it for educational, touristic, or economic reasons - are left to discover, interpret, and aggregate material of interest themselves. Rose and Levinson [Rose et al., 2004] show that over 60% of today's web searches are so-called informational queries, i.e. queries that essentially ask for multiple perspectives from multiples sources. Such queries are not necessarily easy to formulate.

The cultural heritage search and navigation facilities envisaged by MultiMatch cater for these information needs by presenting users with a composite picture of complex CH objects. For instance, in reply to a user's request for information on Van Gogh, the MultiMatch engine can present information on Van Gogh from multiple museums around Europe, in multiple languages; it could complement this with pointers to Van Gogh's contemporaries, with links to exhibitions on Van Gogh, to reviews of these exhibitions, to blog entries by visitors to these exhibitions, and to background information taken from online resources or dedicated sites.

The need to provide information seeking users in the CH domain with this kind of result, posed a number of scientific challenges for information retrieval research in areas as diverse as web crawling, multilingual access, multimedia access, semantic processing, and presentation design. The MultiMatch research agenda aimed at making significant developments in each of these areas in order to arrive at a proof-of-concept implementation of a user-centered search and navigation engine for cultural heritage content.

1.1 The Project Structure

The concepts underlying the system are shown in Figure 1. On the left-hand side of the figure, we show users querying the system in different languages for a range of information on the Dutch artist Vincent van Gogh, including critical analysis, biographies, details of exhibitions. The system displays the information retrieved in an integrated fashion, and in a format determined by the particular user profile. On the right-hand side, we show possible sources of this information and the ways in which it can be acquired.

The project developed a system prototype that supports six languages: Dutch, English, Italian, German, Polish, and Spanish, and extendible to others. The R&D work is organized around three streams:

- *User-oriented research* activities focused on analysing the user requirements, defining the necessary system functionality, selecting and preparing content for system development, studying the ontologies used in the CH domain and the semantic encoding to be adopted by the system.
- *System-oriented research* activities included the study and development of software components for the acquisition, indexing, classification, retrieval and presentation of multilingual CH information in diverse and mixed media and their integration in the system prototypes.
- *Validation* activities included evaluation of the system and its components. User groups composed of CH institutions and CH consumers were formed to test the system and provide feedback.

1.1.1.1 Focussed search engine

In its simplest form, a vertical search engine, i.e., one that caters for domain-specific searches, simply filters a subset of the web believed to be relevant to a topic. In a more useful form, a vertical search engine is able to extract information from web pages, allowing for more sophisticated query interfaces and presentation of results adapted to the task.

MultiMatch aimed at taking a significant leap forward from today's vertical search engines, by offering "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, presenting both certified and non-certified information together (while clearly distinguishing one from the other).

The MultiMatch project fits into the category of advanced, domain-specific search engines, with some salient features: i) it is a search engine that combines automatic classification and extraction techniques with semantic web compliant encoding standards; ii) it considers complex user profiles and search scenarios; iii) it is able to search across language boundaries and across different media.

1.1.1.2 Multilingual/multimedia indexing

Instead of returning documents in isolation, MultiMatch provides complex search results that put documents of various media types into context.

For the indexing-end of MultiMatch, complex object retrieval generates special challenges. First, documents of various media types (text, audio, image, video, or mixed-content) and accompanying metadata are indexed. Existing generic standards such as MPEG-7 cater for such data models by incorporating multimedia content and metadata in a single semi-structured document. The indexing strategies used also recognise and cater for multilingual content.

Particular Cultural Heritage relevant Internet domains or subdomains were spidered using a state-of-the-art crawler [Herjtrix, 2005] [Nutch, 2005] and, in parallel, where supported by the CH institution, the engine interfaces with information sources using open standards [OAI, 2005], [Z39.50, 2005]

CH information was also gathered from the Web at large, employing existing focused crawling techniques specifically targeting cultural heritage information [Chakrabarti et al., 1999], [Chakrabarti, 2002].

1.1.1.1 Information extraction and classification

MultiMatch allows users to interpret the wealth of CH information by presenting objects not as isolated individual items, but as situated, richly connected entities. A range of classifications, as well as various links to reviews, experience reports, and general background knowledge, are provided. Documents are classified on the basis of diverse dimensions, such as topical, geographical, and temporal. MultiMatch uses large scale information extraction from documents to identify entities and their relations in large Web corpora [Bruno et al., 2005], [Clough, 2005].

1.1.1.2 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 aimed 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 spoken document retrieval (SDR), 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.

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. MultiMatch developed components for both document and query translation and procedures for matching one against the other. Much effort was dedicated to the building of domain-specific multilingual resources catering for the terminology adopted in the CH domain [Nie et al., 1999].

1.1.1.3 User-centred interaction

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. Users look mainly for concepts (e.g. individuals, facts, places) and far less for features (e.g. mountains, sunset, clouds). A "semantic gap" exists: human beings intrinsically interpret images depending on a subjective viewpoint while computers remain at the most objective and elementary level. To bridge the semantic gap, human intervention is still needed to add high-level features (i.e. metadata) [Wactlar et al., 1999]. 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. Structures for browsing are created, allowing users to explore content or search results following multiple facets. A key research problem for MultiMatch was enabling the user to adequately formulate their query using the language of their choice and specify both low-level and high-level multimedia feature [Petrelli et al., 2006].

1.2 Understanding User Requirements

A good understanding of user requirements is crucial in order to define the system functionalities. The user requirements analysis was based both on previous experience acquired by the Cultural Heritage institutions participating in the project and also on accepted theory in this area. The goal was to identify users and their needs within a predefined and specific context and map, where possible, their requirements to those features which should be offered by MultiMatch. The user requirements study was performed examining data from a number of sources. Interviews in isolation were not sufficient to be able to build a complete picture as users tend to formulate their description of their requirements on the basis of the tools they know. We thus supported the interviews with a set of imaginary but potentially realizable scenarios together with a vision document representing the functionality that should be included in the proposed system in order to give our users a larger picture [Minelli et al., 2006]. Although this study mainly focussed on users of cultural heritage information for professional purposes, we have also studied log data of user queries to a general purpose search engine in order to understand the types of CH query formulated by the general user.

Different classes of users (from the educational, tourism, and cultural heritage professional sectors) were been identified, together with an analysis of the tasks they perform and the scenarios in which MultiMatch can be expected to operate for these users. This study mainly addressed the needs of users that target cultural heritage information for their professional needs. The motivation is that this kind of user already has well-identified requirements and has had experience in trying to satisfy them with the currently available tools. The analysis has aimed at addressing questions such as what users in the cultural heritage domain typically do on a day-to-day basis (i.e. their work tasks), what type of information they need, and how they look for it (i.e. their search behavior), what would these users require from an information system like MultiMatch to enable them to carry out their activities more effectively (i.e. functionality), and how would these users expect MultiMatch to respond to their search requests (i.e. presentation).

In a first stage, we identified a very large set of requirements. We then analyzed this set in order to identify in the order: (1) the most requested (considered as high-priority); (2) those requirements that best matched the previously declared project objectives and vision. Summarizing briefly, we can say that the main findings were that:

- CH professionals do use the internet widely and as part of their daily work routine but they currently depend largely on generic search engines to find the information they need
- 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 exhibitions, different versions of same document
- they tend to be frustrated by the volumes of information available on the same subject and 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 (preferably) or English (optionally).

1.3 System Functionality

According to the user requirements collected during the first stage of the project, the main functionality that MultiMatch offers is powerful and flexible access to Cultural Heritage information available on-line, on the Web as well as in specialized digital libraries or portals. The main user needs are related to (i) the quality of retrieved items – they must be relevant to the user request, and provided by *certified* sources – (ii) easy formulation of the user request, and (iii) easy visualization and aggregation of retrieved items.

The MultiMatch Search Engine enables the user to retrieve cultural objects through different modalities:

- The simplest one is a traditional *free text* search. This search mode is similar to that provided by general purpose search engines, such as Google, with the difference that MultiMatch is expected to provide more precise results – since information is acquired from selected sources containing Cultural Heritage data – and with support for multilingual searches. 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. Multilingual 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 (see Figure 2).

