



**Project no. 033104**

## **MultiMatch**

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

### **D1.2 User Requirements Analysis**

Start Date of Project: 01 May 2006

Duration: 30 Months

Organisation Name of Lead Contractor for this Deliverable: Alinari

Version Final

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

---

## Document Information

Deliverable number: D1.2  
 Deliverable title: User Requirements Analysis  
 Due date of deliverable: 31 October 2006  
 Actual date of deliverable: 15 November 2006  
 Author(s): Sam H. Minelli, Ilaria del Secco, Giovanna Naldi (Alinari), Roberto Bianchettin, Marco Spadoni (WIND), Paul Clough (USFD); Johan Oomen (BandG); Julio Gonzalo, Juan Cigarran Recuero (UNED) Carol Peters (ISTI-CNR)  
 Participant(s): Alinari, USFD, UNED, BandG, WIND, ISTI-CNR  
 Workpackage: 1  
 Workpackage title: User Requirements & Functional Specification  
 Workpackage leader: Julio Gonzalo, UNED  
 Dissemination Level: RE  
 Version: Final  
 Keywords: user requirements, user groups, education, cultural tourism, cultural heritage

## History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
1	19-06-06	draft	Sam H. Minelli, Ilaria del Secco, Giovanna Naldi (Alinari)	Outline of the deliverable
1.1	13-06-06		Roberto Bianchettin Marco Spadoni (WIND)	CH traffic analysis, Section 6
2	14-07-06	draft	Sam H. Minelli(Alinari)	Update of the document structure, assignments, methodology and definition of user groups
			Sam H. Minelli(Alinari)	Traffic analysis reduction and adaptation
			Paul Clough (USFD); Johan Oomen (BandG)	Questionnaire and methodology (to be integrated)
3	17-07-06	draft	Sam H. Minelli (Alinari)	Structure updating (common vision)
4,5,6	19-07-06	draft	Sam H. Minelli (Alinari)	Updating and adding contributions to all sections; general methodology, interview methodology
7	22-07-06; 04-08-06	draft	Sam H. Minelli (Alinari), Julio Gonzalo (UNED) Paul Clough (USFD);	Common vision, methodology, questionnaires, user groups, updates
8	22-09-06	draft	Sam H. Minelli (Alinari),	Review of sections from 1 to 3, input of sections 4,5,6. Definition of the interview data collection and requirements form.
8.1	29-09-06	draft	Juan Cigarran Recuero (UNED)	Educational group defined. Expert Questionnaire included and explained.  Pending: a) Methodology for query log analysis b) Wind questionnaire definition d) analysis of questionnaires

8.3	15-10-06	draft	Sam H. Minelli (Alinari)	Definition of the user requirements
8.5	01-11-06	Pre-final	Sam H. Minelli (Alinari)	Document finalization
9	10-11-06	Final	Sam H. Minelli, Ilaria Del Secco (Alinari), Juan Cigarran Recuero (UNED), Carol Peters (ISTI-CNR)	Review of deliverable and production of final version incorporating reviewer's suggestions

## Abstract

This deliverable presents a detailed user requirements analysis which will provide input for the definition of the functional specifications of the MultiMatch system prototype. A panorama of the potential MultiMatch user is given and the methodology used to identify the user requirements is defined. One hundred person-to-person interviews were conducted with domain experts (educational, tourism, cultural heritage professionals) in order to collect their opinions and needs. The interviews were conducted mainly in face-to-face mode using a questionnaire, and backed-up by a set of scenarios and a vision document in order to give the respondents an idea of the proposed system functionality. In addition to the interviews, we also performed some analysis of logs from the WIND and Alinari portals and examined the results of previous user studies in the Cultural Heritage domain. A large number of potential requirements were identified and analysed. The final result was a subset of those requirements which seemed to represent the major needs of the user groups studied and which also matched the project vision. The output of this analysis is now being used to generate the functional specifications for the MultiMatch system.

## Table of Contents

Abstract .....	3
Executive Summary .....	5
1 Introduction .....	7
1.1 Outline of the Deliverable.....	7
2 The MultiMatch Vision .....	7
3 Methodology .....	9
3.1 Systems Development.....	9
3.2 Requirements Gathering .....	9
3.3 Putting Theory into Practice .....	13
3.4 Interview Methodology.....	16
4 The Users.....	18
4.1 Identifying the Users.....	18
4.2 Characteristics of Expert Users.....	23
5 User Scenarios.....	24
5.1 Scope of the Scenarios .....	24
6 Interviews .....	32
6.1 The Questionnaire .....	32
6.2 Preliminary Results .....	34
7 Past Experience and Query Log Analysis .....	84
7.1 Recent Studies of the Educational and CH Professional Groups.....	84
7.2 CH traffic and Log analysis .....	88
7.3 User Requirements.....	90
8 User Requirements Selection for MultiMatch.....	104
9 Conclusions .....	108
10 References.....	108
Appendix I: The Common Vision .....	110

## Executive Summary

The MultiMatch project (Multilingual/Multimedia Access to Cultural Heritage) plans to develop a multilingual search engine specifically designed for access, organisation and personalised presentation of cultural heritage information. The goal of this deliverable is to specify the approach which will be used in the project to identify the user requirements, i.e. gather and analyse the requirements of target user groups (and individuals), in order to provide input for the definition of the functional specifications of the MultiMatch system.

## User Groups Investigated

Identifying the appropriate user groups and individuals is a key step when defining system requirements; the cultural heritage (CH) partners in MultiMatch have thus worked at identifying the most relevant user groups likely to benefit from the services that will be offered by MultiMatch. Different classes of user (from the educational, tourism, and cultural heritage professional sectors) have 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 addresses 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 behaviour), 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).

It is intended to investigate the needs of the “casual” or non-professional consumer of CH information for the purposes of personal interest, entertainment or travel via an online questionnaire now being included in the WIND Libero portal (<http://www.iol.it/>). The results will be analysed to provide feedback for the functional specifications for the second prototype.

## Methodology Adopted

The user requirements analysis is based both on previous experience acquired by the CH institutions (Alinari and Sound&Vision) and also on accepted theory in this area. The goal has been to identify users and their needs within a predefined and specific context and map, where possible, their requirements to features which should be offered by MultiMatch. The user requirements study has been 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 (reported in Appendix I). Although this study has mainly focussed on users of cultural heritage information for professional purposes, we have also studied log data from the WIND portal in order to understand the types of CH query formulated by the general user (in this case the Italian user).

## Main Results

In a first stage, we identified a very large set of requirements, which is presented in Section 7.3. We then analysed this set in order to identify in the order: (1) the most requested (and thus considered as high-priority); (2) those requirements that best matched the previously declared project objectives and vision. These requirements are listed in Section 8 and are now being transformed into MultiMatch functional specifications in Deliverable 1.3. 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).

These requirements can all be considered in line with the MultiMatch vision.

# 1 Introduction

This document describes the approach used to perform the user requirements analysis for the MultiMatch project. The output of this work package will drive the design, implementation and evaluation of the MultiMatch system.

Identifying the appropriate user groups and individuals is a key step when defining system requirements; the cultural heritage partners have thus worked at identifying the most relevant user groups likely to benefit from the services that will be offered by MultiMatch. Different classes of users (from the educational, tourism, and cultural heritage professional sectors) have 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. 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 behaviour), what these users require from an information system like MultiMatch to enable them to carry out their activities more effectively (i.e. functionality), and how these users would expect MultiMatch to respond to their search requests (i.e. presentation).

## 1.1 Outline of the Deliverable

The goal of this document is to identify users and their needs within a predefined and specific context and to map, where possible, their requirements regarding features to be offered by MultiMatch. The functional specification for the system will be derived from the analysis of user requirements, gathered from representatives in the following user groups: educational, tourism and content/service providers. The requirements analysis will primarily focus on the needs of professional end users who access the services/content offered by third party organisations such as the cultural heritage partners (currently BandG and Alinari). The needs of non-professional users will be investigated in a second stage via online questionnaires and hands-on sessions with the first prototype. This analysis has been based not just on experience but also on the accepted theory in this area.

The Deliverable is organised as follows: In the next section we briefly describe the main ideas underlying the MultiMatch project in order to establish the context within which we are working – this is expanded in the Common Vision of Appendix I. In Section 3, we describe the methodology that has been adopted for the requirements collection and analysis. Sections 4 and 5 are dedicated to a description of the MultiMatch user while Sections 6 and 7 provide an analysis of their requirements. Section 8 discusses the results of this investigation and provides the input that will be used for the definition of the functional specifications in Deliverable 1.3. Section 9 provides some concluding remarks. This Deliverable is accompanied by 2 Annexes: Annex I lists the people and institutions interviewed and the interview modality used; Annex 2 contains the questionnaire used for the interviews.

## 2 The MultiMatch Vision

The MultiMatch technical annex specified several research lines, which must now be made concrete in WP1, mapping input from partners and users into a functional specification for the MultiMatch search engine.



In order to illustrate the vision underlying the MultiMatch project, let us consider an example scenario:

*“During a social dinner, in Geneva (July 2006), Sam was looking at Swedish banknotes with Martha: one of them portrayed the Italian sculptor Giacometti. Sam remembered a famous picture of Giacometti and his sculptures where the artist resembled his works of art. He proposed to Martha that they try to find this picture of Giacometti where the artist is shown together with one of his statues; the picture is famous because although the artist is walking he also appears as a statue. The search was to be done using a mobile device and Google. Sam and other dinner guests tried different combinations of Italian and English keywords and Boolean operators to retrieve that exact image of Giacometti: ‘Giacometti walking and statue’; ‘Giacometti mentre cammina e statua’, etc. However, the Google Image search generated only Giacometti’s works of art or his portraits. No combination of keywords apparently worked. Finally, Martha tried the query with Google Germany and a German query: the result was the exact image that they were looking for.”*

This real-world scenario illustrates one of the motivations behind MultiMatch: the information needed is very likely to be available on the Internet but may only be accessible if the information seeker is able to cross over various boundaries – and an important boundary is that represented by language. Much important information would be available but is rendered transparent because it is expressed or encapsulated in a different language than that normally employed by the user in his/her search. The MultiMatch system will allow the users to use their preferred language to express the information need (or query). The query will activate a multilingual search, gathering the results and organising them appropriately in such a way that the user is able to effectively use them.

Under this vision, the key ideas underlying the MultiMatch vision are: **multiplicity** and **aggregation**.

**Multiplicity:** MultiMatch will display multimedia results to the user in multiple languages, with various options for searching/browsing, and with multiple links between pages and sites. Users will be able to pose queries in their preferred language(s) and retrieve material in all languages handled by the project (i.e. if the search terms are ‘*ritratto di Giacometti e scultura eseguita da lui*,’ the system will also execute queries in English (‘*Giacometti’s portrait and a sculpture made by him*’) and/or other languages). According to the user’s language profile, results in unknown languages will be returned in a way that is interpretable by the user, e.g. with a summary or associated keywords in the user’s preferred language, or even with a translation acquired from an on-line machine translation service.

Users and customers will be able to search text (‘*find critical texts on Giacometti and his period*’), images (‘*Giacometti’s portraits and his sculptures*’), audio and video (‘*Documentary about the life of Giacometti and the places where he lived*’): image search facilities will include text and content-oriented matching; audio and video search will include the capacity to search transcribed speech (at least in English in the project prototype).

**Aggregation:** The system will aggregate results from diverse sources, and depending on the type of query, can include the results of one or more MultiMatch specialized searches. Specialized search services will be activated to interact with the user to retrieve optimised search results: if the user is looking for online cultural events, the specialized search interface will make it possible to specify the type of event (seminar, theatrical representation, dance, etc.), the place, date, etc.

Appendix I gives a set of screen mock-ups to show how we envisage the functionality to be provided by the system. In the following section, we describe the methodology which has been adopted to gather and analyse the user requirements.



## 3 Methodology

### 3.1 Systems Development

Software (and information systems) are generally developed within a lifecycle following an orderly set of activities (see, e.g. [Somerville, 2000; Hoffer et al., 2005; Pressman, 2001; Maciaszek, 2005; Preece et al., 2002:165-199]). The lifecycle identifies aspects such as the modelling approach and development methodology. A typical development lifecycle will consist of the following phases: business analysis, system design, implementation, integration and deployment, and operation and maintenance. The development (or lifecycle) model indicates *how* software is produced. Modern approaches are typified by an iterative and incremental nature in which software is produced in many iterations with each iteration delivering an incremental (and improved) version. Modern methods of this type include: the spiral model [Boehm, 1988], the IBM Rational Unified Process or RUP [Krutchen, 2003], model-driven architecture [Kleppe et al., 2003] and agile software development [Agile, 2003].

#### System Development in MultiMatch

As will be indicated in Deliverable 3.1 (System Architectural Specification), MultiMatch is adopting an agile approach which differs from other approaches in that it aims to deliver frequent versions of the software to the customer (or user) rather than formal deliverables. The approach has the following stages: user requirements, acceptance tests, test-driven development, refactoring and continuous integration. In this document we are mainly concerned with the initial user requirements phase of development, the primary task of which is requirements elicitation (or engineering). In the following sections, we will describe how we plan to gather requirements, perform data collection and define the functional specifications.

### 3.2 Requirements Gathering

The process of determining requirements is to gather or capture *what* a system should do (not how). A requirement is “a statement of a system service or constraint” [Kotonya and Sommerville, 1998] or “a statement about an intended product that specifies *what* it should do or how it should perform” [Preece et al., 2002:204]. The whole point of gathering requirements is to: (1) identify users’ needs and (2) generate a set of stable requirements [Preece et al., 2002:202]. The first step aims to understand as much as possible about the users, their work, and the context of their work so that the system being built will meet their goals. The second step aims to produce a set of requirements which are based on user needs and provide a foundation from which to continue with the design stage.

A requirement is “a statement of a system service or constraint” [Kotonya and Sommerville, 1998]. A service statement describes how a system should *behave* in response to an individual user or group of users (a community). A constraint statement expresses a restriction on the system’s behaviour (e.g. the system must run on a particular operating platform). Requirements are typically categorised as functional and non-functional. The service statements can constitute functional requirements and typically fall into requirements which describe the scope of the system, business functions and required data structures. The constraint statements constitute non-functional requirements which can be divided into requirements addressing aspects such as: usability, reusability, reliability, performance, and efficiency [Lethbridge and Laganier, 2001]. A further categorisation of requirements is the following:

1. **Functional requirements** capture what a product/system should do. In particular, MultiMatch is interested in collecting requirements about methods of executing searches and generating search results.
2. **Data requirements** capture attributes about the data involved during interaction with a system. MultiMatch is interested in identifying requirements related to persistence, amount and type of contents.

3. **Environmental requirements** refer to the circumstances in which an interactive product/system will be expected to operate. The most important environmental requirement related to MultiMatch is that it will be accessed via the Web.
4. **User requirements** capture the requirements of the intended user group (e.g. the abilities and skills of a user, the type of user, user preferences etc.). The collection of attributes for a “typical user” is called a *user profile*.
5. **Usability requirements** capture the usability goals and associated measures for a particular product/system. Other requirements will be collected for the second prototype: the expert rating will be used to create a product that should be usable and useful (by identifying which features are definitely needed for the final product, and which ones are not crucial but would be nice to have.)

### 3.2.1 Requirements Elicitation and Data Collection

Traditional methods of collecting system requirements include the following [Hoffer et al., 2005:164]:

- Individually interviewing people who work with current systems about future system needs.
- Interviewing groups of people with diverse needs to find synergies and contrasts among system requirements.
- Observing workers in action to see how data are handled and what information people need to do their jobs.
- Studying business documents to discover reported issues, policies, rules and directions as well as concrete examples of the use of data and information in the organisation.

Current practice suggests that the results from using different data collection methods should be combined to create a more complete set of requirements. This is called *triangulation* (see, e.g. Ingwersen and Järvelin, 2005:93; Preece et al., 2002:317-337]. Preece et al. [2002:214] describe a summary of data gathering techniques that can be used in requirements gathering which include the following:

- **Questionnaires** are a series of specific questions designed to elicit specific information from users. They produce quantitative and qualitative data and are good for reaching large numbers of people with relatively low resources. However, questionnaires must be designed carefully to ensure that questions are correctly interpreted by respondents. An additional drawback of questionnaires is that response rates may be low.
- **Interviews** involve asking someone a set of questions and can be held face-to-face or over the phone. They are good for exploring issues, as questions can be guided and clarified by the interviewer. Scenarios can be used in interviews to get people to describe their day-to-day activities, but a more accurate approach in this regard is naturalistic observation (see below.) Interviews mainly produce qualitative data, but some quantitative data can be generated. However, this approach is time-consuming.
- **Focus groups and workshops** are similar to interviews but are conducted with a group of people. They are used to promote discussion and issues regarding requirements. This method is good for collecting multiple viewpoints and can highlight areas of conflict or consensus. However, the group sessions need to be moderated in order to ensure everyone is able to express their views. Focus groups mainly generate qualitative data.
- **Naturalistic observation** involves shadowing users in their own environment as they carry out their work tasks. The process involves spending time with the users and ranges from no involvement from the user at one end (outside observation) to full cooperation from the user at the other (participatory observation). Observing people at work can highlight aspects which other techniques cannot reveal. However, this approach is often very time-consuming and generates a large amount of data.
- **Studying documentation** relates to the studying of written rules and procedures. It is often a good way of learning about how organisations operate and the constraints in which users are

performing their tasks. It requires no time commitment from the users, but users may have developed shortcuts or alternative methods and hence not actually follow the official documented procedures in their day-to-day work.

Further techniques for data collection include:

- **Analysis of log files:** the operations carried out by existing operational systems are often captured in log files (e.g. transactions carried out by a Web server). These can be examined to provide information that complements other methods of data collection. However, this method only explains what has happened and cannot explain *why* something has occurred.
- **Competitor analysis:** this approach is typically used as part of a business modelling phase. The aim of competitor analysis is to compare industries and features [Goto and Cotler, 2005: pp260-276]. It provides a snapshot of a marketplace from a customer's viewpoint including the services and features offered by companies. For example, when redesigning websites, competitor analysis is often performed as part of the web design process. The goal of the analysis is to evaluate the features, technology, content, usability and overall effectiveness of services available to customers or users within a domain.

Deciding on which data collection techniques to use is often difficult and depends on a number of factors, e.g. whether the information required is qualitative or quantitative, the current stage of the project lifecycle (i.e. at the beginning it is likely that there are fewer questions so it may be better to explore issues with interviews rather than questionnaires), available resources, access to stakeholders in the project, the nature of the data gathering technique, the task to be studied and the type of information required.

### 3.2.2 Requirements Analysis

Once a set of requirements has been drafted, these must then be analysed to remove overlap or conflict. Requirements must also be grouped and organised appropriately to facilitate appropriate change control in following iterative cycles of system development (e.g. the process by which new requirements are added or defunct ones removed). Requirements are typically expressed as natural language statements, e.g. "the system will schedule the next phone call to a customer upon the telemarketer's request" [Maciaszek, 2005:60]. Each requirement must be identifiable and is typically assigned a unique identifier or sequential number within a document hierarchy. This is important as it enables evaluation of the system against specific requirements and facilitates change control. Requirements can be structured hierarchically in a parent-child relationship (e.g. 1, 1.1, 1.2, 1.2.1 etc.). This allows requirements to be viewed at different levels of abstraction.

When requirements have been identified and clearly numbered, a requirements dependency matrix or interaction matrix can be constructed. The matrix lists requirements identifiers in sorted order in the row and column headings and then allows the easier identification of conflicting and overlapping requirements. Once overlaps and conflicts have been removed, a subset of working requirements must be selected from the complete set which will be taken to design and implementation. Strategies for selecting requirements are based on prioritising them according to risk, feasibility and available resources. Ideally the priority of requirements should also be ascertained from the customers (or users) via discussions or questionnaires. Requirements can be prioritised as, e.g. "high", "medium", "low", "not sure".