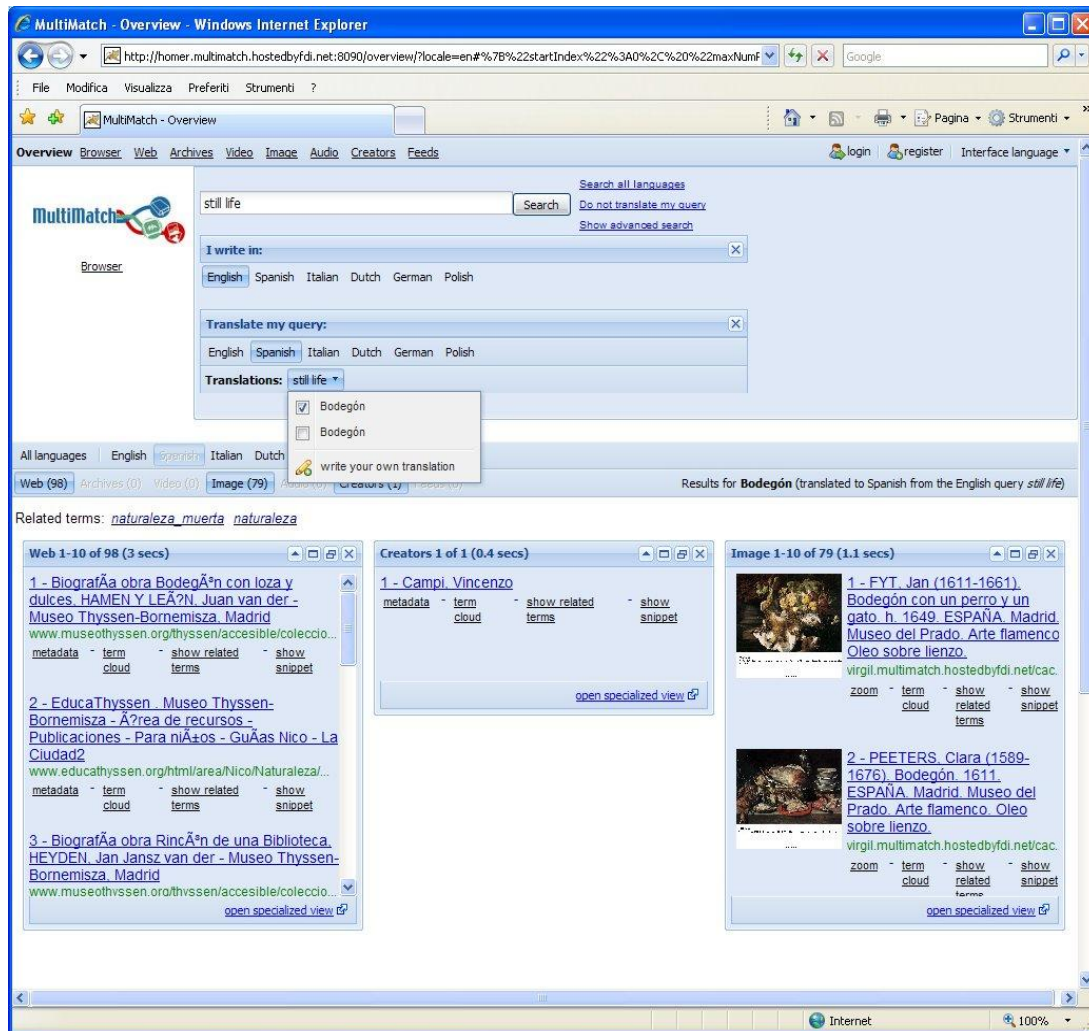


Figure 2 - Example of a multilingual search

- Multimedia searches, based on similarity matching and on automatic information extraction techniques.
- Metadata based searches. The user can select one of the available indexes built for a specific metadata field – initially only *creators* and *creations* – and can specify the value of the metadata field (e.g. the *creator's name*) plus, possible additional terms.
- A browsing capability allows users to navigate the MultiMatch collection using, among others, a web directory-like structure based on the MultiMatch ontology.

Finally, MultiMatch supports multimedia searches, based on similarity matching and on automatic information extraction techniques.

From the results of the expert users survey [Minelli et al., 2006] concluded that, on average, CH professionals tend to classify searches for information about creators (authors, artists, sculptors, composers, etc.) and creations (works of art and masterpieces) as their most common search tasks. Therefore, in MultiMatch we decided to focus on two types of specialized searches for creators and creations, although specialized searches focused on other relevant categories were also considered.

1.3.1 Search Interaction Levels in MultiMatch

A key objective is to provide a system that can be easily adapted to different user needs. For this reason, MultiMatch searches can be made at three main levels of interaction: (i) Default search mode, (ii) Specialised search mode, and (iii) Composite search mode.

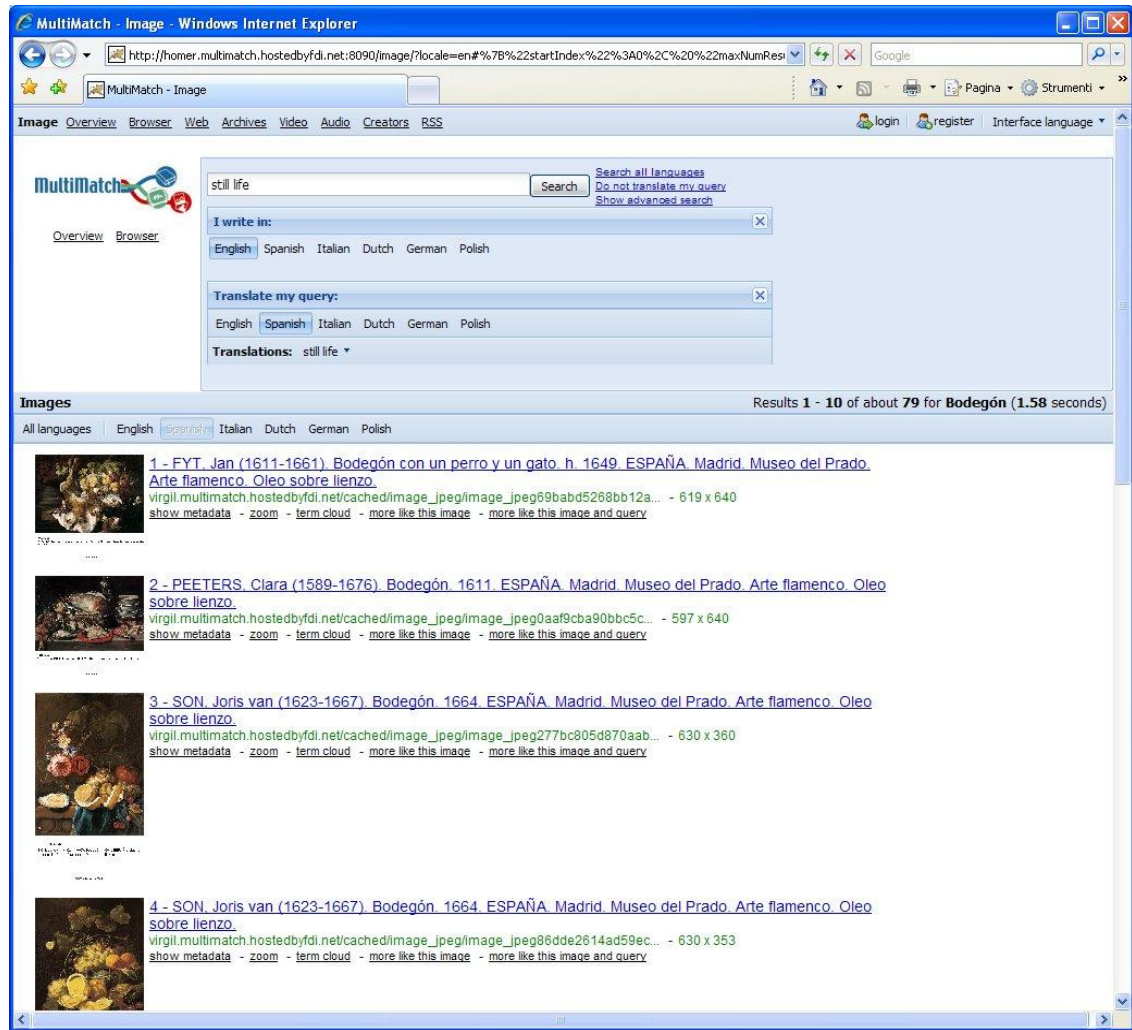


Figure 3 - Example of image specialized view

The simplest search mode is the *default* MultiMatch search level. This is provided for generic users, with a limited knowledge of MultiMatch system capabilities, or with very general search needs. In this case, no assumption is made on the user query, and MultiMatch retrieves information from all indexed material. In this way, given a general query, MultiMatch will retrieve all the cultural objects, web pages and multimedia content that best suit the query. Merging, ranking and classification of these results will be also performed by the system.

The default search level must be understood as a way for the users to express their search needs when they are not looking for information about a specific cultural item (such as a creator or a creation) or when they do not want information in a specific media but want to retrieve all relevant information related to their free text queries.

This interaction level, thus, involves the retrieval of not only cultural objects (i.e. creators and creations) but also web pages, images and videos related to the query. For instance, the query "flowers" should retrieve all cultural objects related with to this topic, such as "the sunflowers" or "waterlily" artwork, but also a ranked list of web pages, a list of relevant news or a list of images and videos.

Users with a more precise knowledge of MultiMatch system functionality, and with specific search needs, may use one of the *specialized* interaction levels available. These allow the user to query MultiMatch specific search services (for instance, video search, image search, etc.) and retrieve all the relevant information available via the selected search service. In this way, MultiMatch will include standalone image, video and metadata-based searches, each with its own search fields, display and refinement options. It will also include a set of browsing capabilities to explore MultiMatch content. Figure 3 gives an example of image similarity search. The specialized interaction level will allow the user to use specific query services, such as metadata-based search, image and video search or browsing.

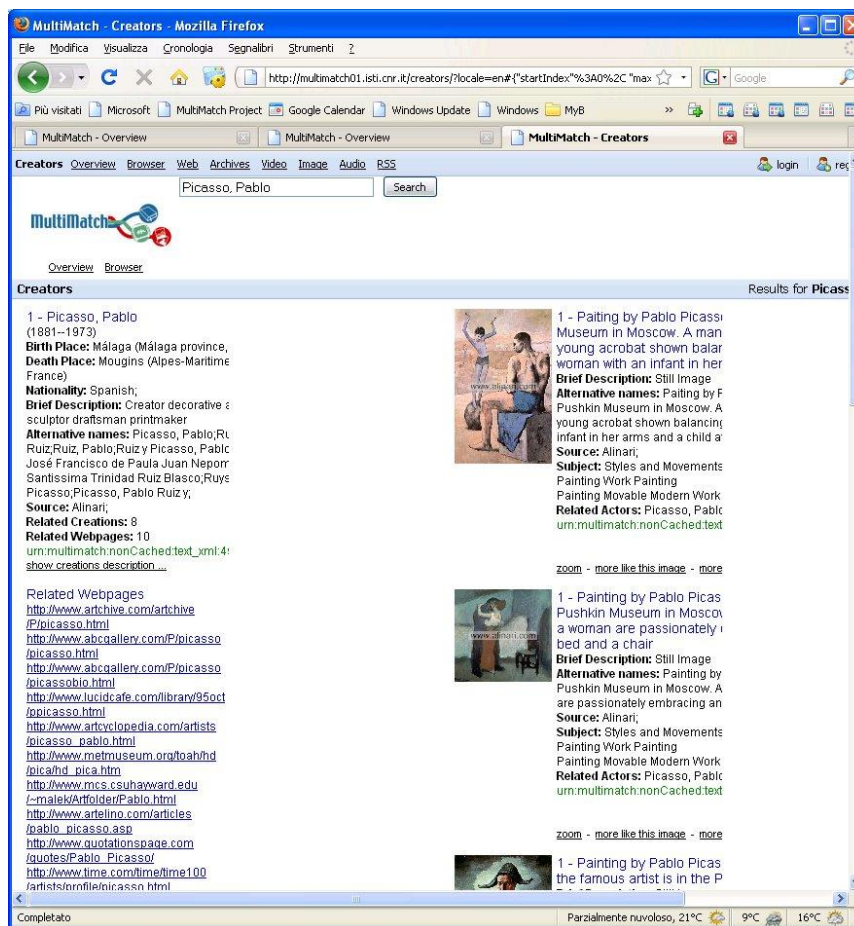


Figure 4 - Browsing Creators

The general idea of metadata based search is that, for a given type of cultural entity (for instance, creators), the whole collection of web pages can be used to mine information about each particular entity that is not present in the individual documents. For instance, the set of all documents talking about Van Gogh can be used to create a profile of the terms most closely associated (i.e. co-occurring more frequently) with Van Gogh. This profile can be subsequently used to compare Van Gogh with other creators. The implication is that for each type of entity considered, MultiMatch must have an index

containing such descriptions. As an example, let us consider that the user intends to search using the *creators* metadata field. The user is expected to type in an author's name plus (optionally) additional free text to retrieve relevant information related to the creator (i.e. artist, composer, writer...). Once the author is input, the system will retrieve different types of information accessible for the user using navigational items such as tabs: for example, it may retrieve web pages with the title plus snippets (as shown in Figure 4) a graphical depiction of the author's network of relationships to other authors, an author's tag cloud with those keywords most representative of the author according to the indexed material, a list of works of art related to the author, a list of cultural heritage sites which host content related to the author.

The *Composite search mode* supports queries where multiple elements can be combined. For example, it is possible to search using the metadata fields associated with each document, but combining this restriction with free text and/or image similarity searches.

Section 2 Relationships with the state of the art

The development of the MultiMatch system implied addressing a number of significant research challenges in a multidisciplinary context. R&D expertise was required in a diverse set of system- and user-oriented research areas. On the system side, these related 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 included 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 report in the initial phases (D.1.1.1) and to provide updates during (D1.1.2) and at the end of the project (D1.1.3)..

The aim was to describe the state of the art in the principal sectors of research covered by MultiMatch in order to establish the scientific and technology baseline from which the consortium partners started. We identified seven main areas: existing technology for cultural heritage, search engines, classification and information extraction, multilingual/multimedia indexing, image collection browsing, multilingual/multimedia retrieval, and user interaction and interface design. Each area is reviewed in a separate chapter of the State of the Art.

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. During the project lifetime, there were major technological developments in this sector. We focused on four main areas where development has been most evident (i) digital library software; (ii) metadata interoperability; (iii) digitization standards; (iv) the impact of Web 2.0 on Cultural Heritage activities.

- With respect to digital libraries, the most important recent development was 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. The consortium thus remained in close contact with the Europeana team in order to monitor and compare progress. While the current version of MultiMatch is far ahead of the first release of Europeana for complex multilingual and multimedia retrieval, it has not been obliged to address the enormous issues of scalability with which Europeana has been battling.
- 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. We thus studied some of the recent advancements in this area. 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, the schemas adopted for MultiMatch were designed to contain all the elements needed to describe the cultural heritage objects within the domain of the project.
- Probably that most relevant digitization standard for MultiMatch is the development of JPEG2000. The consortium has followed carefully the development of this standard in order to ensure compatibility.
- Another issue investigated was the impact of the Semantic Web and Web 2.0 on cultural heritage. There have been profound changes in user requirements since the beginning of the MultiMatch project. 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.

Focussed Search Engines

A search engine can 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. 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 the State-of-the-Art, we briefly surveyed current search engine technology focusing on the areas of main interest to MultiMatch: domain-specific or vertical engines specialising in multimedia and multilingual search and retrieval and in the final version we highlighted the advances made by MultiMatch in this area. MultiMatch's achievements relate well to the current state of the art. Rival multimedia search engine prototypes 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). In the State-of-the-Art we presented 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. 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

We studied the state-of-the-art in the indexing of cultural heritage documents in various languages and of various media types. In particular, we investigated those open problems and challenges that are of most direct relevance to indexing cultural heritage documents as envisioned by the MultiMatch project. We believe that MultiMatch has contributed to advancing the state of the art in the areas of multimedia indexing: structuring and indexing features for spoken audio, handling noise and processing audio from the internet, video classification, complex objects representation, as evidenced in the relevant deliverables and project publications.

Image Collection Overviews and Browsing

Image collection browsing and overviewing 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. We thus evaluated browsing principles and technologies that could be useful in the context of MultiMatch. The objective was to ensure that the MultiMatch 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. Both these problem areas were studied in depth in MultiMatch

User-centred Interaction and Interface Design

The interface acts as the intermediary between users of information retrieval (IR) systems and the search system. We studied users' information seeking behaviour in order to provide informative insight into user interface design. The focus was 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 that were investigated included 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. By examining and testing a variety of designs with potential user groups – including the latest Sematic web interfaces, MultiMatch was able to build an interactive, innovative interface that is first and foremost successful at meeting its users' needs.

Section 3 Impact of the project on the Industry and Research Sectors

MultiMatch's primary scope is to impact the **research field**. The main effect of the MultiMatch project is expected in the research/testing field and in the evaluation of research activities. MultiMatch results will benefit directly and indirectly research groups, industrial partners, content providers

The second channel for the exploitation of the results is represented by the use of MultiMatch in **specific applicative projects**.

Finally MultiMatch could provide products and services for **commercial exploitation**. Both specific components or the complete system could be exploited. We are able to provide a definition of which components and services can be offered, suggesting related exploitation opportunities, but at present the MultiMatch project cannot be exploited as a complete system: additional functionalities should be integrated and further work is needed to engineer MultiMatch (production of documentation, engineering of the components and of the system). We believe MultiMatch technologies could have a market impact in the coming years.

The following deliverables have reported the MultiMatch exploitation planning and the market analysis:

Deliverable 8.2.1 Initial exploitation planning report and market survey (preliminary version): delivered August 2006;

Deliverable 8.2.1 Initial exploitation planning report and market survey (final version): delivered June 2007;

Deliverable 8.2.2 Final exploitation planning report and market survey: to be delivered December 2008

The initial exploitation plan (Del. 8.2.1) defined the exploitation strategy and methodology to be put in place in order to introduce the project outcomes into the market strategies. The overall system as well as individual components were considered. A survey of the competitive solutions on the market was carried out and a first outline of the market analysis of the business targets that the project intended to address was defined. The plan focused on the benefits that MultiMatch could offer to end users and content suppliers. Attention was also paid to Sociological, Technical, Economic and Political (STEP) issues, to Strengths, Weaknesses, Opportunities, Threats, (SWOT) issues and to the exploitation Risk Analysis. An initial IPR report was defined.

The final exploitation plan (D.8.2.2) updated the previous MultiMatch exploitation plan, detailing how the project results can be exploited: how MultiMatch can impact the research field, how it can be employed in specific applicative and how it could provide services for commercial exploitation. D8.2.2 defines which are the exploitable products of MultiMatch, how they could be exploited and which are the constraints of such exploitation. It updates and summarises the intellectual property analysis and describes opportunities for end users and content providers drawing scenarios for future exploitation. It reports individual and joint partners' exploitation plans. The document also reviews the external environment, considering the impact of changes in MultiMatch competitors and potential partners and explores the segmentation of the markets where MultiMatch may operate.

The project exploitation has been coordinated with dissemination activities in order to promote the adoption of the project "vision" and to demonstrate the feasibility of the project approach. Through dissemination events useful contacts have been collected for exploitation purposes (see list of events in Section 4)

Section 4 Dissemination of Knowledge

MultiMatch aimed at developing a search engine specialized for the CH domain; it was expected, however, that the project would have an impact beyond the boundaries of cultural heritage domain, wherever there was a need for complex, domain-specific multilingual search. An adequate and appropriate dissemination of the project results has been crucial in this respect in order to guarantee that information on the project results was disseminated not just to the initially targeted user communities but also to the wider potentially interested society. A detailed dissemination plan was prepared at the beginning of the project and was maintained throughout the entire project period and updated to reflect possible refocusing of the scientific, technical, users' oriented and business work.

The strategy, methodology and dissemination instruments outlined in the dissemination plans were employed by the project consortium to reach MultiMatch user groups/scientific communities and create contacts for the exploitation of the project results. In particular the dissemination activities aimed to: increase awareness of the MultiMatch system and its potential benefits for people interested in or working in the Cultural Heritage domain; encourage CH institutions and content providers to become members of the MultiMatch user groups and to provide content to the MultiMatch archives; obtain feedback from them on the evaluation of the MultiMatch prototypes; investigate the possibility of future exploitation of the project outcomes.

Deliverable 8.1.1 Initial plan for using and disseminating knowledge (delivered August 2006)

Deliverable 8.1.2 Final plan for using and disseminating knowledge (delivered December 2007).

The dissemination activities were based on several disseminations instruments (web –based and paper based).

The following subsections provide relevant details (such as dates, quantitative data, references, web addresses) on such activities and instruments employed.

4.1 Overview Table

The following table gives an overview of the activities that have taken place during the project lifetime. The description of each activity is given in detail in the following subsections.