### 3.2.3 Task Description and Analysis

The following definition of task analysis is taken from Wikipedia<sup>1</sup>:

*"Task analysis is the analysis or a breakdown of exactly how a task is accomplished, such as what sub-tasks are required. This information can then be used for many purposes, such as improving the design of tools or procedures that aid in performing the task."*

---

<sup>1</sup> [http://en.wikipedia.org/wiki/Task\\_analysis](http://en.wikipedia.org/wiki/Task_analysis) [site visited: 02/08/2006]

Task analysis is mainly used to investigate an existing situation and to envision new systems or devices. The aim of the technique is to understand what people are doing and why. The most widely used version is hierarchical task analysis which is described in more detail in [Preece et al., 2002:231-234]. This identifies the activities performed within a task but not the sequence in which activities are performed. This is captured using methods such as a procedural analysis or a use case.

There are at least three different ways of describing tasks: (1) scenarios, (2) use cases and (3) essential use cases [Preece et al., 2002:222-234]. A large amount of research and study has been undertaken in this area, see for example [Ingwersen and Järvelin, 2005] and [Hackos and Redish, 1998].

The *scenario* is an “informal narrative description” of human activities or tasks in a story. This is a natural way for people to describe their tasks and typically does not include information about particular systems or technologies to support the task. These descriptions can then be analysed to extract requirements (and also to build up models of the domain, e.g. building class diagrams based on extracting nouns from the narrative).

The use case is also focused on user goals, but with an emphasis on user-system interaction rather than the task itself. The use case is a structured description of what a new system should do or what an existing system already does. The use case provides a complete description of a particular system function and is a sequence of actions an actor performs within a system to achieve a particular goal. A use case is a “case of using the (prospective) system” (used to specify user’s functional requirements) and a scenario specifies a flow of events. These are written in natural language or expressed in an activity diagram.

Essential use cases combine both scenarios and use cases and represent a more general case than a scenario encapsulates.

### 3.2.4 Understanding Users and Tasks

Task analysis aims to understand the users, the tasks that they perform and the environment in which they perform these tasks. The whole point of any new product or service is to help people to do *things* better. These “things” can include activities such as carrying out work tasks, browsing for information and entertainment. Studying users can answer questions such as the following [Hackos and Redish, 1998:26]:

- What are the individual characteristics of the users that may affect their behaviour with the system designed? For example, users will have different learning styles which will affect the way in which an interface is used and information is managed (see, e.g. [Ingwersen and Järvelin, 2005:214-217]).
- What experiences and knowledge do they bring with them to perform the tasks the job requires? For example, what language skills do they have? How long have they been doing the tasks and how did they learn to perform their tasks?
- What do they know about the subject matter and the tools they use to currently perform the tasks today?
- What is their experience with using existing tools and technologies? What tools do they currently use? Are they happy with those tools and how would they like to see the tools extended?
- What are their actual jobs and tasks? What types of searches do they perform? How do they go about looking for information to fulfil their tasks?

Users can include primary users, secondary users and communities of users. They can have varying levels of experience and be classified as novices, advanced beginners, competent performers and experts. To understand the users, a list of potential users and user groups are defined and, for each user/task, a user/characteristic matrix can be created to help model a community of users. The typical characteristics of a user group are then defined and modelled.

The types of task analysis that are often performed include workflow analysis and job analysis. The former defines how a particular process is accomplished; the latter aims to understand all the work that

one person in a certain position does. MultiMatch will strive to understand the users and their tasks as one aspect of the requirements gathering process.

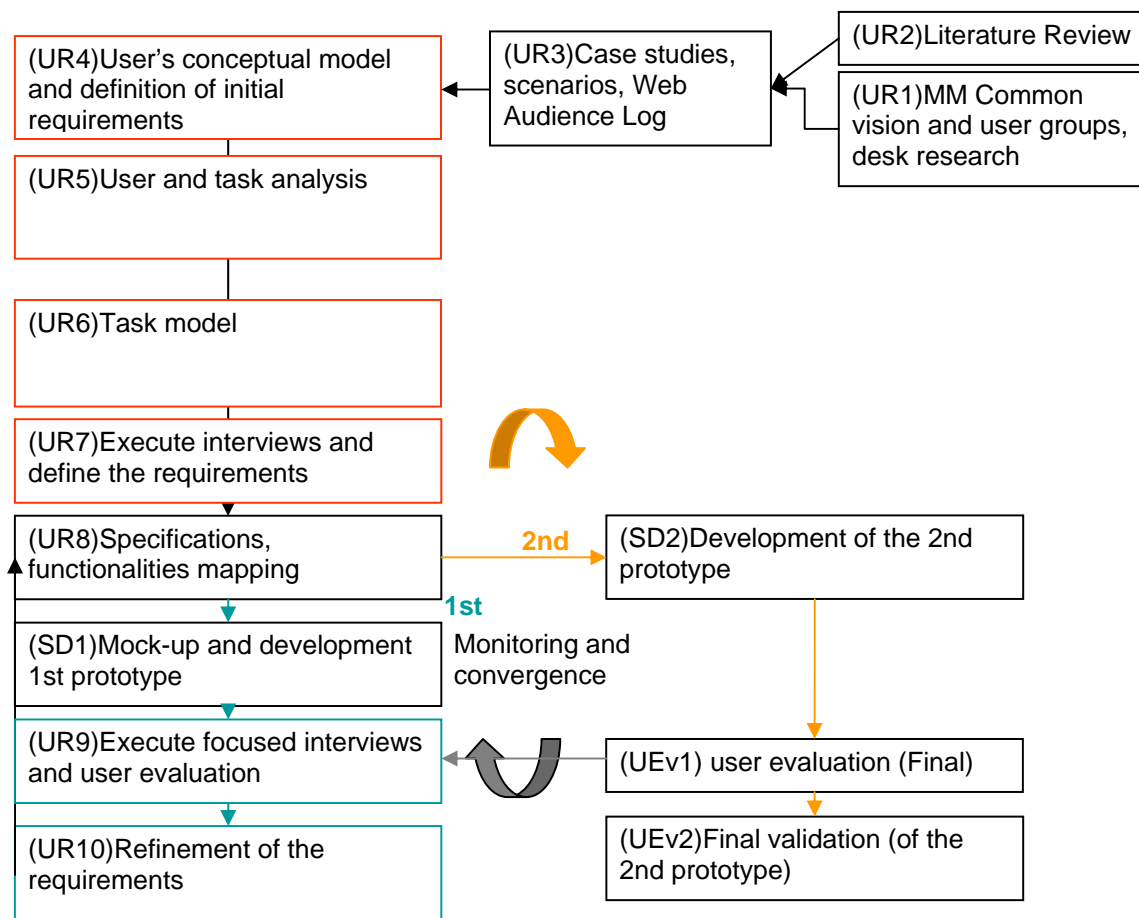
The process of requirements gathering is typically incremental where some aspects of the current and required systems are determined, the requirements are structured and prototypes developed, feedback from users about the prototypes is given, and further analysis is conducted, leading to more refined requirements and consideration of further needs of the customer and system. Eventually this work converges into current operations and the requirements of the new system. In MultiMatch, requirements are gathered in two cycles.

### 3.3 Putting Theory into Practice

This section will describe the practical steps to be applied in the definition of the user requirements from the three user groups on which MultiMatch is focused.

The methodology that we adopt (see Figure 1) takes into account two steps of the MultiMatch prototype development: initial requirements will be collected via interviews. The interviews will aim to validate of the initial MultiMatch vision and to identify any aspects that had been ignored or overlooked. During the second round of interviewing, usability and additional aspects regarding validation will be included in the questionnaires.

The target user groups have been defined and the questionnaires have been collected focusing on their profiles. In the rest of this deliverable, we will refer to User Requirements with UR and a reference number where an activity is related to it<sup>2</sup>. In this section and in Figure 1, we list the steps involved



**Figure 1: Methodology for the definition of the interviews and user requirements. [Maciaszek L.A, 2005].**

<sup>2</sup> Similarly, throughout this document we will refer to Software Development with SD and User Evaluation with UEv.



### UR1 - Literature review

The literature review will refer to the state of the art, the existing knowledge related to existing scenarios and to past project results<sup>3</sup>.

### UR2 - Common vision and user groups, desk research

The common vision has been defined and existing user groups (professional users and customers) have been surveyed by Alinari, BandG and UNED (see Annexes) to identify, through desk research, the main aspects and features which will be addressed by MultiMatch.

### UR3 - Case studies, scenarios, Web Audience Log

On the basis of the previous two activities, specific case studies and scenarios have been created and web data (log statistics) used to verify them. This activity has aimed at identifying user trends and interests and at modelling realistic scenarios (which will be presented in Section **Errore. L'origine riferimento non è stata trovata.**).

Through scenarios related to the user groups that have been selected as being the main MultiMatch users and customers (educational, cultural tourism, cultural heritage), we illustrate for various contexts the support that MultiMatch could give to the users pursuing their goals as students or teachers, tourists, and cultural heritage stakeholders.

The scenarios serve to match the project goals and features against prospective user groups, and to clarify within the project the application system features and design options.

As evidenced in the figure, part of the requirements will be generated from:

- The Web Audience log analysis from the CH partners in the project, including a detailed analysis of the usage of their site search facilities, and a detailed analysis of the queries which drive users to their sites in major search engines (Google, Yahoo!, MSN).
- Desk research (Analysis of the outcome of user studies in previous related projects) and literature

### UR4 - Users' conceptual model and definition of initial requirements

After collecting information about the user groups, and creating case studies or scenarios, we generate a conceptual model on which we define the initial requirements for MultiMatch system.

### UR5 - User and task analysis

The outcome of the previous activities should be used to generate a set of tasks and their analysis (relevance to MultiMatch, feasibility, priority).

- User
- User's goals
- User's environments

### UR6 - Task model

Task modelling will draft the user profiles and tasks in one global model

---

<sup>3</sup> For example, Alinari has taken advantage of its experience studying user needs in the IST project aceMedia for knowledge assisted, adaptive multimedia content management, see <http://www.acemedia.org>

- User profiles
- Task analysis
- Environment profiles

### **UR7 - Execute interviews and define the requirements**

An initial interview activity will collect expert scenarios (in order to refine existing ones) and requirements. With the scenarios, we draw the most relevant tasks and provide a raw modelling of the user profile and environment profile and on this basis we create the questionnaire that scans the scenarios and user tasks as hypothetical requirements to be validated or enhanced. The results of the interviews will be used in order to define the requirements of MultiMatch.

During the development of the first prototype, in-depth focused interviews will be performed to converge and refine the requirements.

To activate a fast feedback process, the interviews have been executed by web interviewing and direct interviews (phone calls, face-to-face) with available expert groups and professional users. The educational group has been explored by UNED (and USFD), the cultural tourism group by BandG, the CH professionals by Alinari (with experts already on the field).

We planned different approaches for the execution of the interviews:

- web/mail (e.g. the WIND portal, Libero)
- direct phone interviews
- face-to-face interviews
- focus group sessions

The interview response collection is a time consuming activity. In particular, the direct face-to-face interview approach is sometimes very slow (involving contacting people, finding a suitable time for an interview, explaining the questions). However, the collected and automatically (where feasible) processed results have been analysed and initial requirements will be collected in this deliverable.

### **UR8 - Specifications, functionalities mapping**

The first expected result of these activities is the definition of the user requirements and the mapping of their needs to MultiMatch functionalities. This activity is crucial as it provides the input to the development of the MultiMatch design.

One major issue is that developers need to check the user scenarios for their technical validity, i.e., if the functions and qualities illustrated by the scenarios are valid according to technical ideas and specifications. The scenarios may be visionary, but must be in compliance with the project goals.

### **SD1 - Mock-up and development of 1st prototype**

An initial interface mock-up and consequent first development of MultiMatch will be done on the basis of the requirements that have been identified to match the project goals and priorities.

### **UR9 -Execute focused interviews and user evaluation**

Using the first prototype, we will reprocess the requirements (UR8) with revised questionnaires to verify whether the first prototype is in line with MultiMatch scopes and the user requirements. In this second stage, more in depth and face to face (plus focus groups and workshops) interviews to experts will be performed picking up on the functionalities which MultiMatch is going to provide. At the same time end user opinions (evaluations, usability and feedback) will be collected.

We will execute the interviews again (2nd interview activity), with direct / focus group interviews based on the working system, comprehensive of usability aspects



### **UR10 - Refinement of the requirements**

The user requirements will be refined and the high priority requirements will be identified.

The first functional specifications (UR8) will provide input for the technical work packages. The second specification (UR11) will collect feedback from development work packages using the results from the evaluation of the first prototype, and producing a second refined functional specification which aims at solving the possible problems detected while maintaining technical feasibility.

### **SD2 - Development of the 2nd prototype**

The initial system will be finalized and consolidated integrating all planned components.

### **UEv1 - User evaluation**

System development will be evaluated and monitored via user testing and feedback.

### **UEv2 - Final validation (of the 2nd prototype)**

The final validation of MultiMatch will verify its efficiency, level of performance and usability

The second interview process will serve not only to refine the functional requirements but also will provide interface usability feedback.

One of the advantages offered by this methodology (creation of group scenarios and execution of interviews) is to facilitate communication between all stakeholders in the project (such as developers, users, domain experts, and managers) and to evaluate the real applications from different perspectives. The scenarios will be:

- accessible to all stakeholders
- less liable to misunderstanding than abstract high-level descriptions
- inspiring
- probing the project idea for viability in various user contexts and domains
- practical

The interview methodology will be presented in the following subsection.

## **3.4 Interview Methodology**

The instruments used to perform the interviews are mainly:

- Web/e-mail/direct phone interview
- Face to face interviews with the selected user groups
- Focus groups or discussion groups
- The outcome of the interviews will generate the MultiMatch requirements. The questionnaires have been defined with respect to the following guidelines:
- Brevity and interest: the respondent is rarely keen to answer long and boring questions
- Closed questions: only professional and expert users will be asked to answer open questions such as a request to suggest particular functionality
- Privacy protection: only professional users and experts will be requested to provide personal data (age, gender, etc.). Collecting personal data from the general public is not possible for reasons of privacy.

- Semi-automatic data collecting: automated analysis systems (not correctly tuned) could misinterpret the results.

Face to face interviews and the focus group interviewing processes need to be planned and the interviewer must be trained. The big difference between these approaches and the web/mail/phone interview is the granularity of the results and the speed of the process: the latter option is generally faster but less detailed than the face to face and the focus group approaches.

MultiMatch intends to mainly employ a web/mail/direct phone interview procedure to collect the most relevant requirements in order to produce initial input for the first prototype. The requirements obtained in this first stage will be further refined (focusing on specific aspects) during the project lifetime using the second two approaches. However, in this deliverable we describe the methodology that will be adopted in all three approaches.

In this section, we will describe the interview workflow. We will refer to User Interaction with UI and a reference number to identify it.

### UI1 - The presentation session with the respondent

MultiMatch executed initial interviews with sector experts (the list of respondents are reported in Annex I) starting from the Advisory board and enlarging it to other field experts to collect their requirements. The interview process was composed of a **presentation session** where the interviewer introduced the user to the project goals while, in the second phase, the user will be introduced to an operational version of the MultiMatch system with a more complete presentation.

The initial interviews were performed by direct contact. The second phase of interviewing activity will perform face to face and focus groups interviews with the following supporting material:

- A consent and privacy protection form to be signed by the user
- A standard questionnaire form
- A set procedure for interviewing and analysing the results
- An introductory section that defines the services to be offered by MultiMatch

The direct interview process is complex but guarantees precision and fine granularity of requirements definition. Direct interviewing requires experts and only partial automated analysis can be performed to interpret the results.

Once the user has an idea of the system envisioned by the project, the interview will be executed with the **questionnaires**.

### UI2 - System testing

The working system will be made available for the second interviewing activities; in the first phase we provided a project outline. During the second phase, the user will be asked to test the current version of the system with a set of assigned tasks.

### UI3 - The questionnaires

A questionnaire has been prepared and used (see Annex II.)

The questionnaire has been designed for the following groups:

1. Cultural Heritage professional users
2. Cultural tourism
3. Educational

The questionnaires that have been generated for the first round of interviews will be extended for the second round to gather other information such as usability and satisfaction.

The user evaluation will generate **feedback** that will be both used to refine the requirements and the questionnaires themselves (only for the first interview set).

- Via email: The targeted respondents will receive an email message with a short description of the MultiMatch service and the goals of the interview. In this message the URL of the platform with a link to the evaluation module will be provided. These targeted respondents will partly have been approached already about participation in the MultiMatch, but nevertheless the expected response rate is comparatively low if not under existing contact agreements (email surveys yield on average a response rate of 10 to 20%). See guidelines, table below.
- Face-to-face / phone calls: Preselected groups will be approached face-to-face. These groups will have agreed beforehand to participate in the interview. During the face-to-face meeting the respondents will receive an introduction to the MultiMatch service and the goals of the interview orally. Afterwards they will be asked to go through the evaluation module and complete the web-based survey. This approach is expected to yield a very high response rate (about 100%).
- Via the internet / web forums: An announcement of the interview together with a link to the MultiMatch platform and the evaluation module will also be published on the websites of the partner institutions, so that other interested members of the target group can also take the survey. A disadvantage of this approach is the low degree of control over the respondents. The expected response rate is even lower than that of the email distribution. But certainly the number of responses could be potentially very large.

## UI5 - Verify and monitor

This activity aims at collecting the results, analysing them and evaluating the quality of the method.

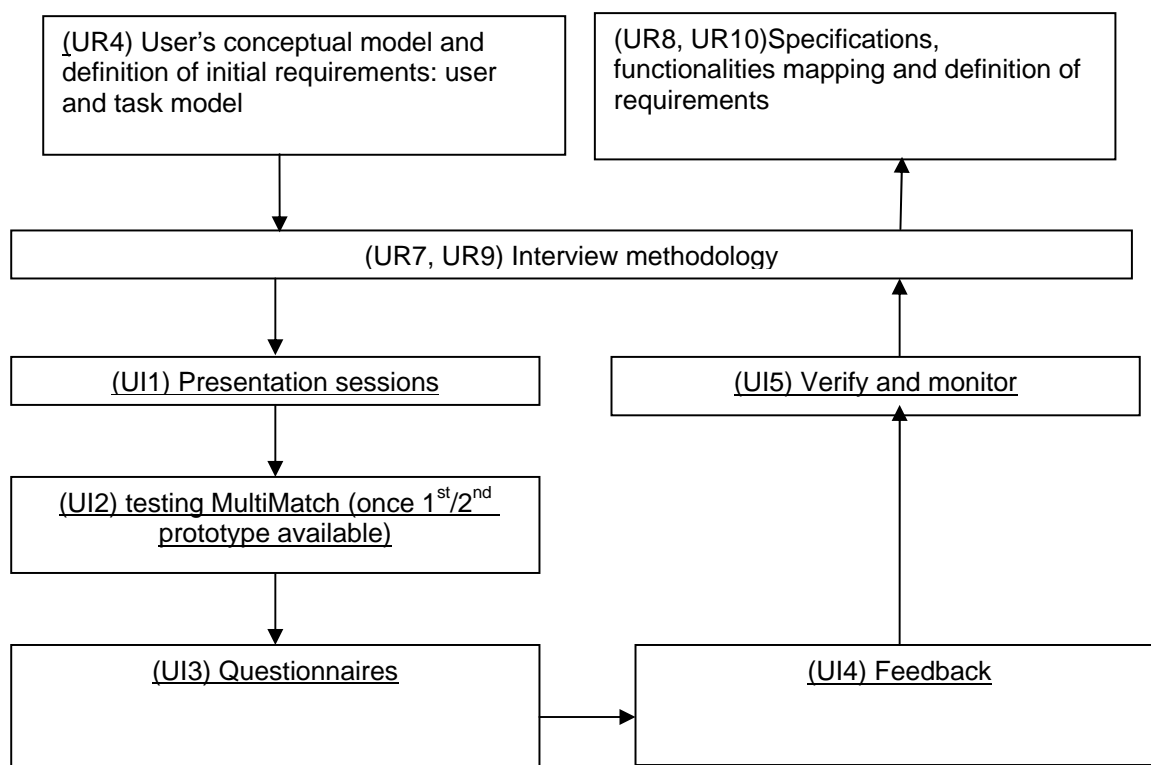


Figure 2: The interview work flow method.