Planned/actual Dates	Type of activity	Type of audience	Countries addressed	Size of audience	Partner responsible/involved
Mths 1 - 30	Project brochures	General	Worldwide distribution	large	ISTI-CNR, Alinari
Mths 7 - 30	Participation in Conferences / Workshops	Research & Industry	Belgium, Canada, Czech Rep., Hungary, Italy, Portugal, Netherlands, U.K., Germany, France, USA, China, Asia, Slovenia, Spain, Denmark, Greece	variable	All (see section 1.1.2)
Mths 1 - 30	Market competitors monitoring	Research & Industry	The Netherlands, UK, Germany, Italy, Denmark, Cyprus	variable	Various
Mths 15, 18, 19,	Attendance at open days, focus	Industry, general public, higher	UK, Spain, China, Germany, Ireland, Italy,	variable	USFD, UA,

26, 27	groups, seminars, research visits	education researchers, CH professionals, tourist user groups	The Netherlands, Asia		UNED,DCU, BVMC, ISTI-CNR, ALINARI (section 1.2.3)
Mths 3, 19, 20, 29, 30	Organisation of MultiMatch Workshops	Research, Educational, CH, markets	Israel, Italy, Germany, Denmark, Cyprus.	variable	Various (see below for details)
Mth 18	Press Releases	General public	Spain, Italy	large	Alinari, BVMC, ISTI-CNR (section 1.2.9)
Mth 14 & 29	Sponsorship of event	Research	Czech Rep., Denmark	variable	Alinari
Mths 1 - 30	Project website	Publicly available	worldwide	Large	ISTI-CNR, Alinari
Mths 7 - 30	Collaborations and clustering	European projects	Various	Variable	All (section 1.2.11)
Mths 4 - 30	Articles, conference/ workshop papers	Research	Various	Large	Various(see section 1.2.7)
Mths 1 - 30	User group contacts (direct e-mailing, phone calls, face-to-face meetings, conferences)	Education / CH professionals / Tourist	Europe	Variable	Alinari (see section 1.2.12)
Mths 1-30	Newsletters	Research	International	Large	ISTI-CNR

4.2 Project Presentations / Representation at Events

Date	Place	Event	Link	Description
October 2008	Napa Valley, California	ACM 17th Conference on Information and Knowledge Management (CIKM 2008),	http://www.cikm2008.org/	Paper presented by UvA.
October 2008	Würzburg, Germany,.	FGIR Workshop Information Retrieval (WIR 2008),	http://lwa08.informatik.uni-wuerzburg.de/Wiki.jsp?page=FGIR08	Paper presented by UvA

October 2008	Napa Valley, California	Second Workshop on Information Credibility on the Web (WICOW 2008).	http://www.cikm2008.org/	Paper presented by UvA.
September 2008	Aarhus, Denmark,	CLEF 2008 Workshop, Workshop on Cross-Language Information Retrieval and Evaluation,	http://www.clef-campaign.org/	Papers presented by DCU, UvA
July 2008	Singapore, Asia	SIGIR 2008 Workshop on Analytics for Noisy Unstructured Text Data (AND 2008)	http://www.sigir2008.org/	Paper presented by UvA
July 2008	Singapore, Asia	31st Annual International ACM SIGIR Conference (SIGIR 2008)	http://www.sigir2008.org/	Paper presented by UvA
24 July 2008	Singapore, Asia	2nd SIGIR Workshop on Searching Spontaneous Conversational Speech (SSCS 2008),	http://ilps.science.uva.nl/SSCS2008/	Paper presented by UvA, DCU
21-25 July 2008	Patras, Greece	18th European Conference on Artificial Intelligence (ECAI 2008),	http://www.ece.upatras.gr/ecai2008/	Paper presented by UvA
18-20 June 2008	London, UK	Sixth International Workshop on Content-Based Multimedia Indexing	http://cbmi08.qmul.net/	Paper presented by UvA, DCU, Beeld en Geluid
15-20 June 2008	Columbus, Ohio, (Association for Computational Linguistics)	ACL-08: HLT	http://www.ling.ohio-state.edu/acl08/	Paper presented by UvA
5 - 7 June 2008	Slovenia	Culture on line Conference	http://www.ekultura.si/	ISTI-CNR MultiMatch project presentation. Dissemination of the project.
1 June 2008	Tenerife-Spain	SIEDL2008, First Workshop on Semantic Interoperability in the European Digital Library ,	http://multimedia.semanticweb.org/siedl/	Presentation by B&G
21-25 April 2008	Beijing China,	WWW2008,17th International World Wide Web Conference	http://www.www2008.org/	DCU
14-15 April 2008	Maastricht-the Netherlands	DIR 2008 – 7th Ditch-Belgian Information Retrieval Workshop	http://www.lti.u-gent.be/DIR2008/index.php?page=cfp	Paper presented by UvA

9-12 April 2008	Montreal, Canada	Museums and the Web Conference	http://www.archimuse.com/mw2008/	USFD & ISTI-CNR presented a demo of the MultiMatch system
30 March – 3 April 2008	Glasgow, UK	ECIR 2008-30th European Conference on Information Retrieval	http://ecir2008.dcs.gla.ac.uk/	Papers presented by DCU, USFD & UNED,
25-26 Feb 2008	Rome, Italy	EPOCH Final Event	http://www.epoch-net.org/	ISTI-CNR presented the MultiMatch project
31 Jan – 1 Feb 2008	Frankfurt am main, Germany	EDL Project Conference	http://www.edlproject.eu/conference	ISTI-CNR represented MultiMatch
11 Jan 2008	Hyderabad, India	CLIA-2008 Workshop	http://search.iiit.ac.in/CLIA2008/index.html	Paper presentation prepared by DCU
29-30 October 2007	Rome, Italy	DILIGENT 2007. Workshop- “European Information Space: e-Infrastructures, Services and Applications”	http://www.diligentproject.org/index.php?option=com_content&task=view&id=195&Itemid=235	MultiMatch presentation by ISTI-CNR
24-26 October 2007	Toronto , Ontario, Canada	ICHIM 2007- International Cultural Heritage Informatics Meeting	http://www.archimuse.com/ichim07/	Paper presented by USFD & BandG
16-21 September 2007	Budapest, Hungary	ECDL 2007-11 th European Conference on Research and Advanced Technologies on Digital Libraries.	http://www.ecdl2007.org/	A MultiMatch poster prepared by ISTI-CNR and UNED was presented
13-14 September 2007	Gent, Belgium	VVBAD Conference- Information 2007	http://www.vvbad.be/informatie2007	MultiMatch presentation by BandG
10-14 September 2007	Modena, Italy	ICIAP 2007- International Conference on Image Analysis and Processing	http://www.iciap2007.org	ISTI-CNR participated with a paper on MultiMatch
September 12-14, 2007	The Hague, Netherlands	EDLNET-European Digital Library Thematic Network	http://www.europeandigitallibrary.eu/edlnet/	MultiMatch presentation by ISTI-CNR
7-8 September 2007	Lisbon, Portugal	International seminar – Disclosure and preservation: Fostering European Culture in the Digital landscape	http://bnd.bn.pt/seminario-conhecer-preservar/index-eng.html	ISTI_CNR presentation ” Interoperability and Multilinguality: Guaranteeing Information Access over Language Boundaries”

23-27 July 2007	Amsterdam-The Netherlands	SIGIR 2007-30 th Annual International Conference	http://www.sigir2007.org/	Two Paper presentations (UVA and DCU)
9-11 July 2007	University of Amsterdam-The Netherlands	CVIR 2007-ACM international Conference on Image and Video Retrieval	http://www.civr2007.com/	Paper presentation UvA
28 June 2007	Prague, Czech Republic	LaTeCH 2007 Language Technology for Cultural Heritage Data	http://ilk.uvt.nl/latech07/	DCU presented paper on MultiMatch (through Alinari) is official sponsor of LaTECH 2007
6-10 June 2007	Florence-Italy	CEPIC (Coordination of European Pictures Agencies Press Stock Heritage)	http://www.cepic.org/	MultiMatch presentation by Alinari
31 March 2007	Pistoia, Italy	CH workshop, EVA Pistoia		Alinari: project dissemination (brochure distribution) and contacts collection
17-20 March 2007	Hannover	CeBIT,		Alinari: project dissemination (brochure distribution) and contacts collection
13-15 February 2007	London , UK	Search Engine Strategy conference		Alinari: project dissemination a(brochure distribution) and contacts collection
29-30 January 2007.	Padova, Italy	3rd Italian Research Conference on Digital Library Systems		ISTI-CNR presented a paper on MultiMatch activities http://ims.dei.unipd.it/events/2007/ircdl/3rd-IRCDL-program.html
14-15 December 2006.	Florence, Italy	On-line Cultural heritage		Project dissemination by Alinari (brochure distribution)
4-5 December 2006.	Rome, Italy	MICHAEL . International Conference Museums libraries and archives online		Poster presentation and distribution of brochures. Contacts with potential user groups
1 December 2006	Utrecht, NE	Workshop of Dutch-funded CATCH project (Continuous Access to Cultural Heritage)		UVA participated in the workshop for Clustering activity

27-28 November 2006	Vienna (Austrian National Library)	EDL workshop		presentation of the project by BandG
14-15 Nov 2006	Rome, Italy	Multi-quality Approach to Cultural Heritage		presentation of the project by ISTI-CNR
8-10 November 2006	Amsterdam, The Netherlands	Infotrends Amsterdam: Digital Imaging '06		Alinari: project dissemination (brochure distribution) and contacts collection
31 October 2006	Florence, Italy	JPSEARCH meeting		Alinari: contribution to standard+ project dissemination
28 October 2006	Florence, Italy	Alinari museum opening-day		Alinari: project dissemination (brochure distribution)
18-19 October 2006	China	The Europe-China Conference on Intellectual Property in Digital Media (IPDM06)		Alinari: project dissemination (brochure distribution)
21-24 September 2006	Budapest, Hungary	Cost 292/IEEE conference		Alinari: project dissemination (brochure distribution)
13-15 September 2006	Paris, France,	Apple EXPO		Alinari: Face-to-face Meetings with exploitation scope; brochure dissemination during the event.
29-30 June 2006	Luxembourg	DigiCult All Projects Meeting		presentation of the project by ISTI-CNR