## 4 The Users

### 4.1 Identifying the Users

The identification of users is aimed at defining the main characteristics of the user group that will influence the way the MultiMatch service can be used and the general requirements. We identify three main MultiMatch user groups (educational, cultural tourism, cultural heritage).

Initially we focused on fixed user groups of expert consumers of cultural heritage information in order to define common sets of approaches for carrying out the user requirement studies and system evaluation. We intend to extend the investigation to include non-professional users in a second step mainly via online questionnaires but also by organising hands-on sessions with the first prototype at relevant events such as the annual workshop of the Cross-Language Evaluation Forum (CLEF).

As soon as the first MultiMatch prototype has been released, we will enlarge the membership of the user groups in order to validate a full set of user needs for the system. Workshops, newsletters and events will be used as instruments to get in touch with the user groups and interview them. UNED, BandG, and Alinari will cover one user group each, consisting, respectively, of: educational, cultural tourism and cultural heritage professionals. The user groups include institutes that are affiliated to larger networks. For example, Alinari will contact the MINERVA<sup>4</sup> National Representatives Group.

We now define the user role, the user groups and the potential customers:

- The **users** are: individuals using MultiMatch services and products for their working tasks.
- The **user roles** are defined by what a user does to content, it is a role in the content processing chain.
- User **segment groups** are people who represent a demographic segment, a typical job or business in the content processing chain.
- The **customers** are: tour operators, public sector institutions and business organisations that potentially can sign agreements and pay for using MultiMatch service. The customers could also be third parties that will be identified by the exploitation plan Workpackage 8 (advertisers, publishers, sponsors).

While the user roles are generic, the user segment groups are specific for certain business models, cultures, or demographic segments. The following table shows the relation of user groups and user roles. In the MultiMatch technical annex, we aggregated the user segment groups in three larger and simplified groups: educational, cultural tourism, cultural heritage professionals. We introduce the segment grouping as the MultiMatch aggregation is broad and would otherwise require clarifications in the context where it is used.

MultiMatch aggregated target group	User Role	User segment group	Explanation
Educational	Educator	Professor, teacher	Creates courses and teaches students using multilingual e-learning courses and tools
	Learner	Student, researcher	Performs activities to achieve knowledge of some cultural topic using multilingual e-learning facilities and resources
Cultural Tourism	Consumer	Tourist (booking) agencies Publishers Globe trotters Holiday travellers Business travellers	Consumes content, searches and selects the content; for various reasons (cultural tourist)

<sup>4</sup> Minerva is creating a network of Member States' Ministries to harmonise activities regarding digitisation of cultural and scientific content, to create a common European platform, recommendations and guidelines on digitisation, metadata, long-term accessibility and preservation. <http://www.minervaeurope.org/>

		Vendors of mobile services and devices Individuals watching a film, slide show or photos; Readers of books or magazines; Visitors of an exhibition; Web surfers; etc.	
Cultural Heritage	Creator	Photographer; Film maker; Author; Artist;	Creates content by taking photos, shooting film, recording music, writing text, etc
	Composer	Web designer; Advertising agency; Editor of a book; Editor of a magazine; Journalist; Film producer; Author of Learning material; etc.	Produces content by composing it from raw material, such as cutting and composing video material to produce a film, or by using still images to create an advertisement or a web page.
	Manager	Content Manager, Cataloguers that need to retrieve information and annotate contents; Archivist; IPR managers to verify copyright of contents	Manages the content base, archive and catalogue by inserting and deleting content, creating metadata, editing the index, etc.
	Broker	Service provider; Content provider; Web portal provider; etc.	Offers content to clients, matches between client request and available content.

**Table 1: MultiMatch users.**

Target customers of MultiMatch

1. Non-profit organisations (libraries and archives) providing access to digitised materials, such as librarians, providing digital services
2. Professionals (universities and publishing houses) preparing cultural added value objects (i.e. books, travels, cultural service, etc.) on the basis of high level digitised materials and tools of content retrieved by means of MultiMatch.
3. For-profit professional Cultural Heritage content owners

Table 2 summarizes success criteria of the MultiMatch service against user and customer benefits:

	Service usability	Service sustainability
<b>Users</b>	<ul style="list-style-type: none"> <li>▪ High rate of users (eventually registered) with respect to the number of web site visitors</li> <li>▪ High rate of queries with respect to the number of the Cultural Heritage Contents gathered</li> <li>▪ High rate of contents downloaded / accessed with respect to the hits</li> </ul>	
<b>Customers</b>		<ul style="list-style-type: none"> <li>▪ Access of all project participants to the service after the initial deployment phase of the service has begun (see exploitation plan, deliverable D8.2)</li> <li>▪ High rate of agreement with new customers in relation to the number of institutions addressed during the dissemination activities</li> <li>▪ New services made available for further market exploitation</li> </ul>

**Table 2: Success criteria of MultiMatch service**

### **Educational user groups**

The Educational user groups associated with MultiMatch are those related to the Cultural Heritage domain according to the roles defined above:

1. **Educator Role:** This role is mainly performed by those users who teach and create specific CH courses using Internet and CH digital repositories. In our particular case, we will consider university professors and high school teachers as members of this group. Course teaching and designing implies browsing and navigation through CH digital repositories and web pages to extract and to summarize the information to be taught and compiled. This groups also faces multilingualism issues.
2. **Learner Role:** Considers people related with CH who performs activities oriented to reach higher knowledge of some cultural topic using multilingual e-learning facilities and digital resources. This role is basically performed by University and High School students who use CH digital courses to improve their knowledge about specific topics. The role is also performed by CH researches (i.e. working for Universities or private institutions) who use CH digital resources to produce any kind of material oriented to improve the general knowledge in the CH area (i.e. research papers, Ph D. Thesis, journal articles, conferences).

### **Cultural tourism user groups**

- Tourist (booking) agencies
  - Companies selling package holidays and looking for cultural information (sights, events, costs) on various designations
  - The information requests differ from low-level through high-level
- Publishers
  - Publishers of travel companions, (i.e. Let's Go!, Lonely Planet), Maps, (i.e. Falke) and editors of tourism oriented websites (i.e. [www.virtualtourist.com](http://www.virtualtourist.com), [www.expedia.com](http://www.expedia.com)).
  - The information requests differ from low-level through high-level
  - The tourism oriented websites are sometimes keen to embed services of external services in their environment.
- Globe trotters
  - Individuals travelling often and to distant places, with a keen interest in learning about the culture
  - The information requests differ from low-level through high-level
- Holiday tourists
  - Families, individuals travelling for relaxation and with limited time reserved for visiting heritage sites, museums etc.
  - The information requests are expected to be low level and focus on main events/sights
- Business travellers
  - Individuals who travel for work and might occasionally have spare time to do sightseeing
  - The information requires are expected to be low level and focus on main events/sights
- Vendors of mobile services and devices
  - For example navigation tools such as TomTom, MapQuest Personal Navigation Device
  - They might be interested in offering the information in MultiMatch as one on the information layers to be linked to navigation interfaces used in cars.

### **CH Professional User groups**

Description of the user groups that could be involved in the validation process.

Professional users of CH system



- Cataloguers
  - This group of users needs the access to CH information to annotate, catalogue and gather information with respect to an image to be catalogued. Collections are generally not provided with references on who is represented in the pictures, etc.
  - The information to be generated goes from low level (i.e. clothing, places, objects, etc.) to high level (who are the people represented, which is the event represented, etc.)
- Educational Area
  - The educational group of CH users, in the archives and library domain are responsible for collecting content and creating new educational products (web sites dedicated to history, fine arts, science, etc.)
  - The educational professional users needs are the collection of consistent contents, high quality and validated data, historical references, etc.
- Content validation
  - The content validation process is managed by field experts (fine arts experts, history experts, etc.)
  - The validation process is needed to disambiguate and contextualize the contents.
- Metadata manager (thesauri, multilinguality, etc.)
  - Cataloguing can be done once a metadata management architecture exists. Until few years ago, the most effective architecture for metadata managing, was the structured thesauri (Alinari has created its own on the basis of the guidelines of the University of Florence and of the national Cataloguing Centre ICCD).
  - A flexible, interoperable and scalable management metadata system is the most important part for a digital cultural archive that owns some millions of documents (images, text, videos, etc.) and manages the data of other archives.
- IPR manager
  - The IPR professional users define the property right management procedures with respect to the norms of the countries of the users where the content will be used.
  - The most important need is the provision of licensing agreements on content usage and the verification of the application of the contractual agreements signed.
- Exhibitor
  - Exhibitors use search engines to complement their research activity, to put together coherent and related information, to maximize the exhaustiveness of content for the exhibition they are going to organize.
  - Some exhibitors need also to retrieve from external sources (archives, museums, etc.) content that is not locally available.
- Marketing manager
  - Marketing is not usually visible in the cultural heritage context but the role of the marketing user group is crucial: the presentation of the results, the added value and the cost process are cardinal aspects for successful exploitation and usage.



## 4.2 Characteristics of Expert Users

This section will attempt to define which questions are most relevant for the interviewing process.

The interviews in the initial phase have been addressed to field experts and were executed directly (by direct personal mail, face-to-face, by phone).

As already stated, three different groups of expert users were considered:

1. Educational (e.g. teachers, researchers and course designers).
2. Cultural tourism (e.g. booking agencies, globe trotters, business travellers, holiday tourists)
3. CH professionals (e.g. content managers, researchers, tools developers, catalogue experts)

UNED, BandG, and Alinari performed the interviews.

MultiMatch User Group	Elements considered	Why (MultiMatch relevance)	Reference in questionnaire
<ul style="list-style-type: none"> <li>▪ <u>EDU</u>: Teachers, researchers or course designers</li> <li>▪ <u>Cultural tourism</u>: booking agencies, globe trotters, business travellers, holiday tourists</li> <li>▪ <u>CH</u>: Cataloguers; Research Area User; Educational Area User; Metadata manager (thesauri, multilinguality); IPR manager</li> </ul>	Profile	User profile: age, primary interests, use of the Web and of cultural information. A range of ages is needed to identify whether there are significant differences with respect to ICT background.	E1
	Knowledge and skills	Level of knowledge can influence speed of access to the system. Multilingual and aggregation of contents, summarization and interest level of expertise.	E1
	Geographical distribution	Geographical distribution directly affects cultural aspects we want to consider	E1
	<ul style="list-style-type: none"> <li>- Primary search sources</li> <li>- Useful search services</li> <li>- Common types of text search</li> <li>- Automation &amp; classification features</li> <li>- Aggregation features</li> <li>- Interest focussed on CH contents</li> </ul>	Identification of primary searching sources (general web search engines, library catalogues, vertical search engines, official directories, news). Identification of search activities performed (by author, date, location). Identification of the classification and presentation of content that is most requested. Identification of type of contents and media.	E2, E3, E4
	Search facilities	Differentiating the search facilities by media: relevance feedback, spell check, search suggestions, filtering, use of authority lists, use of Boolean operators.	E5
	Expert comments	To collect unforeseen aspects and scenarios or requirements.	E6

**Table 3: Expert users and relevance for MultiMatch questionnaire**

## 5 User Scenarios

This section defines the user scenarios within the target user groups. The theory and motivation behind the creation of scenarios is explained in Section 3.2.3. Scenarios make it possible for users to envisage the potential of the system, and thus identify and formulate requirements that are not based only on current practice.

On the basis of our past experience, we created context scenarios<sup>5</sup>. The scenarios cover the three user groups we are targeting and have been used to communicate the MultiMatch goals and proposed functionality with respect to these user groups. They also help to clarify within the project the most desirable system features and design options. A next step in the process will be a technical validation of the requirements expressed by these scenarios by workpackages: WP3(System Architecture Design and System Integration), WP4 (Multilingual/Multimedia Indexing and Information Extraction), WP5(Multilingual/Multimedia Information Retrieval), and WP6 (User Interaction and Interface). The initial collection of user requirements will finally be aggregated, analysed and validated by experts and also with users.

*How are the interviews and scenarios applied to generate the user requirements?* The scenarios are stories that come from an imaginary world where MultiMatch exists. The subsequent interviews support some aspects that the scenarios evidence.

A **context scenario** tells the story of a user in a work of personal context, achieving a goal supported by the system. The story is told in terms of

- *Actors and roles, including system roles*
- *Tasks and goals*
- *What triggers a task, when is a task done*
- *Input, output of the task*
- *Resources, tools used*
- *Workflow, business*

Not covered are details of the interaction between user and system, or data input/output, rules etc. Such details are the focus of scenarios of use.

### 5.1 Scope of the Scenarios

We have collected the following scenarios covering the different target groups. Each scenario provides a list of user needs from which we will select those that best match the project vision and goals.

#### List of scenarios

##### Scenario 1: Martha – Giacometti and statue

Reference	Description
Scenario	During a social dinner, in Geneva (July 2006), Sam was looking at Swedish banknotes with Martha: one of them portrayed the Italian sculptor Giacometti. Sam remembered a picture of Giacometti and his sculptures where the artist resembled his works of art. He proposed to Martha that they try to find this picture of Giacometti where the artist is shown together with one of his statues; the picture is famous because although the artist is walking he also appears as a statue. The search was to be done using a mobile device and Google. They entered the key words in Italian: <i>Giacometti mentre cammina e statua</i> and the

<sup>5</sup> Alinari: experiences learnt from: aceMedia project (Barbara S. Beltz).

	system retrieved the exact image that they were looking for which had been tagged with German metadata and was stored on a German website together with a brief summary in Italian of the associated German text.
User needs:	<ul style="list-style-type: none"> <li>• Retrieval capabilities in other languages than the user's mother language and cross-language summarization</li> <li>• Retrieval of material by subject;</li> <li>• Support of various formats for preview on Internet, final edition and streaming;</li> </ul>
Content	Images and text
User group	Cultural heritage non-professional
Technical Environment	Mobile device, laptop

### Scenario 2: Juan is looking for flowers

Reference	Description
Scenario	Juan is preparing an essay about the use of flowers in art. To do this he needs to mine the web searching for web pages which talk about authors that use flowers in their artworks and all the artworks representing any kind of flowers. Initially Juan decides to use Google and Yahoo! to perform this task but he quickly realizes that, although these search engines perform a good retrieval on the query "flowers", he has to do extra mining over the search results to filter and classify them. Moreover, he has to extract the critical information (i.e. names and titles) manually from the final selected relevant pages. He is upset because he has to finish the work for next Monday and he thinks that it is going to be impossible. Fortunately, he remembers a comment about MultiMatch made to him by his girlfriend Ruth, so he decides to test the search engine to see if it could help him to accelerate his work. Typing the query "flowers" in MultiMatch he realizes that MultiMatch not only retrieves web pages for him (i.e. like Google or Yahoo! do) but also a complete list of authors and artworks related with flowers and the complete list of cultural heritage sites where these artworks can be visited. He is very happy because he has obtained the complete list of items he was looking for with only one query.
User needs:	<ul style="list-style-type: none"> <li>• General query support based on natural language.</li> <li>• Retrieval from specific and cultural heritage related indexes.</li> <li>• Different types of retrieval (web pages and cultural objects).</li> <li>• Classification.</li> <li>• Information compilation.</li> <li>• Information extraction over web pages to create the cultural objects.</li> </ul>
Content	Images and text
User group	Educational
Technical Environment	Desktop computer

### Scenario 3: Juan discovers MultiMatch preferences

Reference	Description
Scenario	Juan continues exploring MultiMatch trying to exploit all its capabilities to finally improve his final work. Unfortunately, his girlfriend asks him to go out for a walk. He decides to explore MultiMatch to find an option allowing him to store his search results. Clicking a "sign in" option he discovers the possibility to register himself as a user and access a complete history of all his searches. Also, as a registered user he can annotate search results and store his search preferences (query language, retrieval languages, type and size of files to retrieve, etc.). Again, he is very happy because MultiMatch has solved his data storing problems and he can continue with his work later on.
User needs:	<ul style="list-style-type: none"> <li>• Registration and sign in features.</li> <li>• Search user profile customization.</li> <li>• History and annotation options over search results.</li> </ul>
Content	Images and text
User group	Educational
Technical Environment	Desktop computer

### Scenario 4: Juan needs to know a little bit more about Van Gogh Sunflowers

Reference	Description
Scenario	After examining initial search results, Juan decides to improve his knowledge about the artwork "The Sunflowers" from Van Gogh a little bit more. He realizes that it was a previous search result showed by MultiMatch as a cultural object and decides to click on its link to see what happens. MultiMatch launches a new query based on metadata associated with "The sunflowers" cultural object and retrieves specific information about this topic. Juan realizes that MultiMatch has clearly separated and classified web pages according to general categories such as pages about the artwork, reviews of the artwork, news related with the artwork and non-categorized pages. He also can access a profile info box which describes the main features of the artwork. This is done by MultiMatch automatically.
User needs:	<ul style="list-style-type: none"> <li>• Specialized search (i.e. in this scenario author search).</li> <li>• Relevance feedback</li> <li>• Specific categories for each specific type of search.</li> </ul>
Content	Images and text
User group	Educational
Technical Environment	Desktop computer

### Scenario 5: Juan needs to know a little bit more about Van Gogh Sunflowers (Part II)

Reference	Description
Scenario	Juan continues examining the MultiMatch "Sunflowers" results and realizes that it would be a good idea to know about all the artworks most closely associated with "The Sunflowers" in order to study similarities and things in common. He finds a tab called "Artwork Space" which seems to answer all his needs. He clicks on it and he finds an artwork network which relates "The Sunflowers" artwork with other masterpieces according to their main features such as the museum where they are exhibited, art period, painting techniques and so on. Also he finds a set of relevant terms related with the artwork which could be used to drive new queries focused on specific and artwork related topics.
User needs:	<ul style="list-style-type: none"> <li>• Information mining techniques.</li> <li>• Visual representation of relationships between items.</li> <li>• Summary of the artwork based on relevant terms</li> </ul>
Content	Images and text
User group	Educational
Technical Environment	Desktop computer

### Scenario 6: Juan needs to know a little bit more about Van Gogh Sunflowers (Part III)

Reference	Description
Scenario	Juan is very happy with MultiMatch but he wants to complement all his search results with those provided by general search services and probably not covered and indexed by MultiMatch. Using the "more" tab Juan discovers that MultiMatch has the possibility of aggregating search results taken from multiple search sources. For the query "The Sunflowers" MultiMatch retrieves web results not only from general sources such as Google and Yahoo! but also from encyclopaedic resources such as Wikipedia or Answers.com. All the search results are not merged together and are showed to Juan separately in order to clearly identify the source from which they were extracted. As a preference, Juan can select the search sources to be aggregated from a complete list of search engines.
User needs:	<ul style="list-style-type: none"> <li>• Search results aggregation.</li> </ul>
Content	Images and text
User group	Educational
Technical Environment	Desktop computer