4.3 Participation in Focus Groups, Seminars, Research visits, Forums

- During the seminar: “Buscadores y la investigación en humanidades Dirigido por José Antonio Millán” held on 11 June 2008 in Pinar- Madrid- Rafael Carrasco (UA) gave a talk with the title: “¿Qué buscamos en una biblioteca digital?”. The topic was about search engines, including MultiMatch.
http://cedros.residencia.csic.es/docactos/4150/programa_de_mano/programa_de_mano04150001.pdf,
- During a seminar held in the School of Computing- National University of Singapore, on 31st July 2008, Gareth J.F.Jones (DCU) gave a talk about the MultiMatch project: “The MultiMatch Project: Searching Multilingual Multimedia Cultural Heritage Collections.
- Joint meeting with ethnoArc, Pasquale Savino (ISTI-CNR) introduced the MultiMatch project and discussed possible cooperation with etnoArch, Berlin, March 13, 2008.
- Seminar targeting the tourist user group, Julio Gonzalo (UNED) and Rafael Carrasco (UA) introduced the MultiMatch project, Alicante, Spain, November 26-28, 2007.
- Seminar at the Dublin Computational Linguistics Research Seminar Serie (DCU), Dublin, Ireland, 7 December 2007.
- Research visits and seminars: Microsoft Research Asia, Beijing, China. Inst. of Computing Technologies, Chinese Acad. of Sci., Beijing, China, October 2007, DCU.
- Chamber of Commerce open day at University of Sheffield, Sheffield, UK: Jennifer Marlow (USFD) gave a talk about MultiMatch at a University of Sheffield event called “Innovate” that targeted the Chamber of Commerce, October 8, 2007.
- Focus group with MLA.The focus group was part of an MSc project on evaluating MICHAEL (esp. CLIR functionality), Sheffield, UK. Paul Clough (USFD) was involved in the focus group with Kate Fernie (MLA). MultiMatch was presented to the CH Professionals attending the event , July 18, 2007
- Focus Group USFD-MLA, London ,UK. 12 CH people attended the event; MultiMatch was presented to the CH professionals attending the event, June 26, 2007.
- Meetings with University of Florence during 2006: Presentation of the MultiMatch project by Alinari.
- BandG participated in the Collectiewijzer steering board (consortium of Dutch national museums and archives) www.collectiewijzer.nl. MultiMatch was presented to this steering board in September 2006 .
- A collaboration was established by BandG between MultiMatch and DEN (Cultural Heritage Netherlands). D2.1 is available on the homepage of their website www.den.nl. The work was mentioned in the DEN Conference, held in 2006 in Rotterdam. Other websites picked up this news (see, <http://forum.archieven.org/index.php?PHPSESSID=6a98ebc11ea74b70beb9066c3fcc94f3&topic=6828.new> for example:
- Meeting in Florence (Alinari) with Mr. Siro of Autonomy Italy- a multimedia search engine company- to talk about business strategies, 12 September 2006.
- Marketing forum in Southampton (UK), 6-9 September 2006. Alinari met four streams (Brand Strategy, Consumer Dynamics, Media Landscape, The Marketer) covering all the latest developments and key issues within the marketing world. Discussion for marketing opportunities and cooperation strategies with several delegates were carried out; project brochures were distributed; meetings with many marketing managers including Tiscali UK, AOL UK, Getty UK, Fiat car Italy took place.

4.4 Organisation of Workshops

Our strategy in organizing the Workshops was influenced by the need to address different target communities (educational user group, CH user group, general public- tourism - user group, scientific community, Digital Library community)

1. Joint DELOS-MultiMatch workshop on '*Semantic-driven Interoperability for Digital Objects in the Cultural Heritage Domain*'⁴ held in Tirrenia, Pisa, 15 February 2007, in conjunction with the DELOS conference. The workshop was oriented towards the Cultural Heritage and the Digital Library research communities. The original idea for organising this workshop resulted from early discussions within MultiMatch and more specifically in the group working on the definition of the most appropriate metadata schema and overall conceptual framework for the diverse types of information that would be handled by the project. The intention was investigate the current state-of-the-art, and discuss those issues that currently hinder the widespread adoption of standards and impede interoperability; thus providing useful input for the MultiMatch metadata working group. The workshop was organised by ISTI-CNR and was attended by about 40 researchers and practitioners in the IT and cultural heritage sectors. MultiMatch members from Alinari, CNR, BandG and USFD participated actively in the event. Neil Ireson & Johan Oomen presented a paper: Capturing e-culture: Metadata in MultiMatch
2. The MultiMatch workshop "*Multilingual / Multimedia Access to Cultural Heritage*"⁵ was held 20 November 2007, as part of EVA/MINERVA 2007 (Jerusalem, Israel); the workshop was organized by ISTI-CNR. Johan Oomen (B&G), Jennifer Marlowe (USFD), Carol Peters & Franca Debole (ISTI-CNR) presented various aspects of the project and demo of the system. The MultiMatch session represented the occasion to target the CH user group and create contacts with European projects for clustering activities: The European Library (TEL), DELOS, EPOCH, MINERVA.
3. The second workshop on *Searching Spontaneous Conversational Speech*⁶ (SSCS 2008) was held in conjunction with SIGIR 2008 on 24 July in Singapore⁷. The goal of the workshop was to bring the speech community and the information retrieval community together. The workshop was not directly addressing MultiMatch, but thematically was closely linked to the project (both the audio and the video collections that were the subject of investigation in MultiMatch contained spontaneous conversational speech). One of the SSCS 2008 organizers (UVA) was from a MultiMatch consortium member. MultiMatch members submitted two papers: a full paper and a poster paper. A MultiMatch member was involved in the panel session.
4. The MultiMatch workshop "Innovative Approaches for searching and using Cultural Heritage Information for Learning and Teaching"⁸ was held 28th November 2007 as part of Online Educa Berlin Conference (Berlin-Germany). The workshop was mainly oriented to the educational user group. The Workshop was organised by Alinari in collaboration with Online Educa Berlin staff. Contributions to the organization of the event came from the MultiMatch consortium (UNED, ISTI-CNR, U.of Alicante and USFD). The workshop was the ideal context to reach the target user

⁴ http://www.delos.info/index.php?option=com_content&task=view&id=524&Itemid=271

⁵ http://www.digital-heritage.org.il/conferences/2007_Tuesday.html

⁶ <http://ilps.science.uva.nl/SSCS2008/>

⁷ <http://www.sigir2008.org/>

⁸ <http://www.online-educa.com/online-educa-berlin-07>

groups: Educators (instructors and teachers) who create courses and teach using multilingual elearning, resources and tools; Cultural content providers aiming to improve distribution of their content inside the educational market; Technology and service providers (multimedia and software designers) providing a bridge between educators and cultural content providers. The workshop was successfully attended by about 30 participants. Useful feedbacks on the evaluation of the PT1 were collected through the use of ad hoc questionnaires. New contacts for dissemination/ exploitation activities and evaluation scopes have been collected.

5. The MultiMatch workshop “*Multilingual / Multimedia Access to Cultural Heritage*”⁹ was held 4 December 2007, as part of the 2nd DELOS conference¹⁰, (Tirrenia, Pisa-Italy). The event was organized by ISTI-CNR. + Presentations by Ying Zhang (DCU). The workshop was oriented towards the Cultural Heritage and the Digital Library research communities. The event represented the occasion for disseminating and exploiting the project results: collect feedback on the MultiMatch system, gather contacts with new potential content providers.
6. The Workshop on “*Information access to Cultural heritage*”¹¹(IACH) was held 18 September 2008 as part of ECDL (Aarhus, Denmark). It represented the occasion to target both the Digital Library and Cultural Heritage research communities. The goal of the workshop was to promote exchange of ideas concerning creation, storage, retrieval and use of cultural heritage information. In particular, the workshop focused on the question of how keepers of cultural heritage resources can improve information access by applying information technology and principles currently being developed in the area of information access. During this workshop the first prototype was demonstrated and the main results of the MultiMatch project presented. This workshop was organized by UVA, UNED, BandG.
7. The Final MultiMatch workshop: *Multilingual and Multimedia Access to Cultural Heritage Digital Objects: the MultiMatch solution*¹² was held 23 October 2008 as part of VSMM conference¹³ (Limmasol, Cyprus). The main workshop’s goal was to present the results of the MultiMatch project. Participants were involved in the evaluation of PT2: after the demo presentation a questionnaire collected users feedbacks and some of the participants gave their availability to participate in the MultiMatch field trails. The workshop represented the occasion to reach the CH and research community and to propose the starting of clustering activities with other European projects: IMAGINATION, EUROPEANA, MICHAEL. This workshop was organized by Alinari.

4.5 Project Website

The MultiMatch web site was created at the beginning of the project with the following URLs: www.multimatch.eu, www.multimatch.info, www.multimatch.org. The website was updated and maintained for the entire project duration. It is hosted on the ISTI-CNR servers and the dedicated dissemination areas are hosted on Alinari’s servers: (the user does not notice the passage from ISTI-CNR to Alinari servers). During the project lifetime new sections have been added and enriched.

Alinari is responsible for the “News”, “Events”, “Content providers”, “User groups” sections. During the project lifetime the section “Content providers” has been enriched with new content providers; the “News”, “Events”, “User groups” sections have been continuously updated with new information.

⁹ http://www.multimatch.eu/tirrenia_12_07.html

¹⁰ http://www.delos.info/index.php?option=com_content&task=view&id=606&Itemid=337

¹¹ <http://ilps.science.uva.nl/IACH2008/schedule/schedule.html>

¹² <http://www.cs.ucy.ac.cy/conferences/vsmm2008/downloads/MultiMatch-Workshop-Agenda.pdf>

¹³ <http://www.vsmm2008.org/>

ISTI-CNR is responsible for the publicly available areas “Homepage”, “about MultiMatch”, “Consortium”, “Publications”, “Related sites”, “Interesting links” and for the “Workspace” area- restricted to the project Consortium: this is a password protected area which contains all project documentation, information on meetings and phone conferences, events of interest, etc. It also includes a dynamic document management system (BSCW Basic Support for Cooperative Work). The BSCW is used to store all internal working documents, including preliminary versions of the deliverables. The project Wiki, hosted by OCLC, is used as a discussion area.

4.6 Posters/Brochures

Two project brochures have been produced. The brochure was a basic and important dissemination instrument and has been distributed during the participation national and international events, and other kinds of meetings.

The brochures were used to: disseminate the project results, create contacts (for exploitation, clustering, etc.) consolidate the user groups. The first brochure was printed at month 3 of the project. The second brochure was produced after the release of the first prototype.