### Scenario 7: Peter is preparing his visit to Madrid

Reference	Description
Scenario	Peter is organizing a tourist trip to Madrid. He wants to spend all his time in Madrid visiting museums and cultural sites. He knows about the Prado Museum, which will be his first visit, but he also wants to visit other cultural sites. He uses MultiMatch, first asking its specialized search about "Prado". MultiMatch retrieves a complete compilation of information related with authors and artworks that are exhibited there and also a set of web pages categorized as reviews of the site, news pages and unclassified ones. Peter can explore all this information to know more about the museum contents and also can refine his search asking about any of the specific cultural objects retrieved. Finally, he can use the "Site Space" tab to find other museums closely associated to Prado, in the same city, covering the same art periods or the same artists.
User needs:	<ul style="list-style-type: none"> <li>• Specialized search on cultural sites.</li> <li>• Information compilation.</li> <li>• Information mining techniques.</li> <li>• Visual representation of the relationships between items.</li> </ul>
Content	Images and text
User group	Tourism
Technical Environment	Desktop computer

### Scenario 8: Giovanna publishes her book

Reference	Description
Scenario	<p>Giovanna is a lecturer at Florence University. She is collecting images for a book on Italian art. She has made iconographic searches in many Stock Images and needs to produce a dummy with images and text to show the publisher.</p> <p>Giovanna searches for relevant images and texts on the MultiMatch site. MultiMatch can:</p> <ul style="list-style-type: none"> <li>• Associate images with related text</li> <li>• Order images and text with respect to relevance</li> <li>• Provide annotations relative to the digital quality of the images, colour and orientation</li> <li>• Give information on IPR (copyright of images, copyright author, restrictions) and the relative reproduction fees</li> </ul> <p>Giovanna prints the dummy produced with MultiMatch and shows it to the publisher</p>
User needs	<ul style="list-style-type: none"> <li>• Retrieve material by subject</li> <li>• Associate images with related text</li> <li>• Order results according to relevance</li> <li>• Provide information on image features</li> <li>• Provide information on IPR and prices</li> </ul> <p>support various formats for preview on Internet, final edition and streaming;</p>
Content	Images and text
User group	Educational professional
Technical Environment	A PC with internet connection

### Scenario 9: Oliver studies movies

Reference	Description
Scenario	<p>Oliver is a student studying nineteen century film productions. He is going to write an essay about the movies by film director Roman Polanski. He has the entire collection of Roman Polanski 's movies with the exception of one he was told about without knowing the title. He does not know Polish (he is a native German speaker) and the movie is in Polish only.</p> <p>A friend suggests he searches for the movie using the MultiMatch search engine where the results are limited to the few top ranked and he has the possibility of improving the search by accepting the system prompts ('perhaps you are looking for...') and other advanced functionality (multilingual search, context analysis, biographies etc.).</p> <p>Oliver queries the system in German and finds the movie he is looking for via the English title (Dwaj Ludzie z Szafa / Two Men and a Wardrobe). He saves both the queries he entered and the results in a dedicated folder. MultiMatch adds the information that he likes movies by Polanski to his user profile. MultiMatch also presents Oliver with a system generated page containing a biography of Polanski, some articles on his best known works, the complete list of the movies he has made, film festivals where he has won awards, his favourite actors, etc., all in a well formatted layout. Now he has enough material for his essay.</p>
User needs	<ul style="list-style-type: none"> <li>• Limitation of the number of results</li> <li>• Language summarizations from source to target language</li> <li>• System prompts ('perhaps you are looking for...')</li> <li>• Advanced functionality (cross-language search, context analysis, biography etc.).</li> <li>• Related information on the topic searched – created automatically by the system</li> <li>• Ability to save queries and results</li> <li>• Retrieval of material by subject;</li> <li>• Support of various formats for preview on Internet, final edition and streaming;</li> </ul>
Content	Video
User group	Educational, entertainment non-professional
Technical Environment	Desktop computer

### Scenario 10: Daniela is preparing an exhibition

Reference	Description
Scenario	<p>Daniela, who is preparing an exhibition, needs a list of low resolution files on Jean Michel Basquiat's works of art. She searches for them using MultiMatch: the results should be classified by country and by place; and must include only events between 1971 and 1984;</p> <p>She also needs descriptions of each image in two different languages as she is preparing the exhibition in Italy and will organize it at the Spanish embassy there. She wants to know the technical costs, the right-of-use costs and the mode of</p>



	<p>payment to use the MultiMatch page as a brochure to be distributed at the event. Some paintings will be sold during the exhibition and Daniela has the possibility to input her comments and costs on the page generated by the MultiMatch interface.</p> <p>Daniela found an interesting documentary and video interview with Madonna, the pop singer, who had been Basquiat's girlfriend for three months on MultiMatch. She would like to show it during her presentation and contacts the author. Finally, she notices that one of Basquiat's masterpieces is missing from the list. She knows that the painting reproduced the Mona Lisa in some way. So she retrieves a picture of the Mona Lisa and by similarity search retrieves Basquiat's Mona Lisa 1983. MultiMatch also provides enough information for her to contact the owner of the painting so that she can request permission to exhibit it..</p> <p>Daniela finally wonders if she can advertise the exhibition in some way on MultiMatch to her address list and to a wider group of users. MultiMatch allows her to share a page that she has designed with the exhibition brochure (this page will be indexed for those who will search events in the fixed event period).</p>
User needs	<ul style="list-style-type: none"> <li>• A list of low resolution files on an artist works of art.</li> <li>• The list should be classified by country and by place;</li> <li>• The list must include only events between a selected date range (time line) and countries (country selection)</li> <li>• Descriptions of each image in two different languages</li> <li>• Technical costs,</li> <li>• Right-of-use costs and the mode of payment</li> <li>• To print aggregate results in brochure form</li> <li>• Possibility to annotate the results page generated by MultiMatch</li> <li>• Search by visual similarity.</li> <li>• Results saved in personal folder.</li> <li>• Advertisement and dissemination functionality</li> <li>• Retrieval by subject;</li> <li>• Retrieval and preview via Internet, import of video material in editing software; retrieval of related topics by text metadata.</li> </ul>
Content	Image, text, video, translations, cross-language summarization
User group	Cultural heritage, cultural tourism
Technical Environment	Desktop computer

#### Scenario 11: Rita is a SmallArchives manager

Reference	Description
Scenario	<p>Rita is a catalogue manager for SmallArchives. She has received the latest photo archive to catalogue in the content base. She has received an Excel file with textual information about all the photos and needs to organize the information as metadata for the images using a thesaurus, which later enables retrieval by several attributes.</p> <p>Rita has chosen the Iconographic Thesaurus. The program organizes the text file and the images in the Iconographic Dictionary .</p>

User needs	<ul style="list-style-type: none"> <li>Controlled dictionaries and thesaurus to organize the text information</li> <li>Retrieval of material by subject;</li> <li>Various formats for preview on Internet,</li> <li>Final edition and streaming;</li> <li>Information on IPR</li> </ul>
Content	Thesaurus, metadata, images
User group	Cultural Heritage
Technical Environment	PC connected to the Internet

#### Scenario 12: Leonardo is a content broker at BigSearchImages

Reference	Description
Scenario	<p>Leonardo is a content broker at BigSearchImages Ltd. He helps clients to find special content in BigSearchImages' image base. In the past Leonardo has used a personal set of archives (Getty Hulton, Alinari, Corbis, Bridgeman, etc.) to find the requested contents. He had to search the same contents in many different vertical repositories; this process was very time-consuming. He now uses MultiMatch to collect contents from different authoritative sources from a single search interface. MultiMatch also stores the queries done by Leonardo in the past.</p> <p>One client, who is writing a book, has just asked for some historical photos of every-day work in early industry, which she needs as illustration for her book.</p> <p>Leonardo can retrieve pictures from the MultiMatch content base using a thesaurus. It contains some keywords that match with the client's request. Leonardo can browse the resulting subset of images or further limit it by adding query expressions.</p> <p>Once Leonardo has found an image that fits, he asks MultiMatch to search images that are similar to this example. MultiMatch searches by comparing some graphical features of the example image (colours, layout, objects such as human figures) but also looking for semantically related descriptors in the metadata. After a short time, MultiMatch starts to list thumbnail images that are similar. Some are indeed similar to what Leonardo expects, others are not. He picks another two suitable ones and asks MultiMatch to search using the 3 selected examples. Now the images retrieved by MultiMatch fit closely to what Leonardo is looking for.</p> <p>Leonardo's next client asks for 20 images that show well-known sites (such as Niagara Falls or the New York City skyline) and have a certain atmosphere about them: the client needs cold / calm colours and layout. Leonardo starts his retrieval by thinking of well-known places, browsing photos of these, selecting some that are in landscape format and (to him) radiate calmness. When he has found a few such images showing different places he asks MultiMatch to retrieve images that are visually similar to these examples.</p>
User needs	<ul style="list-style-type: none"> <li>Collect contents from different authoritative sources from a single search interface</li> <li>Ability to store the queries formulated in the past.</li> <li>Ability to browse the resulting subset of images or further limit it</li> <li>Possibility to search for similar images or for similar concepts looking for semantically related descriptors in the metadata.</li> </ul>

	<ul style="list-style-type: none"> <li>• Ability to search by examples.</li> <li>• Ability to search by atmosphere: needs. cold / calm colours and layout.</li> <li>• Browsing support,</li> <li>• Retrieval support,</li> <li>• Ability to search by visual similarity to one or several examples,</li> <li>• Ability to search by concept</li> </ul>
Content	Photographs, any subject
User group	Cultural heritage
Technical Environment	PC

## 6 Interviews

One hundred interviews were performed with expert users and this section will discuss the requirements that emerged from an analysis of the results. In Section 8, we identify which requirements are the most important and should thus be addressed by MultiMatch.

### 6.1 The Questionnaire

As stated previously, the first MultiMatch survey of user needs has considered experts in the three main groups identified (i.e. educational, tourism and cultural heritage professionals). A questionnaire was designed covering the main search features under consideration for implementation in MultiMatch (see Annex II). Here below, we describe each of the features considered and explain why are they important for MultiMatch.

The questionnaire is divided into different parts to address the CH expert users' requirements from different points of view. The aim is to extract information about the user profile, text, image and video search behaviour, and the search facilities, a CH expert user would like to find in a search engine.

The following points describe the intentions of the questionnaire and how they relate to the MultiMatch goals:

- **E1. User Profile.** The main objective is to derive the expert user profile on the basis of his abilities when he is searching and, in general, using the web to access CH information. Specifically, this part of the questionnaire covers:
  - General questions about age and working activities of the respondent (EQs 1.1, 1.2 and 1.3).
  - General questions about internet use (EQs 1.4, 1.5 and 1.6).
  - General questions about generic search engines (EQ 1.7)
  - General questions about language skills (EQs 1.8 and 1.9)

The results of this part of the questionnaire try to answer the following questions:

- What is the personal profile of a MultiMatch expert user?
  - Which services does an expert user require in his daily CH information seeking?
  - Which are the preferred languages for CH information access?
- **E2. Text Search.** The main objective is to understand how an expert user executes a text search. This part of the questionnaire considers which are the users' primary sources when searching, their habits when searching and present two ideal scenarios in which information mining and aggregation are investigated. Specifically, this part of the questionnaire covers:

- General sources for text search in the CH domain (EQs 2.1 and 2.2).
- General types of text search in the CH domain (EQs 2.3 and 2.4)
- Aggregation preferences in the CH domain (EQs 2.5 and 2.6)
- Information mining preferences in the CH domain (EQs 2.7 and 2.8)

The results of the E2 part of questionnaire try to answer the following questions:

- Where do expert users usually perform their CH text searches?
- Which are the most common information needs in CH text searches?
- Is information aggregation useful for CH experts? Which kinds of aggregation are most useful?
- Is information mining useful for CH experts?
- **E3. Image Search.** The goals of this part are similar to those of E2 but related to images, not text. The following aspects are covered:
  - General sources for image search in the CH domain (EQs 3.1 and 3.2).
  - General types of image search in the CH domain (EQ 3.3)

The results of the E3 part of questionnaire try to answer the following questions:

- Where do expert users usually perform their CH image searches?
- Which are the most common information needs in CH image searches?
- **E4. Video Search.** Similar to text and image search but focused on video search. Specifically, this part of the questionnaire covers:
  - General sources for video search in the CH domain (EQs 4.1 and 4.2).
  - General types of video search in the CH domain (EQ 4.3)

The results of part E4 try to answer the following questions:

- Where do expert users usually perform their CH video searches?
- Which are the most common information needs in CH video searches?
- **E5. Desirable Search Facilities.** This part of the questionnaire investigates the search facilities preferred by the expert user when executing a search. In E5, search facilities are divided in three main groups basically oriented to querying, browsing and results displaying and, finally, results processing. Specifically, E5 covers the following aspects:
  - Desirable search facilities at query time considering general searches (i.e. text, image and video) (EA 5.1.1 to 5.1.11)
  - Desirable search facilities at query time considering image and video searches (EA 5.1.12 to 5.1.15)
  - Desirable search facilities when obtaining and displaying results considering general searches (i.e. text, image and video) (EA 5.1.16 to EA 5.1.26)
  - Desirable search facilities when obtaining and displaying results considering image and video searches (i.e. text, image and video) (EA 5.1.27 to EA 5.1.32)
  - Desirable search facilities when handling results considering general searches (i.e. text, image and video) (EA 5.1.33 to EA 5.1.36).

The results of part E5 try to answer the following questions:

- Which are the expert user's preferred facilities when querying a search system?
- Which are the expert user's preferred facilities when browsing and inspecting search results?
- Which are the expert user's preferred facilities when handling and managing the search results?
- **E6. Your Comments.** This part aims at retrieving any requirements not otherwise covered by the questionnaire.

## 6.2 Preliminary Results

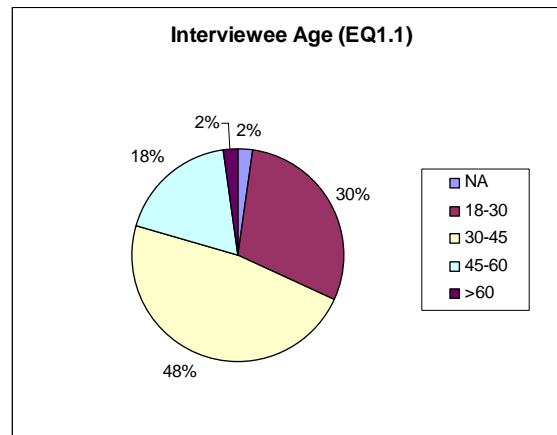
In this section we present the preliminary results obtained from the interviews. The first part of the section discusses the results of the Educational group, the second will study the results of the Cultural Heritage group and, finally, in the third part, results of the Tourism group are presented.

### Educational Group Results (UNED)

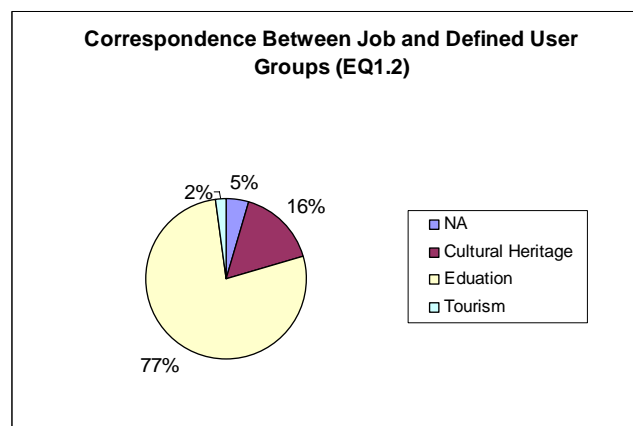
MultiMatch partners interviewed 44 different experts from Cultural Heritage Educational areas. The results obtained are the following.

#### E1. User Profile

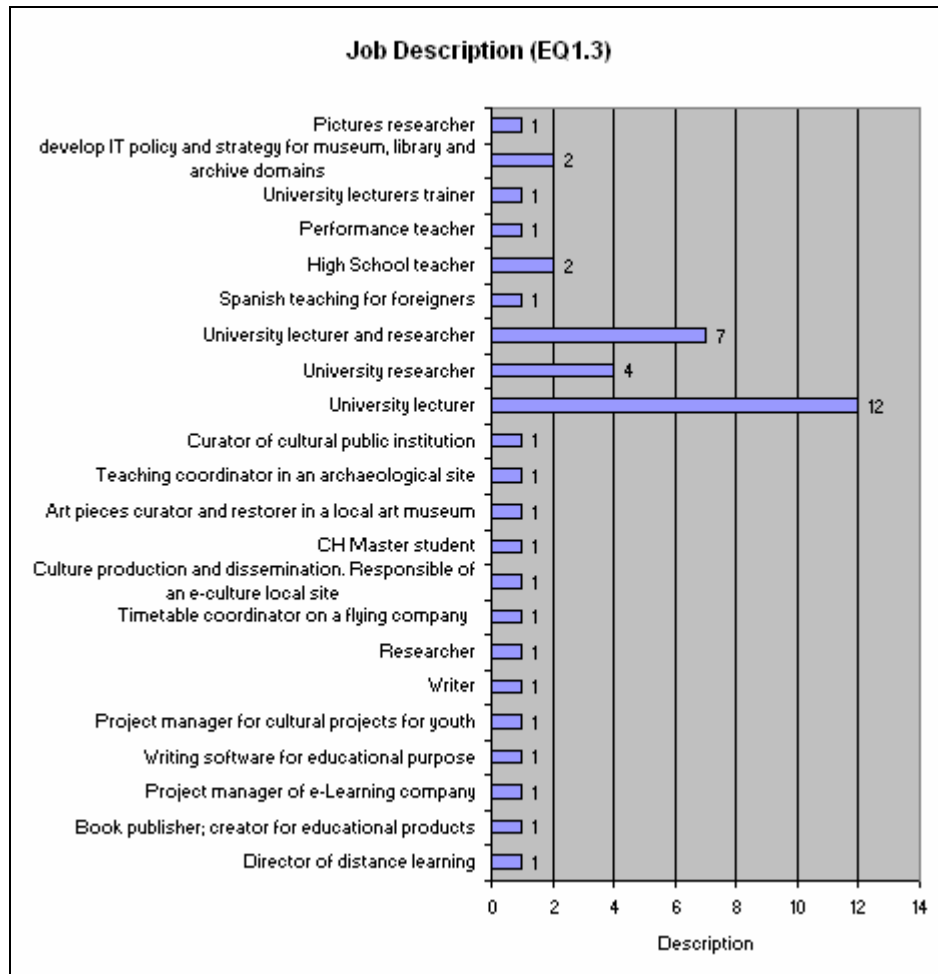
Question	Comments	Figure
EQ 1.1	The most common age range of the expert respondents is between 30-45 (48%) closely followed by 18-30 (30%)	3
EQ 1.2, 1.3	Most of the educational experts associated their job with the educational area (77%) while 21% associated their work with cultural heritage or tourism areas. This is because the educational group includes: <ul style="list-style-type: none"> <li>• Students of areas related with CH that can be also working on CH or for tourist agencies.</li> <li>• Experts that perform their educational activities mainly in CH institutions (e.g. museums, art galleries) to promote and disseminate their contents.</li> </ul> A snapshot of specific jobs can be seen in Figure 5.	4,5
EQ 1.4	95% of experts considered essential or useful (59% vs 36%) the WWW in their daily work	6
EQ 1.5	Most popular sources of cultural information (considered essential or useful) were the following: <ol style="list-style-type: none"> <li>1. Regular visits to specialized web sites (<math>\approx 90\%</math>)</li> <li>2. General search services (<math>\approx 90\%</math>)</li> <li>3. Specialized search services (<math>\approx 70\%</math>)</li> <li>4. Use of databases and catalogs (<math>\approx 70\%</math>)</li> <li>5. Use of links suggested by colleagues (<math>\approx 60\%</math>)</li> </ol> While RSS feeds from specialized sites and newsletters subscriptions were not as useful ( $<40\%$ )	7
EQ 1.6	Cultural information search frequency was as follows: <ul style="list-style-type: none"> <li>• Cultural text searches are the most common, mainly performed daily or weekly by experts (<math>\approx 90\%</math>)</li> <li>• Image and video searches are performed occasionally video searches being less frequent than image searches (30% vs 50% for daily and weekly searches respectively)</li> </ul>	8
EQ 1.7	Preferred search engines were the following: <ul style="list-style-type: none"> <li>• For text search: Google, Yahoo, Vivissimo and library catalogs and centralized DBs</li> <li>• For image search: Google, Yahoo and Corbis</li> <li>• For video search: Google, Youtube, Yahoo and Limewire</li> </ul>	9,10,11
EQ 1.8, 1.9	Although most experts interviewed were native Spanish speakers, results of question 1.9 show that English is the most useful language to get updated and comprehensive cultural information from web sources.	12,13



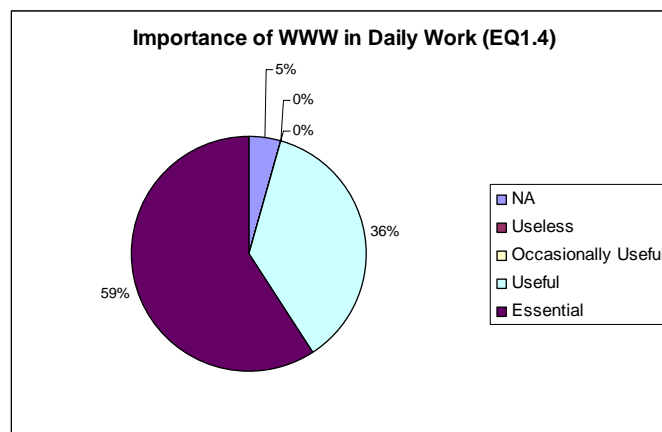
**Figure 3: respondents age distribution (Educational group)**



**Figure 4: respondents' job distribution (educational group)**

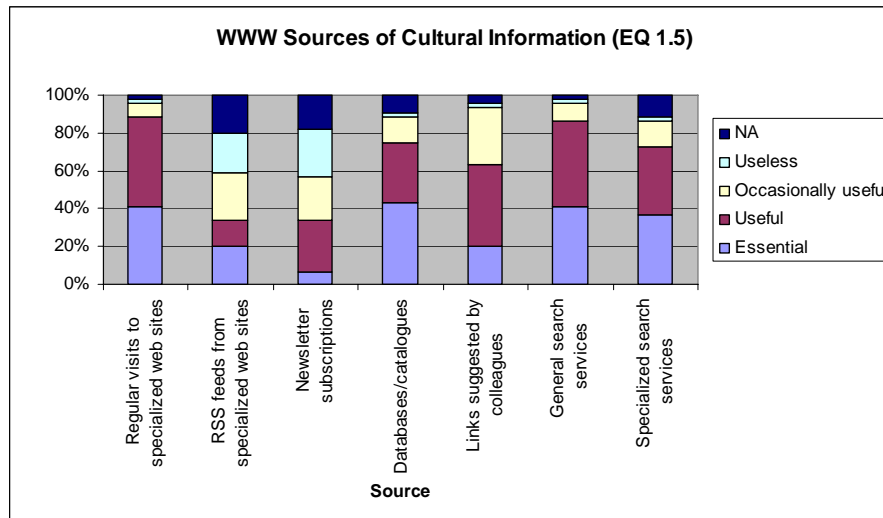


**Figure 5: respondents' job description (educational group)**

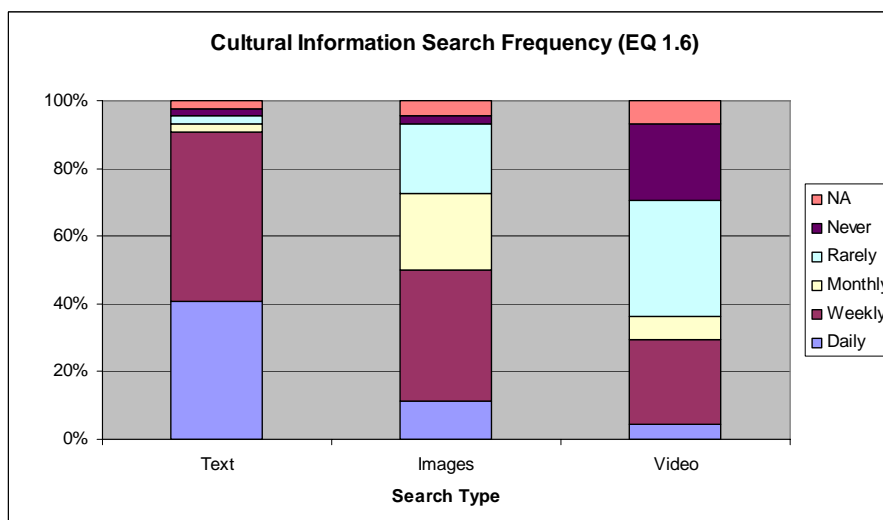




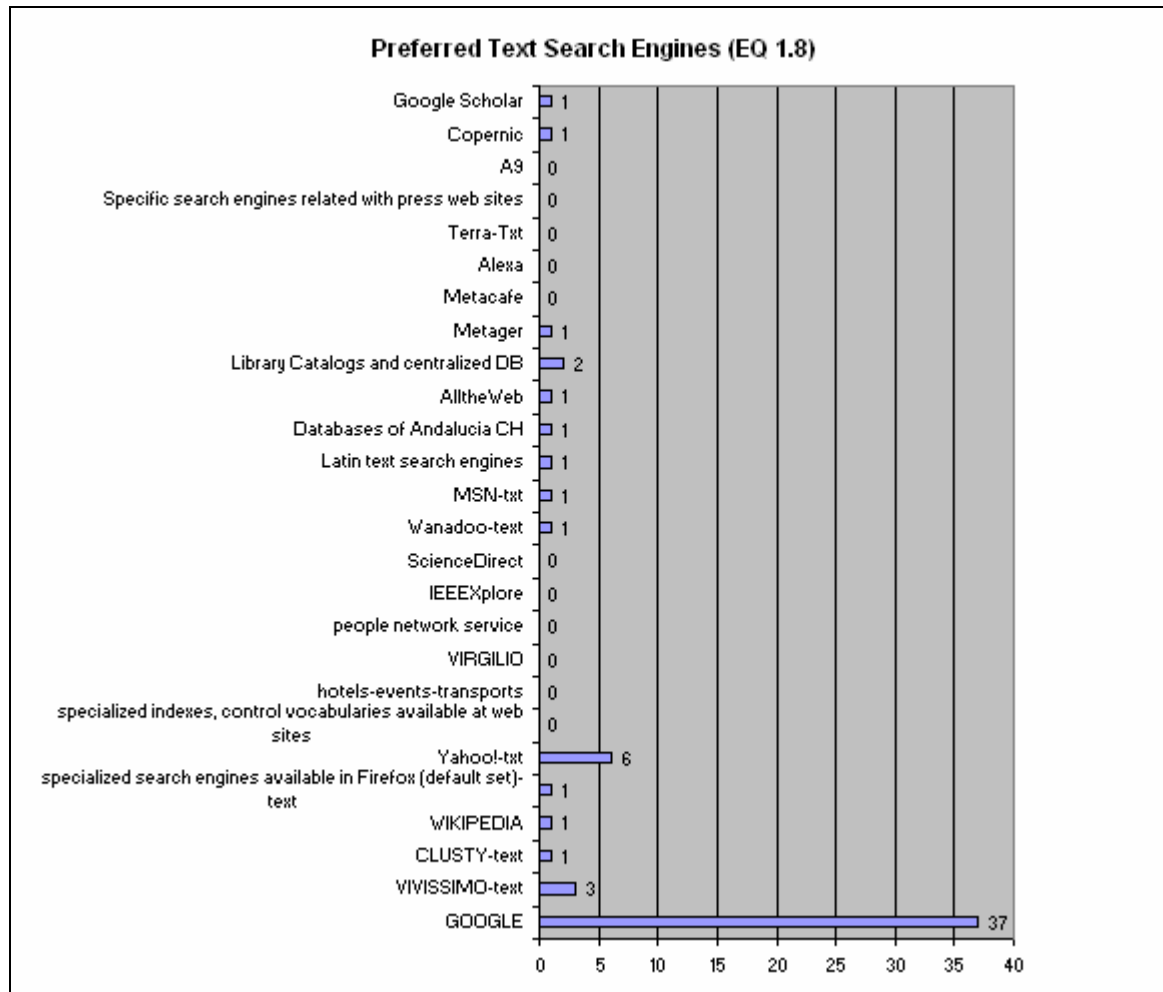
**Figure 6: Importance of the WWW in daily work (educational group)**



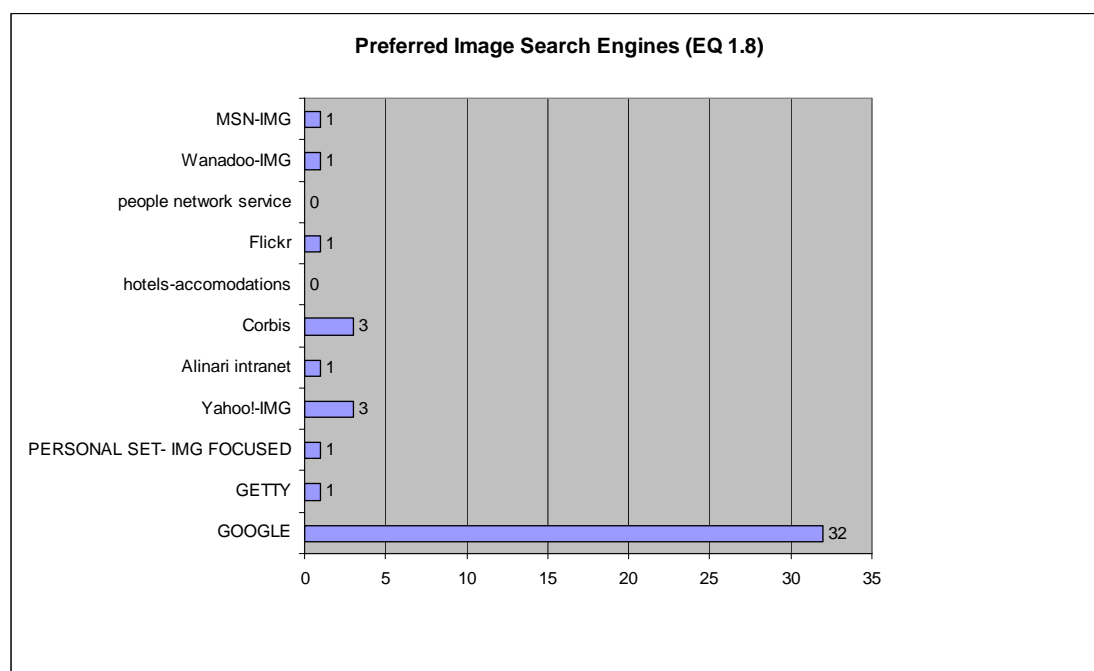
**Figure 7: Main sources of cultural information (educational group)**



**Figure 8: Text, images and video search frequencies (educational group)**



**Figure 9: Preferred text search engines (educational group)**



**Figure 10: Preferred image search engines (educational group)**

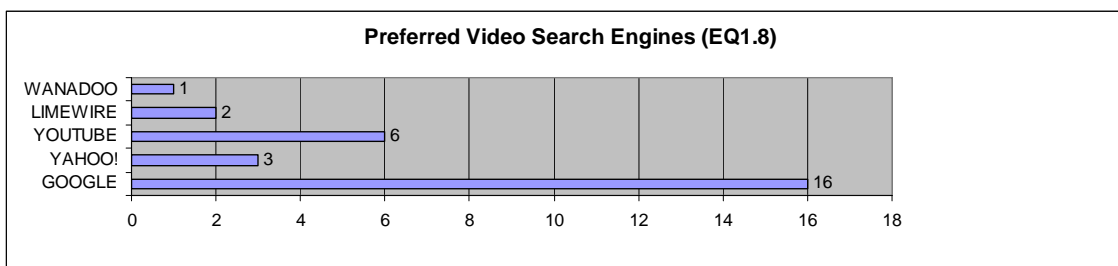


Figure 11: Preferred video search engines (educational group)

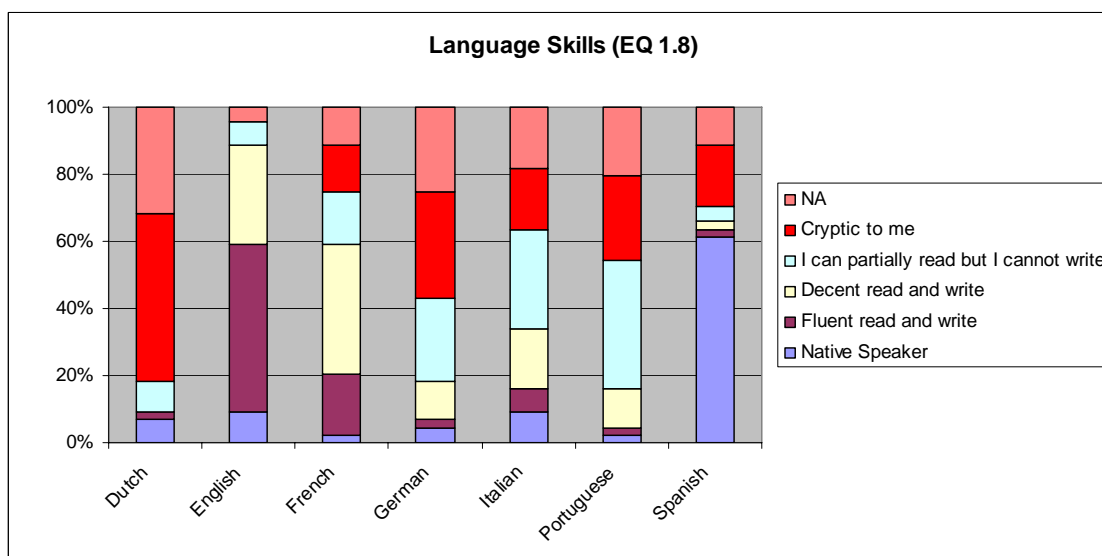


Figure 12: Experts language skills (educational group)

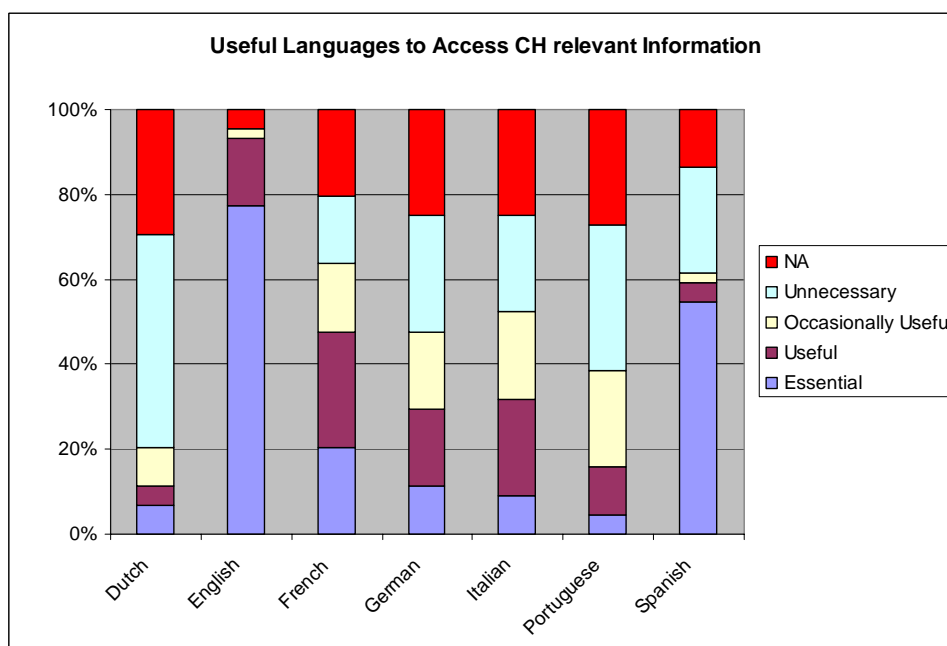
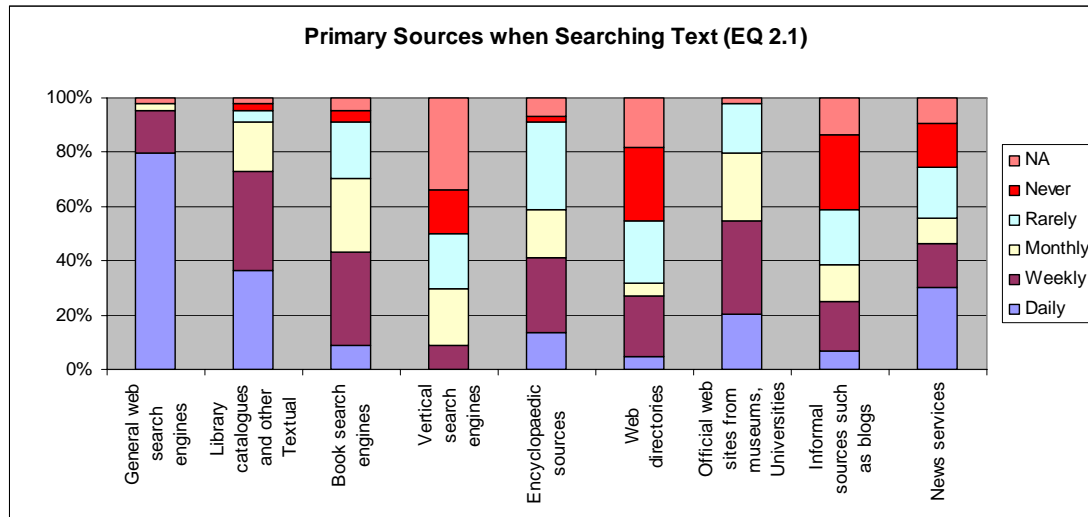


Figure 13: Considerations about which languages are more useful for CH access (educational group)