Several posters have been produced (they are downloadable on the project website). The posters were designed for three main purposes:

- presentations during events/ workshops: during exhibition sessions the posters have been of key importance for presenting MultiMatch system and results.
- web presentations
- inside company/institution dissemination.

Some posters have been designed with a particular focus on the user groups (how to become a MultiMatch registered user etc.) and with textual details; other posters have been designed to map the project goals, the technologies, results and mainly with graphics.

4.7 Publications

A list of publications of MultiMatch-related work by project partners is maintained on the website. The publications produced by the project are listed here:

1. 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.
2. He B., Macdonald C., He J., Ounis I., "An Effective Statistical Approach to Blog Post Opinion Retrieval", ACM 17th Conference on Information and Knowledge Management (CIKM 2008), Napa Valley, ACM, October, 2008.
3. He J., Larson M., de Rijke M., "On the Topical Structure of the Relevance Feedback Set", FGIR Workshop Information Retrieval (WIR 2008), Würzburg, Germany, October, 2008.
4. Tsagkias M., Larson M., Weerkamp W., de Rijke M., "PodCred: A Framework for Analyzing Podcast Preference", Second Workshop on Information Credibility on the Web, Napa Valley, ACM, October, 2008. (won the best paper award at WICOW 2008).
5. Larson M., Newman E., Jones G., "Overview of VideoCLEF 2008: Automatic Generation of Topic-based Feeds for Dual Language Audio-Visual Content (draft)", Working Notes for the CLEF 2008 Workshop, Aarhus, September, 2008.

6. E.Newman and G.J.F.Jones, "DCU at VideoClef 2008", In Proceedings of the CLEF 2008: Workshop on Cross-Language Information Retrieval and Evaluation, Aarhus, Denmark, September 2008.
7. N.O'Hare, P.Wilkins, C.Gurrin, E.Newman, G.J.F.Jones and A.F.Smeaton, "DCU at ImageCLEFPhoto 2008", In Proceedings of the CLEF 2008: Workshop on Cross-Language Information Retrieval and Evaluation, Aarhus, Denmark, September 2008.
8. He J., Zhang X., Weerkamp W., Larson M., "The University of Amsterdam at VideoCLEF 2008", Working Notes for the CLEF 2008 Workshop, Aarhus, Denmark, September, 2008.
9. Larson M., Fernie K., Oomen J., Cigarran J., "Proceedings of the ECDL 2008 Workshop on Information Access to Cultural Heritage", ECDL 2008 Workshop, Aarhus, September, 2008.
10. Carmichael, J., Clough, P., Newman, E., Jones, G. "Multimedia Retrieval in MultiMatch: The Impact of Speech Transcript Errors on Search Behaviour" On-line Proceedings of the Information Access to Cultural Heritage (IACH) Workshop, Aarhus Denmark, September (2008). http://www.multimatch.eu/docs/publications/Carmichael_etal_MultiMatch_IACH2008.pdf
11. Balog K., de Rijke M., Weerkamp W., "Bloggers as Experts", 31st Annual International ACM SIGIR Conference (SIGIR 2008), Singapore, ACM, pp. 753-754, July, 2008.
12. Balog K., Weerkamp W., de Rijke M., "A Few Examples Go A Long Way: Constructing Query Models from Elaborate Query Formulations", 31st Annual International ACM SIGIR Conference (SIGIR 2008), Singapore, ACM, pp. 371-378, July, 2008.
13. 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", In Proceedings of the Workshop on Searching Spontaneous Conversational Speech at Thirty-First Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008), Singapore, July 2008.
14. He J., Weerkamp W., Larson M., de Rijke M., "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.
15. Kohler J., Larson M., de Jong F M G., Kraaij W., Ordelman R J F., "Proceedings of the ACM SIGIR Workshop Searching Spontaneous Conversational Speech", 2nd SIGIR Workshop on Searching Spontaneous Conversational Speech (SSCS 2008), Singapore, Centre for Telematics and Information Technology, Enschede, The Netherlands, July, 2008.
16. Meij E J., Weerkamp W., Balog K., de Rijke M., "Parsimonious Relevance Models", 31st Annual International ACM SIGIR Conference (SIGIR 2008), Singapore, ACM, pp. 817-818, July, 2008
17. Tsagkias M., Larson M., de Rijke M., "Term Clouds as Surrogates for User Generated Speech", 31st Annual International ACM SIGIR Conference (SIGIR 2008), Singapore, ACM, pp. 773-774, July, 2008.
18. Larson M., Newman E., Jones G., "Classification of Dual Language Audio-Visual Content: Introduction to the VideoCLEF 2008 Pilot Benchmark Evaluation Task", In Proceedings of the Workshop on Searching Spontaneous Conversational Speech at 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, July 2008.
19. Weerkamp W., Balog K., de Rijke M., "Finding Key Bloggers, One Post At A Time", 18th European Conference on Artificial Intelligence (ECAI 2008), Patras, Greece, pp. 318-322, July, 2008.

20. Carmichael J., Larson M., Marlow J., Newman E., Clough P., Oomen J., Sav S., "Multimodal Indexing of Electronic Audio-Visual Documents: A Case Study for Cultural Heritage Data", Sixth International Workshop on Content-Based Multimedia Indexing(CBIMI2008), London, 18-20th June, 2008.
21. Weerkamp W. and de Rijke M. Credibility Improves Topical Blog Post Retrieval In ACL08:HLT, June 2008
22. Jones G.J.F. and Li Q. "Focused Browsing: Providing Topical Feedback for Link Selection in Hypertext Browsing". In Proceedings of the 30th European Conference on Information Retrieval Research (ECIR 2008), Glasgow, Scotland, April 2008
23. Clough, P. and Read, S. "Key Design Issues with Visualising Images using Google Earth". In Proceedings of the 30th European Conference on IR Research (ECIR'08), Glasgow, UK, April 2008
24. Marlow, J., Clough, P., Cigarrán Recuero, J. and Artiles, J "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
25. Weerkamp W. and de Rijke M. "Looking at Things Differently: Exploring Perspective Recall for Informal Text Retrieval". In: Proceedings of Dutch-Belgian Information Retrieval Workshop (DIR 2008), 14-15 April 2008
26. He J., Larson M., and de Rijke M. "Using Coherence-based Measures to Predict Query Difficulty". In: C. Macdonald et al., editors, 30th European Conference on Information Retrieval (ECIR 2008), LNCS 4956, pages 689-694, April 2008
27. Balog K. and de Rijke M. "Associating People and Documents". In: C. Macdonald et al., editors, 30th European Conference on Information Retrieval (ECIR 2008), LNCS 4956, pages 296-308, April 2008
28. Marlow, J., Clough, P., Ireson, N., Cigarran Recuero, J., Artiles, J., & Debole, F. "The MultiMatch project: Multilingual/Multimedia access to cultural heritage material on the web". In Proceedings of Museums and the Web 2008, Montreal, Canada, April 9-12, 2008
29. Jones G.J.F., Fantino F., Newman E. and Zhang Y. "Domain-Specific Query Translation for Multilingual Information Access Using Machine Translation Augmented With Dictionaries Mined From Wikipedia". In Proceedings of the 2nd International Workshop on Cross Lingual Information Access - Addressing the Information Need of Multilingual Societies (CLIA-2008), Hyderabad, India, pp 34-41, January 2008.
30. Kludas J., Bruno E., and Marchand-Maillet S. "Can Feature Information Interaction help for Information Fusion in Multimedia Problems?" In Proc. of the first International Workshop on Metadata Mining for Image Understanding (MMIU), Funchal, Madeira, January 2008
31. Snoek C.G.M., Huurnink B., Hollink L. , de Rijke M., Schreiber G., and Worring M "Adding Semantics to Detectors for Video Retrieval" IEEE Transactions on Multimedia 9(5):975-986, August 2007 <http://staff.science.uva.nl/~mdr/Publications/Files/tmm2007.pdf>
32. Ahn D., van Rantwijk J., and de Rijke M. "A Cascaded Machine Learning Approach to Interpreting Temporal Expressions" In proceedings NAACL-HLT 2007, April 2007. <http://staff.science.uva.nl/~mdr/Publications/Files/hlt-naacl-2007-timex.pdf>
33. Casarosa, V. and Peters, C. "DELOS-MultiMatch Workshop on Ontology-Driven Interoperability for Cultural Heritage Digital Objects" Ercim News, No. 69, April 2007 <http://ercim-news.ercim.org/content/view/179/327/>