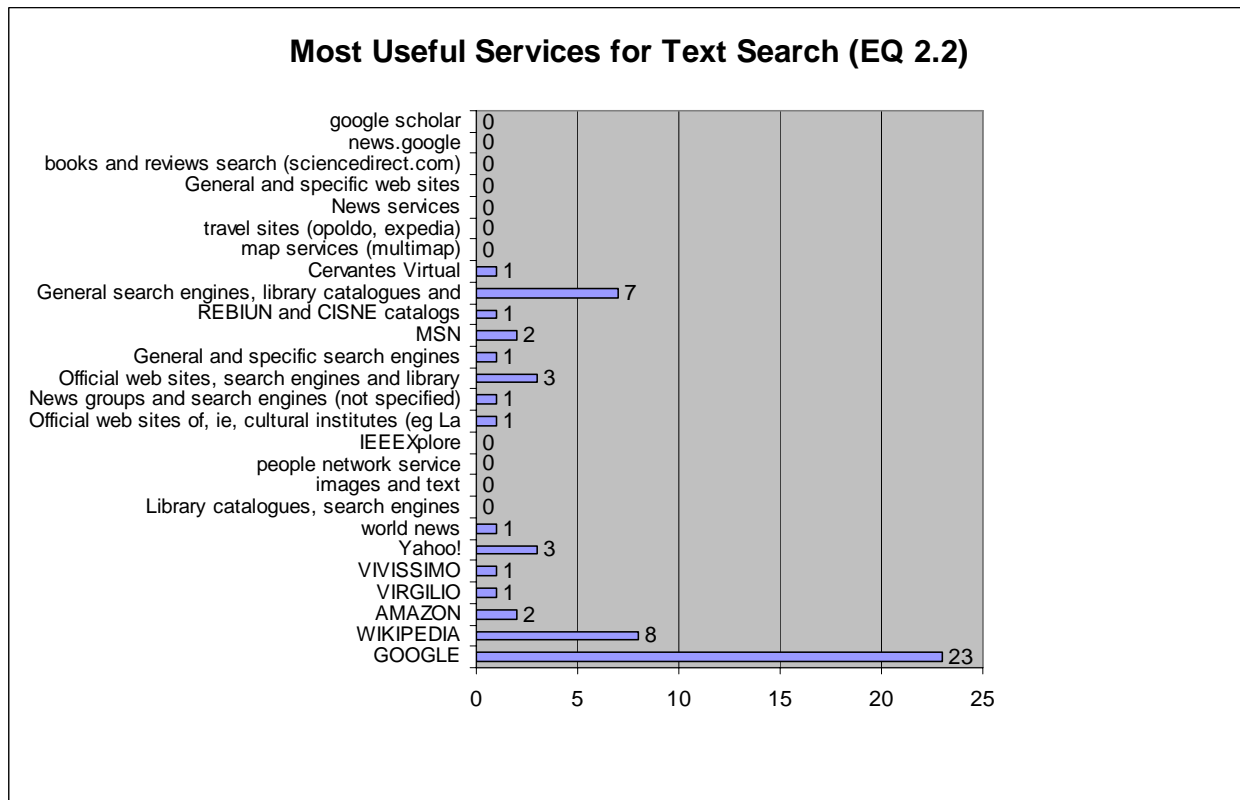
## E2. Textual Search

Question	Comments	Figure
<b>EQ 2.1</b>	<p>The main sources when searching text cultural information were the following (for daily, weekly or monthly use):</p> <ol style="list-style-type: none"> <li>1. General web search engines (<math>\approx 100\%</math>)</li> <li>2. Library catalogs and other textual databases (<math>\approx 90\%</math>)</li> <li>3. Official web sites of museums or universities (<math>\approx 80\%</math>)</li> <li>4. Book search engines (<math>\approx 70\%</math>)</li> <li>5. Encyclopaedic sources (<math>\approx 60\%</math>)</li> <li>6. News services (<math>\approx 50\%</math>)</li> </ol> <p>Other sources such as vertical search engines, web directories and blogs were not considered as useful (<math>&lt;40\%</math> of daily or weekly use)</p>	<b>14</b>
<b>EQ 2.2</b>	<p>The most useful services were:</p> <ol style="list-style-type: none"> <li>1. General search engines (such as Google, Yahoo! or MSN)</li> <li>2. Wikipedia</li> <li>3. Library catalogues and textual databases</li> <li>4. Official web sites</li> </ol>	<b>15</b>
<b>EQ 2.3, 2.4</b>	<p>All the text searches proposed in the questionnaire were accepted as useful in daily, weekly or monthly work. However, results can be ranked accordingly the following importance list:</p> <ol style="list-style-type: none"> <li>1. Search for specific authors (<math>\approx 90\%</math>)</li> <li>2. Search for specific web sites (<math>\approx 90\%</math>)</li> <li>3. Search for books (<math>\approx 80\%</math>)</li> <li>4. Search for scholar publications (<math>\approx 70\%</math>)</li> <li>5. Search for cultural events (<math>\approx 70\%</math>)</li> <li>6. Search for news (<math>\approx 60\%</math>)</li> <li>7. Search for an art period (<math>\approx 50\%</math>)</li> </ol> <p>Text searches proposed by experts were the following:</p> <ul style="list-style-type: none"> <li>• Thematic searches</li> <li>• Search for quotes</li> <li>• Search in databases for specific words or texts related with one or more classic authors</li> </ul>	<b>16</b>
<b>EQ 2.5, 2.6</b>	<p>Regarding information compilation and aggregation expert users considered essential or normally useful the following:</p> <ol style="list-style-type: none"> <li>1. Compiled information about authors (<math>\approx 70\%</math>)</li> <li>2. Compiled information about artistic items (<math>\approx 70\%</math>)</li> <li>3. Compiled information about cultural sites (<math>\approx 65\%</math>)</li> </ol> <p>The other options presented were also considered normally useful but the percentage was always under 50%.</p> <p>Educational group experts didn't make suggestions about other useful kind of compilations.</p>	<b>17</b>
<b>EQ 2.7, 2.8</b>	<p>Regarding information extraction and pattern discovering, expert users considered essential or normally useful the following relations:</p> <ol style="list-style-type: none"> <li>1. Ideas in common between two or more authors (<math>\approx 90\%</math>)</li> <li>2. Authors most closely associated with a given author (<math>\approx 80\%</math>)</li> <li>3. Masterpieces most closely associated with an author (<math>\approx 70\%</math>)</li> <li>4. Places/people/topics most closely associated with an artist (<math>\approx 70\%</math>)</li> <li>5. Artists/masterpieces/CH sites most closely related to a given keyword (<math>\approx 70\%</math>)</li> <li>6. Artists/masterpieces/cultural sites most closely associated with a city</li> </ol>	<b>18</b>

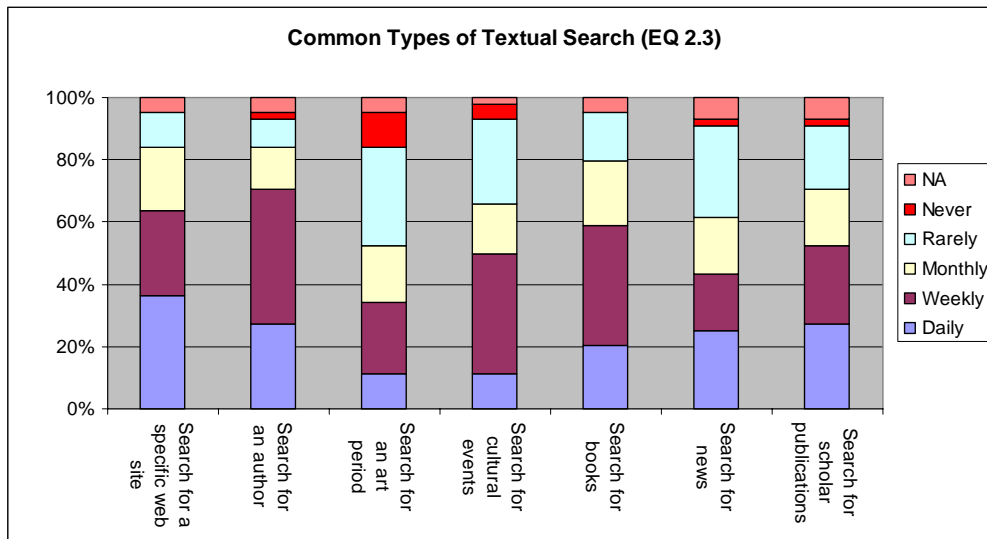
	<p>(≈60%)</p> <p>Options related with authors popularity and opinions were not considered as useful (&lt;50%)</p> <p>Educational group experts didn't make suggestions about other useful kind of compilations.</p>	
--	---	--



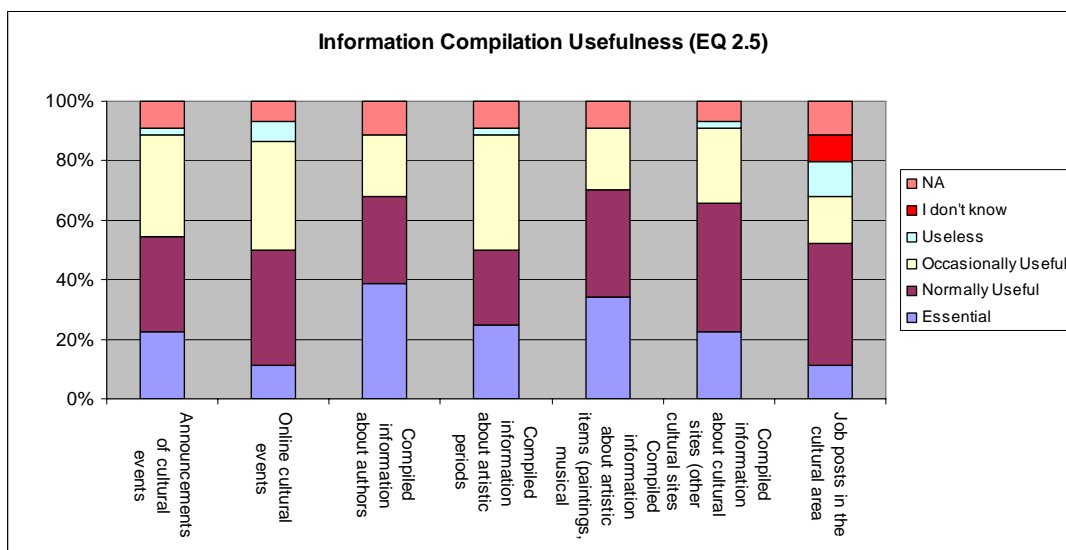
**Figure 14: Primary sources for textual cultural searches (educational group)**



**Figure 15: Most useful services for textual cultural searches (educational group)**

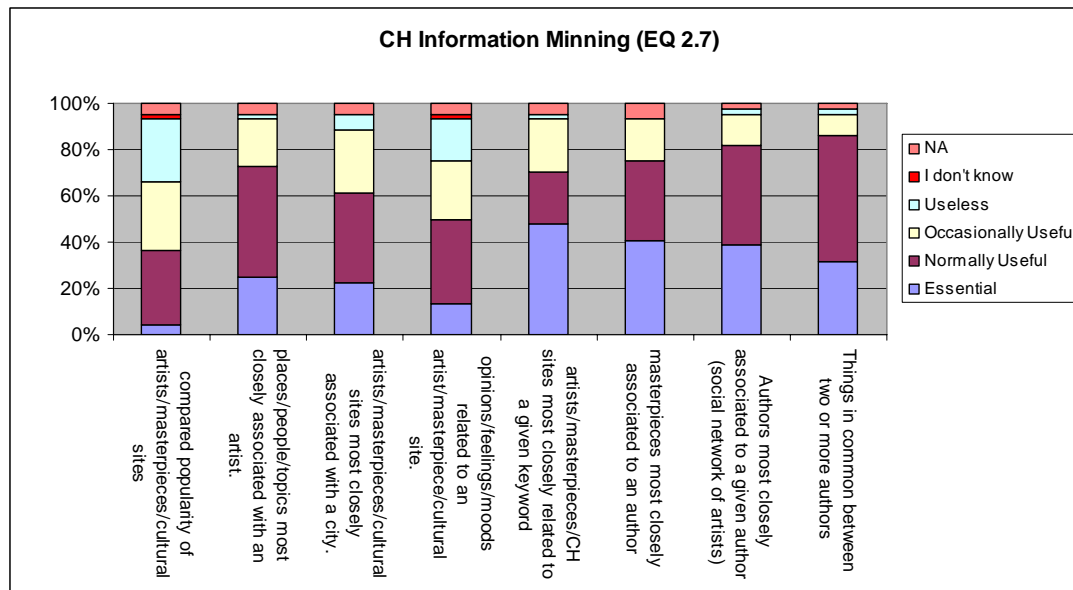


**Figure 16: Common types of textual search (educational group)**



**Figure 17: CH Information compilation usefulness (educational group)**





**Figure 18: CH information mining usefulness (educational group)**

### E3. Image Search

Question	Comments	Figure
<b>EQ 3.1</b>	<p>The primary sources stated by expert users when searching for images were the following (for daily, weekly and monthly use):</p> <ol style="list-style-type: none"> <li>1. General image search engines (<math>\approx 90\%</math>)</li> <li>2. Cultural web sites (<math>\approx 60\%</math>)</li> </ol> <p>Proprietary databases, classified web directories and collective web repositories are also used but with a lower frequency</p>	<b>19</b>
<b>EQ 3.2</b>	<p>The most useful services stated by expert users were mainly focused in the use of general image search engines such as Google</p>	<b>20</b>
<b>EQ 3.3, 3.4</b>	<p>The most common types of image searches were the following:</p> <ol style="list-style-type: none"> <li>1. Search for illustrative images (<math>\approx 70\%</math>)</li> <li>2. Search for photographs of artworks (<math>\approx 65\%</math>)</li> <li>3. Search for artistic images (<math>\approx 55\%</math>)</li> </ol> <p>The other options were also considered useful but the expert user but with a lower frequency (<math>&lt; 50\%</math>)</p> <p>Types of image search proposed by expert users were the following:</p> <ul style="list-style-type: none"> <li>• Search for images of artists</li> <li>• Search for images related with an specific research topic</li> <li>• Search for maps</li> <li>• Search for graphics</li> <li>• Search for film images</li> </ul>	<b>21</b>

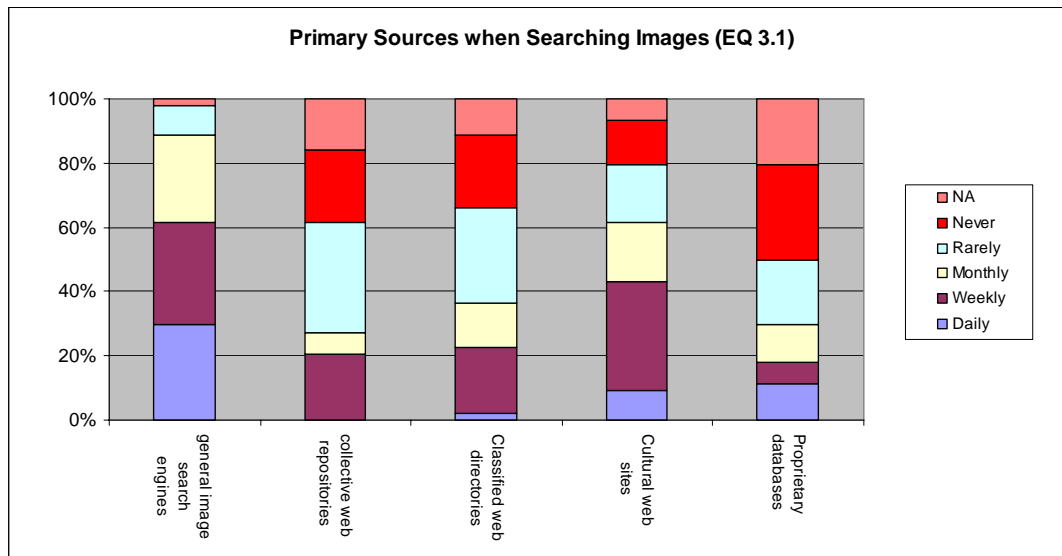


Figure 19: Primary sources for cultural image search (educational group)

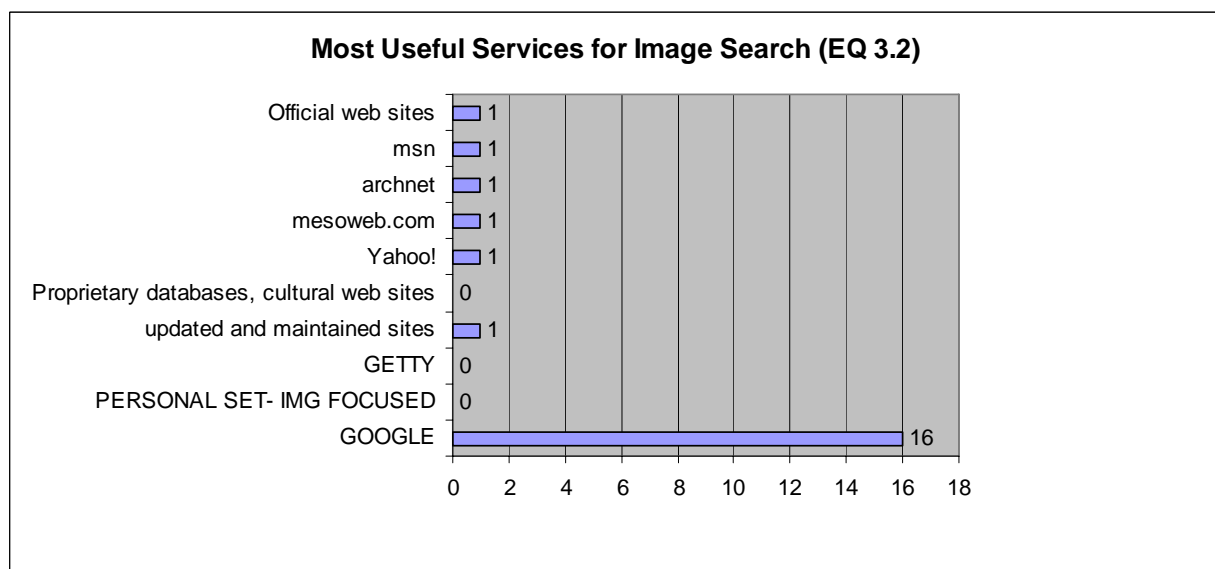
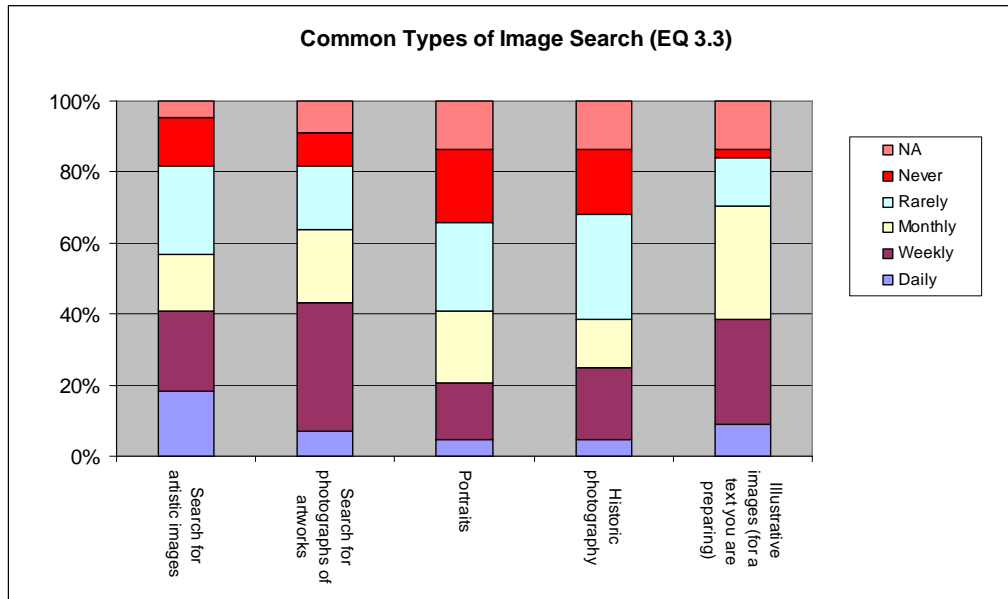