34. Ireson, N., Oomen, J. "Capturing e-Culture: Metadata in MultiMatch" In Proc. DELOS-MultiMatch workshop, February 2007, Tirrenia, Italy
<http://www.delos.info/files/pdf/DELOS%20Multimatch%202007/papersdelostirrenia.pdf>
35. Amato, G., Cigarran, J., Gonzalo, J., Peters, C., Savino, P. MultiMatch Multilingual/Multimedia Access to Cultural Heritage In Proc IRCDL 07, 3rd Italian Conference on Digital libraries, to be published in Springer LNCS series
36. Clough, P., Al-Maskari, A. and Darwish, K. "Providing multilingual access to FLICKR for Arabic users" Results of the Sixth CLEF Evaluation Campaign. Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany, in print. iCLEF paper from 2006 which should be published 2007 in Springer LNCS. (Evaluation methods-i.e user satisfaction-and correlation with search effectiveness) http://www.multimatch.eu/docs/publications/clough_providing_iclef2006-submission-lncs.pdf
37. Clough, P., Marlow, J. and Sanderson, M. "Designing Multilingual Information Access to Tate Online Workshop" held at the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; Workshop: New Directions in Multilingual Access, Seattle, August 2006 (in print)
http://www.multimatch.eu/docs/publications/clough_designing_sigir2006.pdf
38. Zhang Y., Jones G.J.F. and Zhang K. Dublin City University at CLEF 2007: Cross-Language Speech Retrieval (CL-SR) Experiments. In Advances in Multilingual and Multimodal Information Retrieval. Revised papers of the CLEF 2007: Workshop on Cross-Language Information Retrieval and Evaluation, Budapest, Hungary, LNCS, Springer. To appear.
39. Pecina P. Hoffmannova P., Jones G.J.F., Zhang,Y. and Oard D.W. Overview of the CLEF 2007 Cross-Language Speech Retrieval Track. In Advances in Multilingual and Multimodal Information Retrieval. Revised papers of the CLEF 2007: Workshop on Cross-Language Information Retrieval and Evaluation, Budapest, Hungary, LNCS, Springer. To appear.
40. Larson, M., Jijkoun, V., Löffler, J. and Tjong Kim Sang, E. Practical applications of stand-off annotation. Sprache und Datenverarbeitung: International Journal for Language Data Processing, 31(1-2), 2007, pp 115-129.
41. Schuth A., Marx M. , and de Rijke M. Extracting the Discussion Structure in Comments on News-Articles. In: 9th ACM International Workshop on Web Information and Data Management (WIDM 2007), November 2007
42. Fissaha Adafre S., Jijkoun V. and de Rijke M. Fact Discovery in Wikipedia. In: 2007 IEEE/WIC/ACM International Conference on Web Intelligence, November 2007
43. Szekely E., Bruno E. and Marchand-Maillet S. Clustered Multidimensional Scaling for Exploration in Information Retrieval. In 1st International Conference on the Theory of Information Retrieval (ICTIR'07). Budapest, Hungary 2007
44. Fissaha Adafre S. and de Rijke M. Ask the Crowd to Find Out What's Important. In: ICDM'07 Workshop on Data Mining in Web 2.0 Environments, October 2007
45. Marlow J., Clough, P.D. and Dance K., [Multilingual Needs Of Cultural Heritage Web Site Visitors: A Case Study Of Tate Online](#), Proc. of International Cultural Heritage Informatics Meeting (ICHIM2007), Toronto, Ontario, Canada, October 24-26 2007
46. Amato, G., Magionami, V., Savino, P. Region based image indexing and retrieval inspired by text search. In occasion of the International Conference on Image and Analysis and Processing, ICIAP 2007, Modena, Italia, September 2007.

47. Fissaha Adafre S., Jijkoun, V. and de Rijke, M. Entity Retrieval. In: Recent Advances in Natural Language Processing (RANLP 2007), September 2007
48. Kludas, J. Multimedia Retrieval and Classification for Web Content, BCS IRSG Symposium: Future Directions in Information Access, September 2007.
49. Amato, G., Cigarran, J., Gonzalo, J. Peters, C., Savino, P. MultiMatch Multilingual/Multimedia Access to Cultural Heritage. In ECDL 2007 Proceedings, [Lecture Notes in Computer Science](#) 4675 Springer, 505-508.
50. Zhang, Y., Jones, G.J.F, Zhang, S., Wang, B., Guo, Y. and Ma, Y. An Investigation of Question Translation for English-Chinese Cross-Language Question Answering, 2007 China-Ireland International Information and Communications Technologies, Dublin, Ireland, August 2007.
51. Snoek C.G.M., Huurnink B., Hollink L. , de Rijke M., Schreiber G., and Worring M. (2007). IEEE Transactions on Multimedia 9(5):975-986, August 2007.
52. Al-Maskari, A., Sanderson, M. and Clough, P. (2007) Arabic Users' Satisfaction with the Online Information as Obtained from Google. In Proceedings of the Sixth International Conference on Conceptions of Library and Information Science (COLIS2007), in print.(The effect of cultural background on search).
53. Kludas, J., Bruno, E., Marchand-Maillet, S. (2007) Information Fusion in Multimedia Information Retrieval, 5th international Workshop on Adaptive Multimedia Retrieval (AMR), July 2007 -to appear-
54. Larson, M., Eickeler, S. and Köhler, J. Supporting radio archive workflows with vocabulary independent spoken keyword search. In Proc. of Searching Spontaneous Conversational Speech, ACM SIGIR 2007 Workshop, Amsterdam, The Netherlands, July 2007.
55. Jones, G.J.F., Zhang, Y., Newman E., and Lam-Adesina, A.M., Examining the Contributions of Automatic Speech Transcriptions and Metadata Sources for Searching Spontaneous Conversational Speech. In Proc. of Searching Spontaneous Conversational Speech, ACM SIGIR 2007 Workshop, Amsterdam, The Netherlands, July 2007.
56. Minelli, S., Marlow, J., Clough, P., Cigarran, J., Gonzalo, J., Oomen, J.(2007), [Gathering requirements for multilingual search of audiovisual material in cultural heritage](#). To appear in Proc. of Workshop on User Centricity – state of the art (16th IST Mobile and Wireless Communications Summit), Budapest, Hungary, 1-5 July 2007.
57. Jones, G.J.F., Zhang, Y., Fantino, F., Newman, E., and Debole, F. Multilingual Search for Cultural Heritage Archives by Combining Multiple Translation Resources. In Proc. of the ACL Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007), Prague, Czech Republic, June 2007.
58. Minelli S. and Hyde R., The Impact of the Web and a Reconfigured Social Dynamic on LIS digital libraries. Presented at the conference on Diffusion and dissemination of ideas and experiences: Issues of international comparative. Library Science in Krakow at the University of Jagiellonian- June 2007. (Web-orientation of new librarian's services).
59. Larson, M. and Köhler, J. Structured Audio Player: Supporting Radio Archive Workflows with Automatically Generated Structure Metadata. In Intelligent Access to Audio Content, Proc. of RIAO 2007, 30 May -1 June 2007.
60. Fissaha Adafre S. and de Rijke M. (2007). Estimating Importance Features for Fact Mining (With a Case Study in Biography Mining). In Intelligent Access to Audio Content, Proc. of RIAO 2007, 30 May -1 June 2007.

4.8 Newsletters

Information on the MultiMatch project have been published in internationally distributed newsletters:

- ERCIM News — quarterly newsletter published by the European research Consortium for Informatics and Mathematics. Articles appeared in ERCIM News:

No. 66, July 2006 – see http://www.ercim.org/publication/Ercim_News/enw66/peters.html

No. 69, April 2007-see <http://ercim-news.ercim.org/content/view/179/327/> “MultiMatch Workshop on Ontology-Driven Interoperability for Cultural Heritage Digital Objects”.

No. 70, July 2007 – see <http://ercim-news.ercim.org/images/stories/EN70/EN70-web.pdf> the call for papers of the MultiMatch Workshop “Innovative Approaches for Searching and Using Cultural Heritage Information for Learning and Teaching”.

- IM2 Newsletter – Interactive MultiModal Information Management (an announcement appeared in the Issue No.37, May 2006) see www.im2.ch/newsletters/newsletter-pdf/issue37.pdf

4.9 Press Releases

University of Alicante managed contacts with Spanish news magazines to disseminate OnlineEduca Berlin Workshop.

- Article in BVMdC web: <http://www.cervantesvirtual.com/noticias/noticias.jsp?mes=11&anyo=2007#cod4186>
- Article in newspaper "La Verdad": <http://www.cervantesvirtual.com/AdmNoticiasCervantes/ficheros/imagenes/2007/verdad.jpg>; <http://www.laverdad.es/alicante/20071127/cultura/biblioteca-cervantes-participa-proyecto-20071127.html>
- Article in newspaper "Información": http://www.diarioinformacion.com/secciones/noticia.jsp?pRef=2241_5_696673; Cultura-Biblioteca-Virtual-Miguel-Cervantes-participa-proyecto-europeo-Multimatch <http://www.ua.es/dossierprensa/2007/11/27/10.html>

Through Alinari, Online Educa Berlin Workshop had a dedicated space in *Nova* - technological insert of “IL Sole 24 Ore”.

4.10 Contribution to Standards

MultiMatch consortium partners are involved in a number of standardisation initiatives and used these links both to check that MultiMatch developments adopted relevant standards and to provide contributions where needed. Some of the research results of the project are relevant to some open standards (see below) and the partners involved in these standardization initiatives have acknowledged the project references.

Emerging standards (as for example the new component of JPEG2000 called JPSEARCH or the P-Meta standard for metadata exchange) have been taken into consideration by the project partners for integration after testing.

The scope of dissemination in contributing to standards is to provide international awareness with respect to the project technology.

The list of standards in which MultiMatch partners have shown interest:

- JPEG2000-JPSEARCH: for the search and retrieval architecture (Alinari). Alinari attended the WG1 Jpeg meeting and contributed to the definition of the search query workflow

(wg1n3973_JPSearch_Part-1_PDTR) and the definition of the use cases for jpsearch (which is very relevant to MM). Alinari has also followed the discussion about MPEG7-search and the possible consequences to jpsearch.

- P-Meta standard for metadata exchange (BandG)
- MPEG Video coding (WIND)

4.11 Clustering Activities

Relations with other relevant European projects and networks have been set up through the participation in workshops, conferences and other events (see sections 4.2. and 4.4). Links have been established and proposals to start collaborations with the MultiMatch project have been explored.

The following projects, which fully address issues of importance to the Cultural heritage sector (best practices, services, technologies and services), were contacted:

- **TEL**¹⁴ The European Library was reached through direct contacts and contributed as Content provider to the MultiMatch project.
- **EPOCH**¹⁵ Network of Excellence in Processing Open Cultural Heritage (<http://www.epoch-net.org/>). Contents were established and MultiMatch was presented at the EPOCH final workshop.
- **MINERVA**¹⁶ is a network of Member States' Ministries to discuss, correlate and harmonise activities carried out in digitisation of cultural and scientific content for creating an agreed European common platform, recommendations and guidelines about digitisation, metadata, long-term accessibility and preservation. – The MultiMatch consortium reached MINERVA through direct contacts with the Italian Ministry of culture. MultiMatch was presented at two MINERVA events and discussions were held with Minerva representatives.
- **MICHAEL**¹⁷ Through the multilingual MICHAEL service people are able to find and explore European digital cultural heritage material using the Internet. The MICHAEL consortium is made up of the ministries of culture of France and Italy, the Museums, Libraries and Archives Council of the United Kingdom, supported by the private bodies Dedale, AJLSM and Amitié for technological and administrative aspects. The MultiMatch consortium reached MICHAEL through direct contacts and participation in events (like the final MultiMatch workshop in Cyprus where a representative of MICHAEL was invited speaker at the workshop). Hypothesis of future collaboration have been proposed.
- **IMAGINATION**¹⁸ Image-based Navigation in Multimedia Archives. The MultiMatch consortium established links with this project through direct contacts. One representative of the IMAGINATION was invited speaker at the final MultiMatch workshop in Cyprus. Imagenotion¹⁹ is an application developed by IMAGINATION. Possible cooperation between MultiMatch and ImageNotion are were explored, in particular:
 - Exchange of technical experience: MultiMatch multi-lingual search with automatic query translation, personal workspace of images, search for similar images (in color and texture). IMAGINATION semantic annotation, annotation and navigation through image parts, visual search refinement.