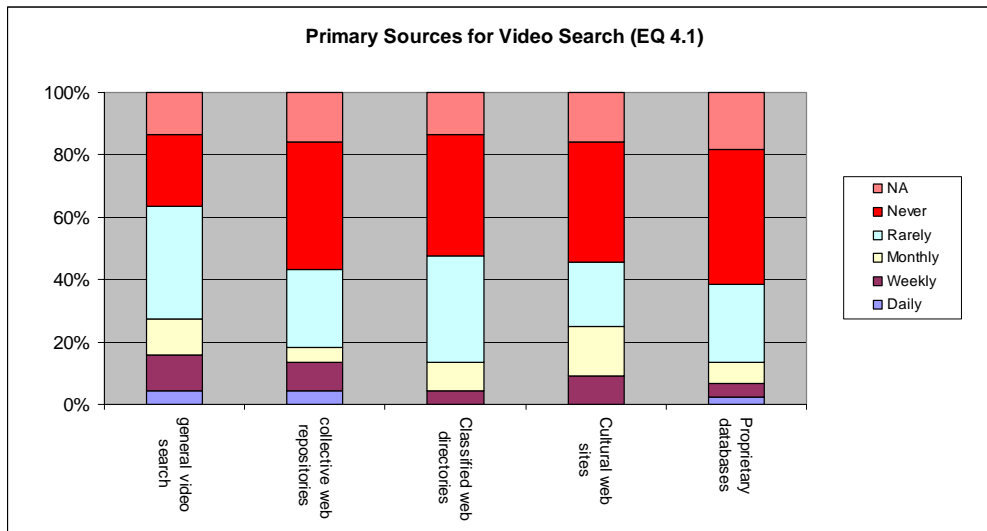
Figure 20: Most useful services for cultural image search (educational group)



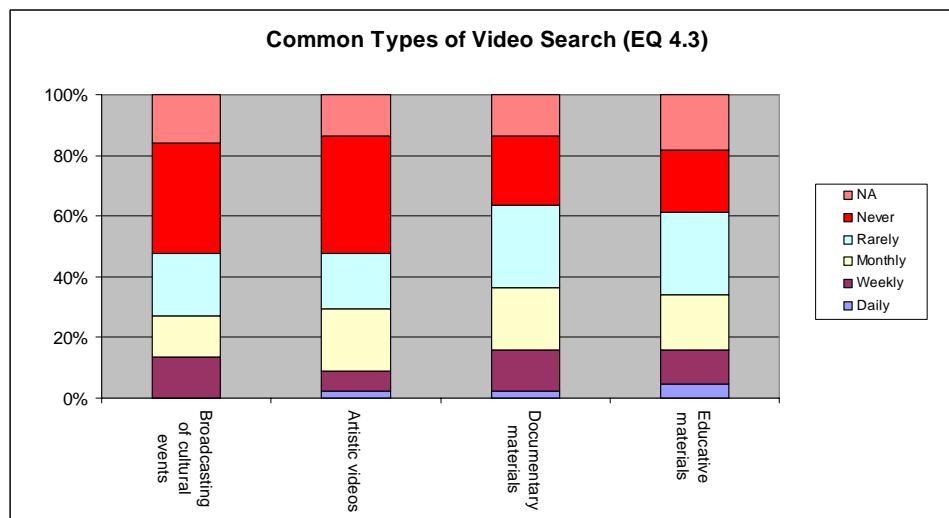
**Figure 21: Most common types of cultural image search (educational group)**

#### E4. Video Search

Question	Comments	Figure
<b>EQ 4.1</b>	<p>The primary sources stated by expert users when searching for videos were the following (for daily, weekly and monthly use):</p> <ol style="list-style-type: none"> <li>3. General video search engines (<math>\approx 30\%</math>)</li> <li>4. Cultural web sites (<math>\approx 30\%</math>)</li> </ol> <p>Proprietary databases, classified web directories and collective web repositories are also used but with a very low frequency.</p>	<b>22</b>
<b>EQ 4.2</b>	The most useful services stated by expert users were mainly focused in the use of general video search engines such as Google	<b>None</b>
<b>EQ 4.3, 4.4</b>	<p>The most common types of video searches were the following (for daily, weekly and monthly use):</p> <ol style="list-style-type: none"> <li>4. Search for documentary materials (<math>\approx 65\%</math>)</li> <li>5. Search for educative materials (<math>\approx 60\%</math>)</li> </ol> <p>The other options were also considered useful by the expert user but with a lower frequency (<math>&lt; 50\%</math>)</p> <p>No relevant feedback by the user was given to enlarge the types of video searches</p>	<b>23</b>



**Figure 22: Primary sources for cultural video search (educational group)**

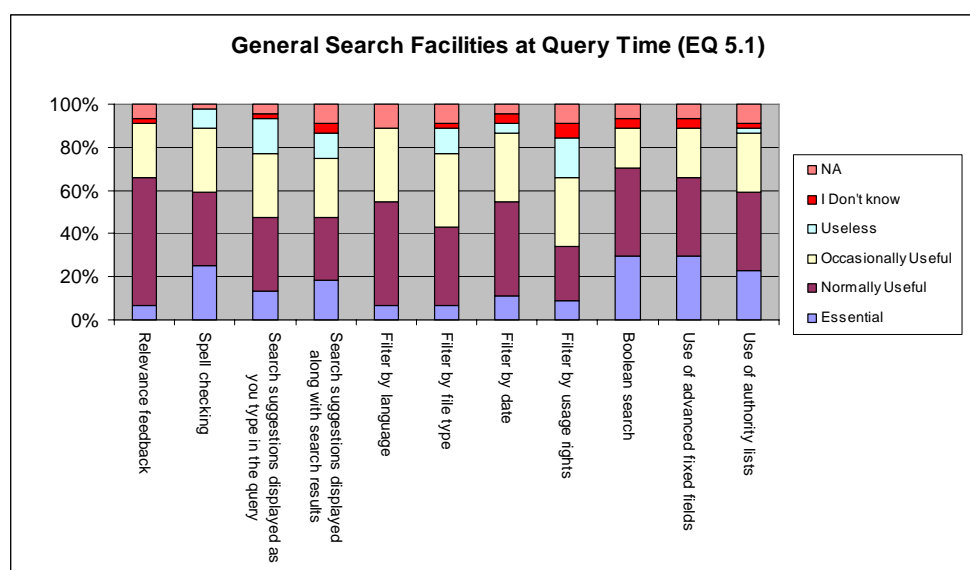


**Figure 23: Most common types of cultural video search (educational group)**

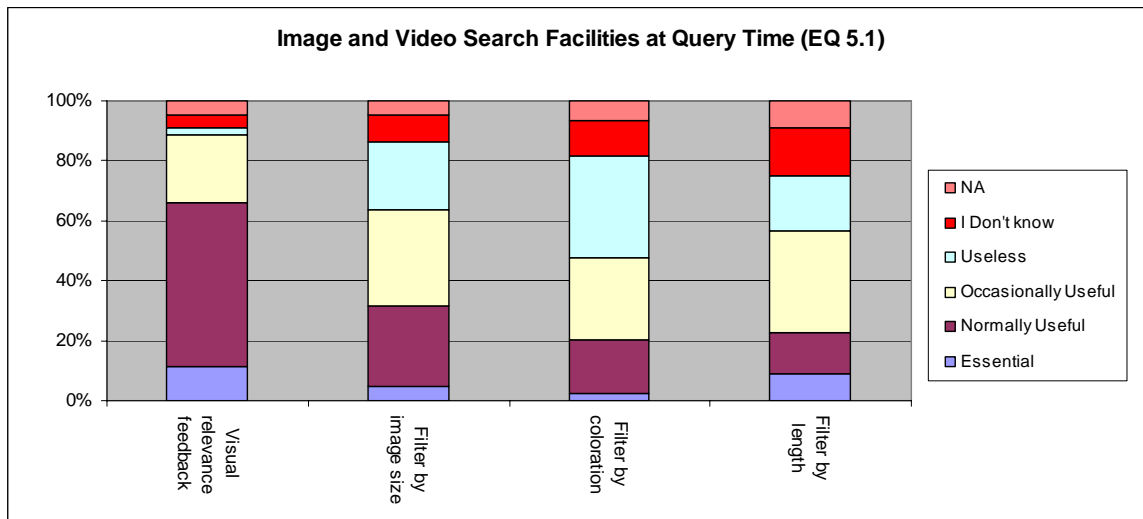
## E5. Desirable Search Facilities

Question	Comments	Figure
<b>EQ 5.1. At query time</b>	<p>The most requested general search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Boolean search (<math>\approx 70\%</math>)</li> <li>2. Use of advanced fixed fields (<math>\approx 65\%</math>)</li> <li>3. Relevance feedback (<math>\approx 65\%</math>)</li> <li>4. Spell checking (<math>\approx 60\%</math>)</li> <li>5. Use of authority list (<math>\approx 60\%</math>)</li> <li>6. Filtering by language (<math>\approx 55\%</math>)</li> <li>7. Filtering by date (<math>\approx 55\%</math>)</li> </ol> <p>Other options such as search suggestions while typing the query or retrieving results, and filtering by type or usage rights were considered useful but with lower frequencies (<math>&lt; 50\%</math>)</p>	<b>24, 25</b>

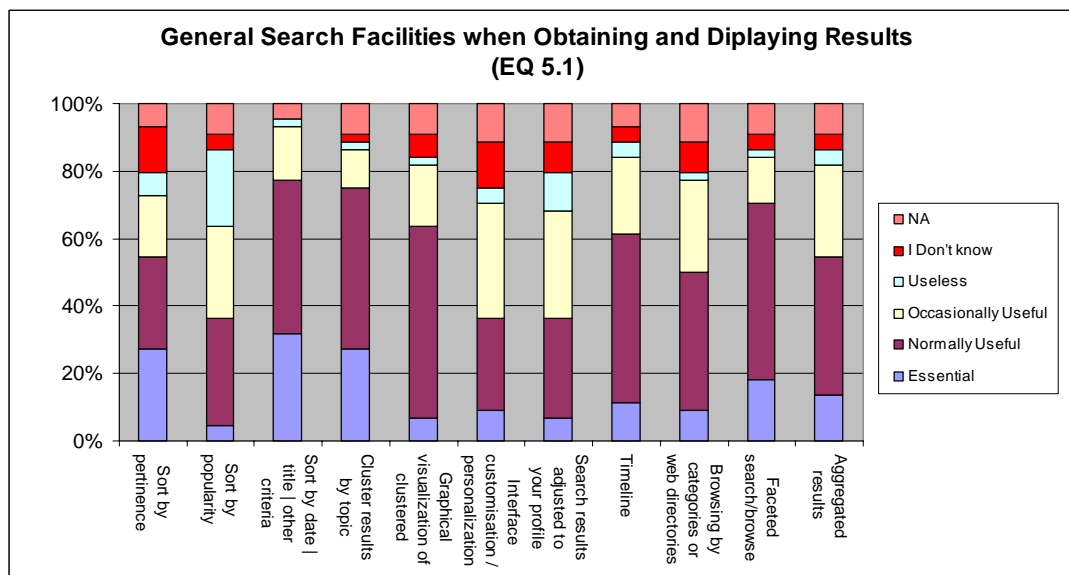
	<p>The most requested image and video specific search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Visual relevance feedback (<math>\approx 70\%</math>)</li> </ol> <p>Visual filtering (such as by coloration, image size or length) were also considered useful but with very low frequencies (<math>&lt; 30\%</math>)</p>	
<b>EQ 5.1 (when obtaining and displaying results)</b>	<p>The most desired general search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Sort by date, title or any other criteria (<math>\approx 80\%</math>)</li> <li>2. Clustering of the results by topic (<math>\approx 75\%</math>)</li> <li>3. Faceted search and browsing (<math>\approx 70\%</math>)</li> <li>4. Graphical visualization of clustered results (<math>\approx 65\%</math>)</li> <li>5. Timeline (<math>\approx 60\%</math>)</li> <li>6. Sort by pertinence (<math>\approx 55\%</math>)</li> <li>7. Aggregated results (<math>\approx 55\%</math>)</li> <li>8. Browsing by categories or web directories (<math>\approx 50\%</math>)</li> </ol> <p>Options related with user profile and search customisation and sorting by popularity were considered useful but with a very low frequency (<math>&lt; 40\%</math>)</p> <p>The most desired image and video specific search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Cluster results by description (<math>\approx 50\%</math>)</li> <li>2. Search for video fragments (<math>\approx 50\%</math>)</li> </ol> <p>The other options proposed (such as different kind of searches and clustering approaches accordingly the contents' visual properties) were considered useful but with lower frequencies (<math>&lt; 50\%</math>).</p>	<b>26,27</b>
<b>EQ 5.1 (when handling results)</b>	<p>The most desired general search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Annotation (<math>\approx 75\%</math>)</li> <li>2. Search history (<math>\approx 65\%</math>)</li> <li>3. Basket (<math>\approx 65\%</math>)</li> </ol> <p>Option related with collaborative search was considered useful but with a 50%</p>	<b>28</b>



**Figure 24: General search facilities at query time (educational group)**

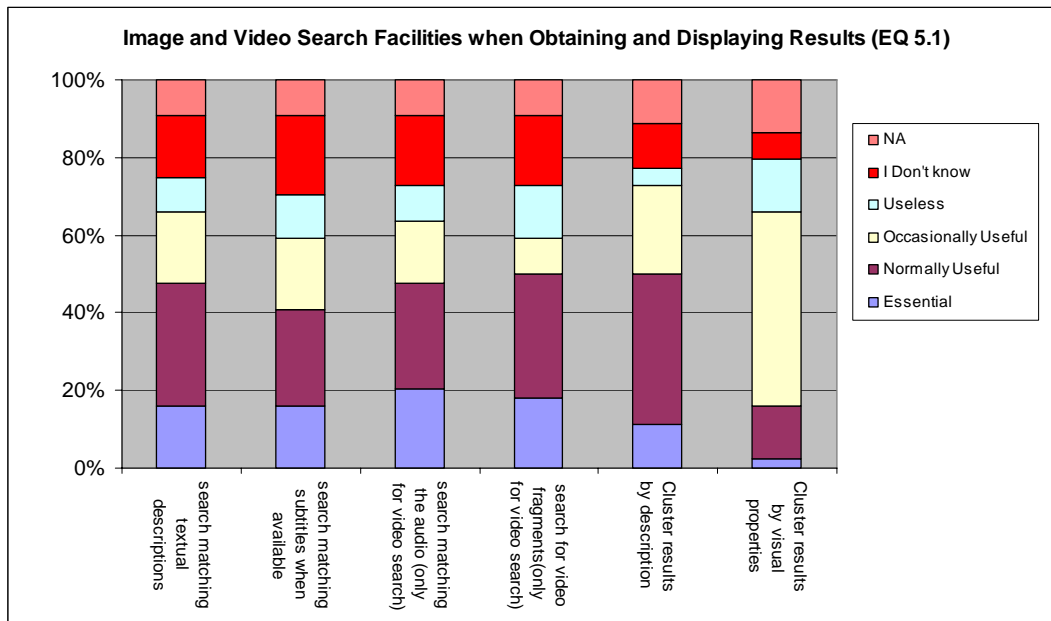


**Figure 25: Specific image and video search facilities at query time (educational group)**

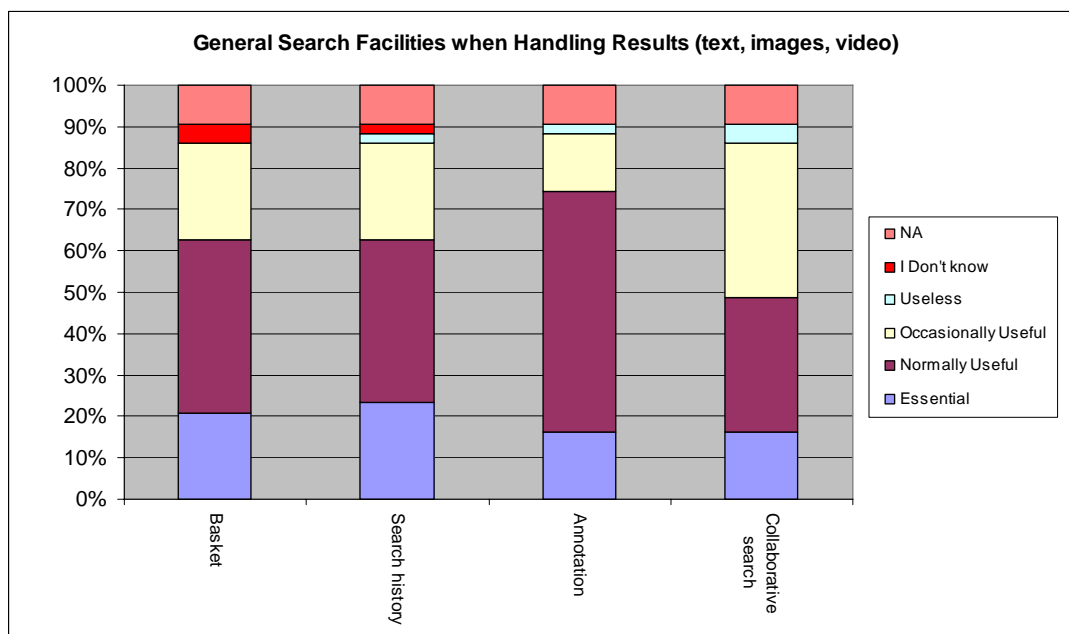


**Figure 26: General search facilities when obtaining and displaying results (educational group)**





**Figure 27: Specific image and video search facilities when obtaining and displaying results (educational group)**



**Figure 28: General search facilities when handling results (educational group)**

## Educational Group Conclusions

<b>E1</b>	<ul style="list-style-type: none"> <li>• The educational experts group is mainly composed by young people directly related with the University and with enough knowledge to manage the Internet in a fluent way. They are familiar with the most general search services and also with specific ones such as library catalogs or databases. Moreover, they consider the use of Internet essential to ease and improve their daily work. However, it is note worthy that some Internet specific services such as newsletters or RSS feeds are not very well known among interviewed experts.</li> <li>• Their searches are usually based on the use of different cultural search services (e.g. cultural web sites, web search engines or catalogs) but not in a connected way. In this sense, they usually look for cultural information in different and separate sources and finally they manually aggregate the results (if necessary).</li> <li>• The most common searches are text searches but image and video are also searched but not so frequently as there is less need for this kind of information.</li> <li>• The most used languages were English and Spanish. These results justify multilingualism in MM. This is because almost all the interviewed experts were Spanish native speakers who considered English almost as useful to them as Spanish when searching for CH information. However, the other proposed languages obtained low rates for their usefulness because the respondents did not have competence in these languages and thus would not attempt to search in them.</li> </ul>
<b>E2</b>	<ul style="list-style-type: none"> <li>• The educational experts group mainly uses general search engines, library catalogues and textual databases and official cultural web sites for text search. They also execute (not very frequently) specific searches using book search engines, encyclopaedic sources and news services. It is noteworthy that specific services such as web directories or vertical search engines are not very well known (20% and 30% of experts did not answer the questions related with these services). Blogs are but it seems they are considered not very useful (probably the experts don't trust them) .</li> <li>• The most common text searches look for information about authors, specific cultural web sites and books. This latter search is mainly conducted using library databases but not book search engines. Scholarly publications, cultural events, news and art periods are also searched but with lower frequencies.</li> <li>• With respect to aggregation, the preferred compilations were for authors but also on artistic items and cultural sites. This last one is mainly from the tourism point of view (e.g. information about nearby cultural sites, hotels, transports).</li> <li>• With respect to information mining and pattern discovery, expert users focused their preferred associations on authors: relationships among authors (things in common) or the masterpieces, authors, places, people and topics most closely associated with them were most important. Information mining about popularity or opinions about authors or masterpieces were not considered as useful by experts.</li> </ul>
<b>E3</b>	<ul style="list-style-type: none"> <li>• The Educational experts group mainly uses general image search engines and official cultural web sites. Specific and general web repositories are also used but with very low frequencies.</li> <li>• Majority of searches are focused on illustrative images (to improve a text) and photographs of artworks.</li> </ul>
<b>E4</b>	<ul style="list-style-type: none"> <li>• The Educational experts group mainly uses general image search engines and official cultural web sites. Specific and general web repositories and proprietary databases are also used but with very low frequencies.</li> <li>• Main searches are focused on documentary and educational materials.</li> </ul>

<b>E5</b>	<ul style="list-style-type: none"> <li>Educational experts detect the following features as very useful to improve their searches:</li> </ul>	
	<b>At query time</b>	
	Query building related	<ul style="list-style-type: none"> <li>Boolean search (they are familiar with this as it is very common in almost all general search engines)</li> <li>Use of advanced fixed fields (as expert users of catalogues and databases they are very familiar with this)</li> <li>Use of authority lists (This is a very common option when retrieving from databases)</li> <li>Relevance feedback</li> <li>Visual relevance feedback</li> </ul>
	Filtering related	<ul style="list-style-type: none"> <li>Spell checking</li> <li>Common filtering options (by date or by language)</li> </ul>
	<b>When obtaining and displaying results</b>	
	Source related	<ul style="list-style-type: none"> <li>Aggregated results</li> </ul>
	Sorting related	<ul style="list-style-type: none"> <li>Common sorting options such as date or title sorting</li> <li>Complex sorting options such as pertinence sorting</li> </ul>
	Browsing related	<ul style="list-style-type: none"> <li>Faceted search and browsing</li> <li>Browsing by categories or web directories</li> <li>Search for video fragments</li> </ul>
	Clustering related	<ul style="list-style-type: none"> <li>Results clustered by topic</li> <li>Results clustered by image/video description</li> </ul>
	Visualization related	<ul style="list-style-type: none"> <li>Graphical visualization of clustered results</li> <li>Timeline</li> </ul>
	<b>When handling results</b>	
	Action related	<ul style="list-style-type: none"> <li>Results annotation</li> </ul>
	History related	<ul style="list-style-type: none"> <li>Search history</li> <li>Basket</li> </ul>

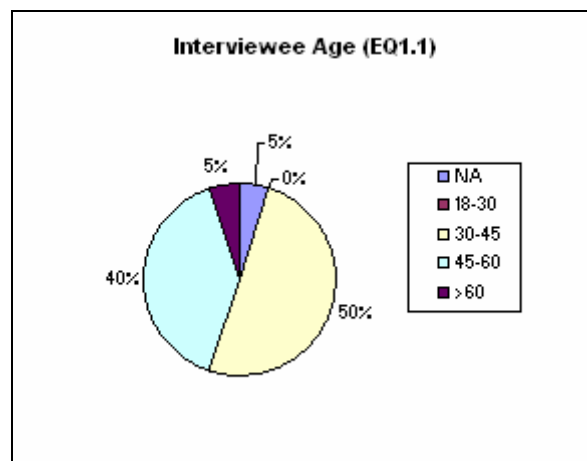
## CH Professional Group Results

The results obtained for professionals working in the CH domain were the following.