¹⁴ <http://www.theeuropeanlibrary.org>

¹⁵ <http://www.epoch-net.org/>

¹⁶ <http://www.minervaeurope.org/>

¹⁷ <http://www.michael-culture.org/>

¹⁸ <http://www.imagination-project.org/>

¹⁹ www.imagenotion.com

- Common workshops
- Use content of each other
- **DELOS**²⁰ Network of Excellence for Digital Libraries. Coordinated a joint programme of activities of major European teams working in digital library related areas. The MultiMatch consortium took part to two DELOS conferences which represented the occasion to target this network and organised a joint workshop with DELOS on interoperability.
- **CLEF**²¹ Cross Language Evaluation Forum. Evaluation tracks to test different aspects of information retrieval system development. All the MultiMatch research partners are very active in CLEF. MultiMatch sponsored an evaluation track in CLEF 2008
- **EUROPEANA**²² The European digital library, museum and archive The MultiMatch consortium established links with this project through direct contacts (one partner for the MultiMatch consortium is partner in Europeana) Hypothesis for future collaborations have been proposed.
- **CATCH**²³ CommunicAtion Technologies for Cultural Heritage. The MultiMatch consortium established contacts with the Catch project during the workshop held in Utrecht 1st December 2006.
- **EDLNET**²⁴ European Digital Library Network. MultiMatch is represented in EDLNET by ISTI-CNR which is part of EDLNET.
- **ethnoArc**²⁵ – Linked European Archives for Ethnomusicological Research. The MultiMatch consortium established contact with the ethnoArc project by participating to a joint in Berlin, March 13, 2008.

4.12 MultiMatch User Groups

MultiMatch distinguished between a core user group (controlled in size and composed of invited members) expected to provide input for the user requirements and contribute to the evaluation of the system prototypes, and a general user group to which any interested user can belong.

The MultiMatch users have in common the desire or necessity to find and use cultural heritage contents (images, video, audio, multimedia). MultiMatch offered them: services to access multimedia/multilingual cultural contents; possibility to request specific system functionality addressing particular user needs; invitation to MultiMatch workshops/information days; indexing of third party contents (professional contents); involvement in the evaluation activity.

The MultiMatch user group membership was continuously enlarged during the project lifetime and reached the expected numbers:

	Education	General Public	CH
Core	5 to 10	5 to 10	5 to 10
General	>50	>20	>20

1. Educational user group

Users contacted

²⁰ <http://www.delos.info/>

²¹ <http://clef.isti.cnr.it/>

²² <http://dev.europeana.eu/>

²³ <http://www.catchproject.net/index.html>

²⁴ http://ec.europa.eu/information_society/activities/econtentplus/projects/cult/edlnet/index_en.htm

²⁵ <http://www.ethnoarc.org/>

Around 50 users educational users have been involved in the educational user group

Dissemination activities

The educational user group was targeted in different ways by:

- distributing advertising materials (leaflets, brochures, etc.)
- direct contacts (typically making presentations to visitors)
- distributing electronic newsletters, on line questionnaires, emails
- participation in events: conferences, meetings, workshops
- organization of MultiMatch Workshops. In particular “Innovative Approaches for Searching and Using Cultural Heritage Information for Learning and Teaching ”held on 28 November 2007 at Berlin, (Germany) represented an important occasion to target this user group and collect contacts.

2. General public (cultural tourism) user group

Users contacted

Some difficulties were encountered addressing this target community. Face to face meetings with some of the users contacted helped us to point out the difficulties encountered and the possible solutions. We set up a dissemination strategy functional to establish new contacts. Finally, using each partner’s channels we involved a list of more than 20 users defined as “general public”.

Dissemination activities

- Direct contacts: we established contacts with Tourist Portals/organizations in order to offer the potentialities of the MultiMatch system. Each partner from the Consortium representing a different Country tried to establish contacts with tourist portals or organizations.
- General public was reachable through the WIND Libero portal which deals with travel and tourism, together with the services related to communities of users interested in those topics.
- Participation in national and international Events (of particular relevance: a Seminar targeting the tourist user group was held November 2007 in Alicante, see section 2.3 for details)

3. Professional users in Cultural Heritage domains

Users contacted

Around 60 cultural heritage users were involved in the MultiMatch user group. Among them museums, CH institutions, CH networks.

Dissemination activities

- Distribution of advertising materials (leaflets, brochures, etc.)
- direct contacts (typically making presentations to visitors)
- phone calls
- distribution of on line questionnaires, emails
- participation in several events: conferences, meetings, workshops
- organization of MultiMatch workshops. All the MultiMatch workshops that the Consortium organized were addressing the CH user group (see 2.4)

Content providers

The CH users contacted were asked to contribute with their data to MultiMatch. Some of them agreed to allow the MultiMatch system to access some of their contents:

1. AISA photographic agency;
2. The European Library (TEL)-ITA (National Library Florence)
3. Michael Plus ICIMSS (PL)

4. Michael Plus ONB (AT)
5. Michael Plus MLA (UK)
6. Teche RAI (ITA)
7. IL Sole 24 ORE (ITA)-news
8. The Jewish National and University Library
9. Centre virtuelle de la connaissance sur l'Europe (FR)

4. Digital Library community

MultiMatch workshops organized during: 1st DELOS conference (Pisa, 15 February 2007), 2nd DELOS conference (Pisa, December 2007), IACH workshop (held at ECDL 18 September 2008) represented the occasions to reach this target community.

MultiMatch is also represented officially by its coordinators (Carol Peters and Pasquale Savino) in the EDLNet thematic network which is building consensus towards creating The European Digital Library.

Scientific Community

MultiMatch addressed the scientific community in order to create awareness about the MultiMatch project and to get feedback on the scientific outcomes of the project.

The consortium's activity aimed at reaching the target scientific community is part of a permanent dissemination strategy which is focused on:

- Publications: scientific conference/workshop papers, articles in scientific international journals (see section 2.7)
- Participation to national and international events: conferences, workshops, industrial days, meetings (see sections 2.2)
- MultiMatch workshops (see section 2.4)

Appendix A – List of public deliverables

Del No	Deliverable title	Resp.	Delivery date
D1.1.1 D1.1.2 D1.1.3	State of the Art Monitoring Reports	ISTI-CNR	End of October 06 End of October 07 End of July 08
D1.2	User Requirements Analysis	Alinari	End of July 06
D2.1	First Analysis of Ontologies in the CH domain	BandG	End of September 06
D8.1.1 D8.1.2 Final	Knowledge Dissemination Planning Reports	Alinari	End of July 06 End of October07
D1.3	Functional Specification of First Prototype	UNED	End of September 06
D2.2.1 D2.2.2	Semantic web encoding Metadata Schema and Mapping Evaluation and Revision	BandG	End of February 07 End of January08
D7.1	Evaluation methodology	ISTI-CNR	End of April 2007
D1.4	Functional Specification of Second Prototype	UNED	End of July 2007
D8.3.1 D8.3.2	Organization of workshops and project dissemination	Alinari	End of October07 End of October08
D7.3	Evaluation of Second Prototype	ISTI-CNR	End of Sept 08
D7.4	Results of field trials	BandG	End of October08

References

- [Bruno et al., 2005] E. Bruno, N. Moenne-Loccoz, and S. Marchand Maillet. Interactive video retrieval based on multimodal dissimilarity representation. In *1st Workshop on Machine Learning Techniques for Processing Multimedia Content, MLMM'05, Bonn, Germany*, 2005.
- [Chakrabarti, 2002] S. Chakrabarti. *Mining theWeb: Analysis of Hypertext and Semi Structured Data*. Morgan Kaufmann, 2002.
- [Chakrabarti et al., 1999] S. Chakrabarti, M. van den Berg, and B. Dom. Focused crawling: A new approach to topic-specific web resource discovery. *Computer Networks*, 31:1623–1640, 1999.
- [Clough, 2005] P. Clough. Extracting metadata for spatially-aware information retrieval on the internet. In *GIR-05 Workshop at CIKM 2005, Bremen, germany*, 2005.
- [Heritrix, 2005] Heritrix. Internet archive's web crawler project. Technical report, 2005. <http://crawler.archive.org/>.
- [Minelli et al., 2006] S. H. Minelli and al. User requirements analysis. Technical Report D1.2, MultiMatch Project, 2006. Internal project deliverable (restricted distribution).
- [Nie et al., 1999] J.-Y. Nie, M. M. Simard, P. Sabelle, and R. Durand. Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the web. In *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, San Francisco, CA, USA*, 1999.
- [Nutch, 2005] Nutch. Open source web-search software. Technical report, 2005. <http://lucene.apache.org/nutch/>.
- [OAI, 2005] Open archives initiative. Technical report, 2005. <http://www.openarchives.org/>.
- [Petrelli et al., 2006] D. Petrelli, S. Levin, M. Beaulieu, and M. Sanderson. Which user interaction for cross-language information retrieval? design issues and reflections. *special issue on Multilingual Information Systems*, 2006.
- [Rose et al., 2004] D. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of WWW 2004*, May 17-22, 2004.
- [Wactlar et al., 1999] H. Wactlar, M. Christel, Y. Gong, and A. Hauptmann. Lessons learned from the creation and development of a terabyte digital video library. *IEEE Computer*, 32(2):66–73, 1999.
- [Z39.50, 2005] Z39.50. information retrieval application service definition and protocol specification. Technical report, 2005. <http://www.loc.gov/z3950/agency/>.