### E1. User Profile

Question	Comments	Figure
<b>EQ 1.1</b>	The most common age range of the professional respondents was between 30-45 (50%) followed by 45-60 (40%)	<b>29</b>
<b>EQ 1.2, 1.3</b>	All the CH experts associated their job with the CH area	<b>30, 31</b>
<b>EQ 1.4</b>	55% of professionals considered the WWW to be essential or useful (36%) in their daily work	<b>32</b>
<b>EQ 1.5</b>	Most popular sources of cultural information (considered essential or useful) were the following: <ul style="list-style-type: none"> <li>Regular visits to specialized web sites (<math>\approx 95\%</math>)</li> <li>General search services (<math>\approx 85\%</math>)</li> <li>Specialized search services (<math>\approx 58\%</math>)</li> <li>Use of databases and catalogs (<math>\approx 74\%</math>)</li> </ul>	<b>33</b>

	<ul style="list-style-type: none"> <li>Use of links suggested by colleagues (<math>\approx 80\%</math>)</li> </ul> <p>While RSS feeds from specialized sites and newsletters subscriptions were not as useful (<math>\approx 40\%</math>)</p>	
<b>EQ 1.6</b>	<p>Cultural information search frequency was as follows:</p> <ul style="list-style-type: none"> <li>Cultural text searches are the most common, mainly performed daily or weekly by experts (<math>\approx 90\%</math>), Image searches are performed daily or weekly by 50% of the experts.</li> <li>Video searches are performed occasionally (15%)</li> </ul>	<b>34</b>
<b>EQ 1.7</b>	<p>Preferred search engines were the following:</p> <ul style="list-style-type: none"> <li>For text search: Google, Yahoo!,</li> <li>For image search: professional data bases, Google, Yahoo! and Corbis</li> <li>For video search: Google, Youtube, Limewire</li> </ul>	<b>35, 36</b>
<b>EQ 1.8, 1.9</b>	<p>The majority of professionals interviewed were native Italian (35%) Spanish (30%) English (15%), German (10%) speakers, results of question 1.9 show that English (85%) is the most useful language to get updated and comprehensive cultural information from web sources.</p>	<b>37, 38</b>



**Figure 29: respondents age distribution (CH professionals)**

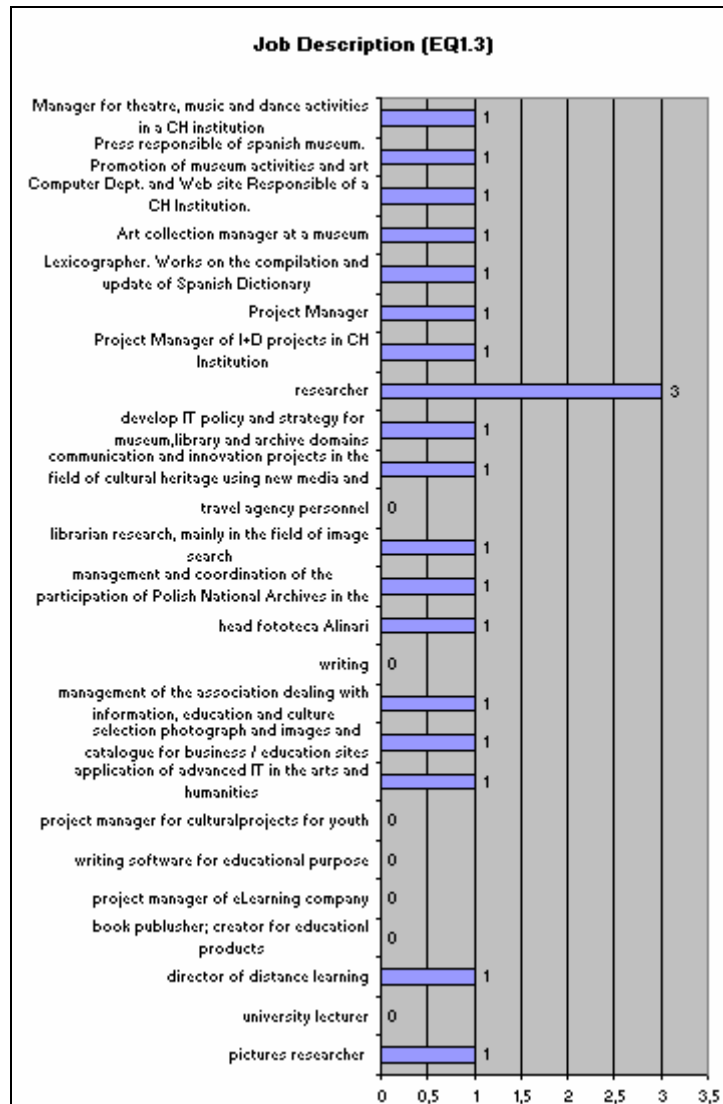


Figure 30: respondents' job description (CH professionals)

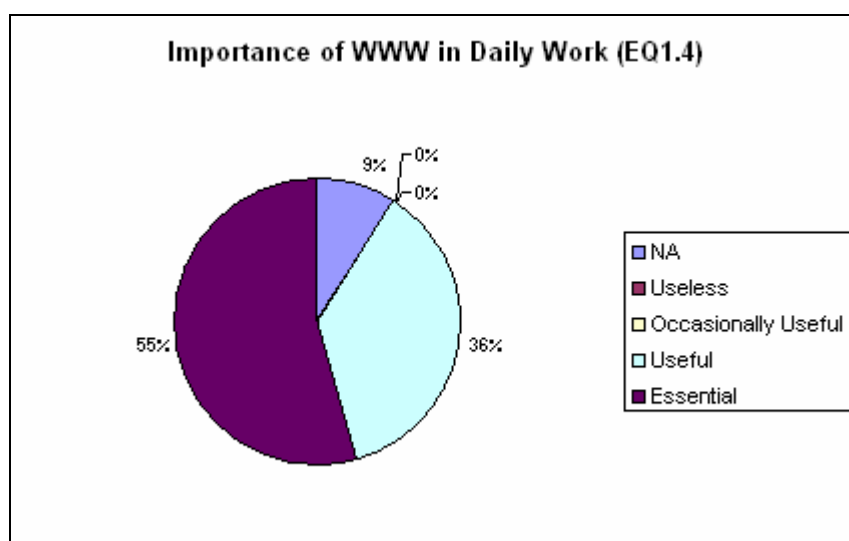


Figure 31: Importance of the WWW in daily work (CH professionals)

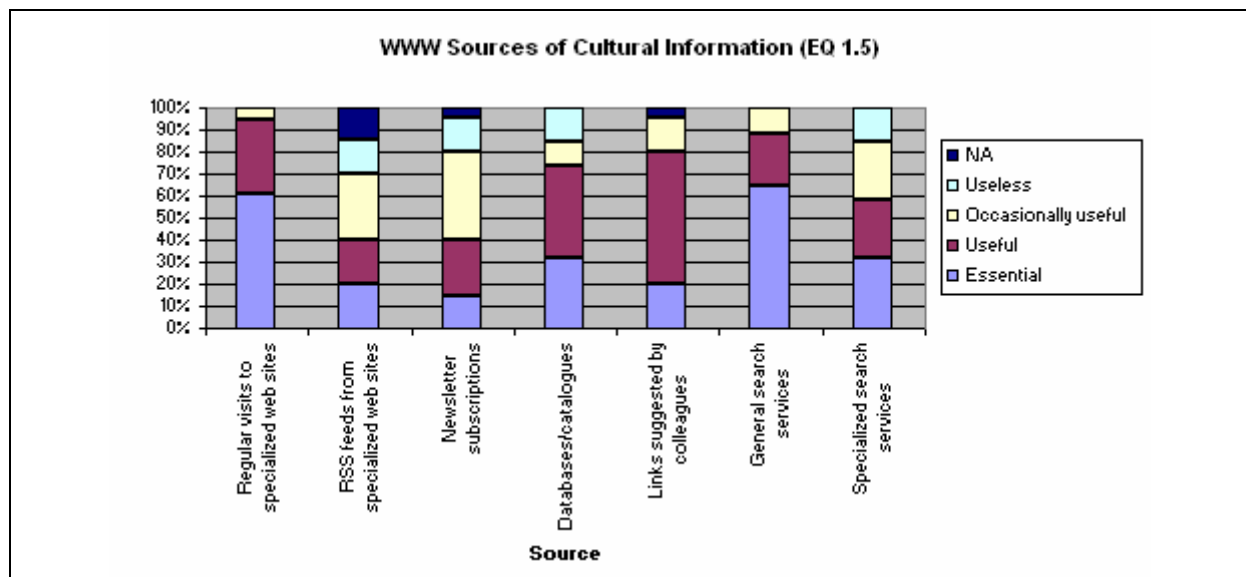


Figure 32: Main sources of cultural information (CH professionals)

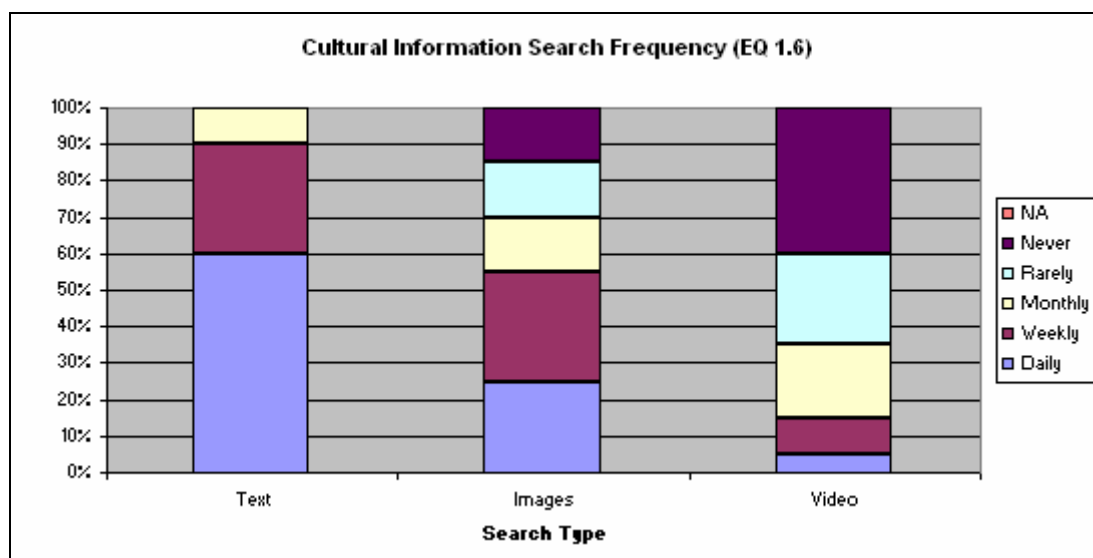


Figure 33: Text, images and video search frequencies (CH professionals)

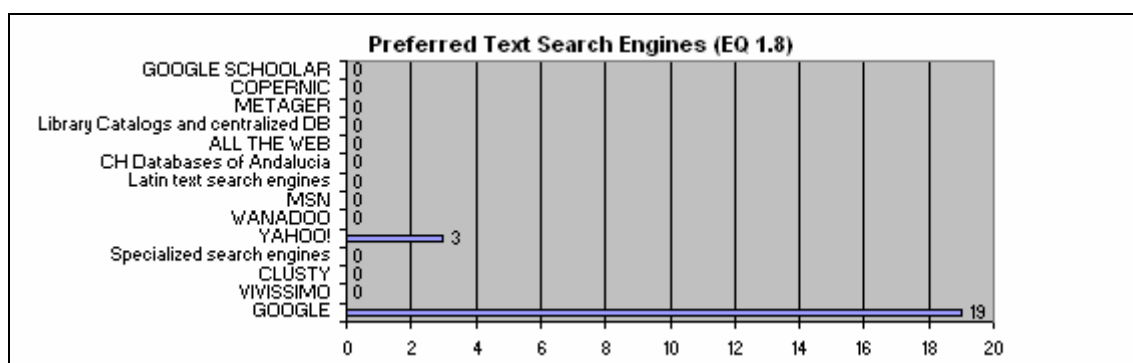


Figure 34: Preferred text search engines (CH professionals)



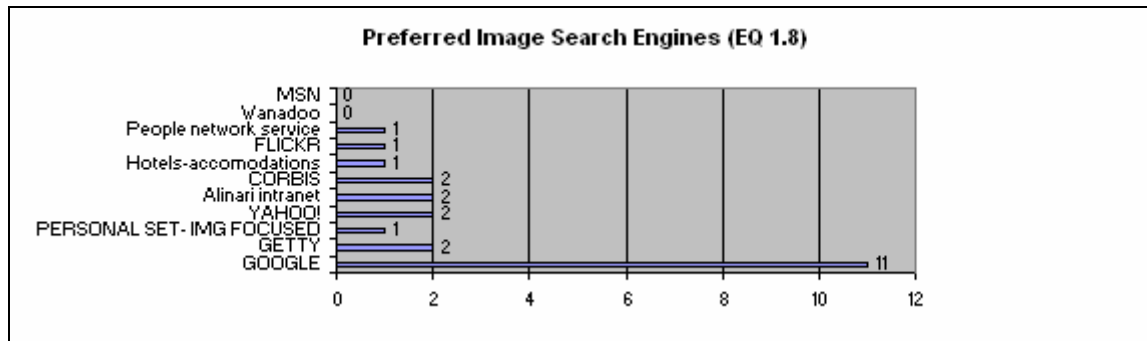


Figure 35: Preferred image search engines (CH professionals)

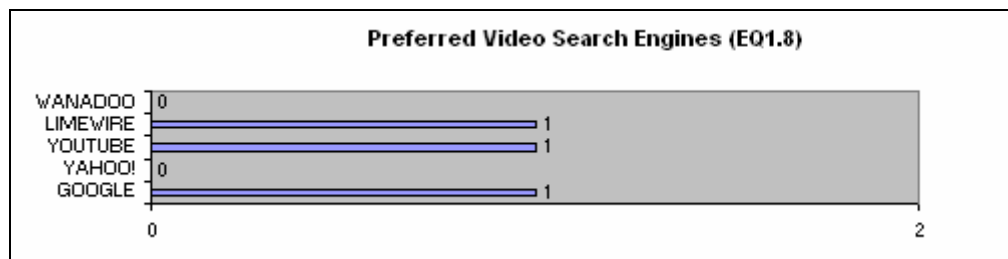


Figure 36: Preferred video search engines (CH professionals)

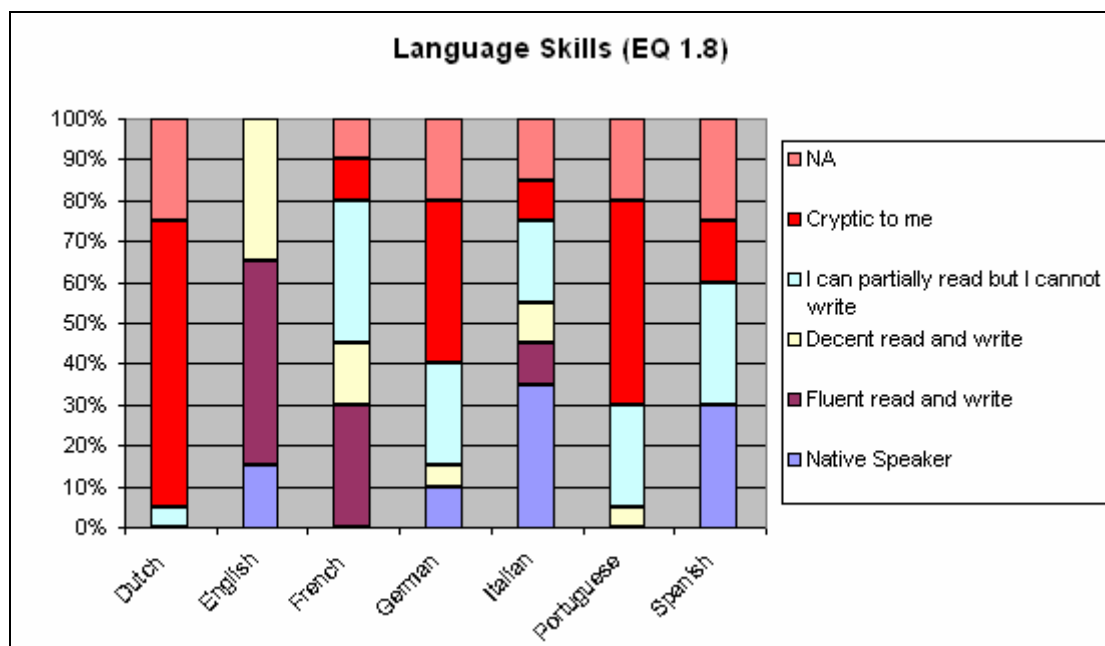
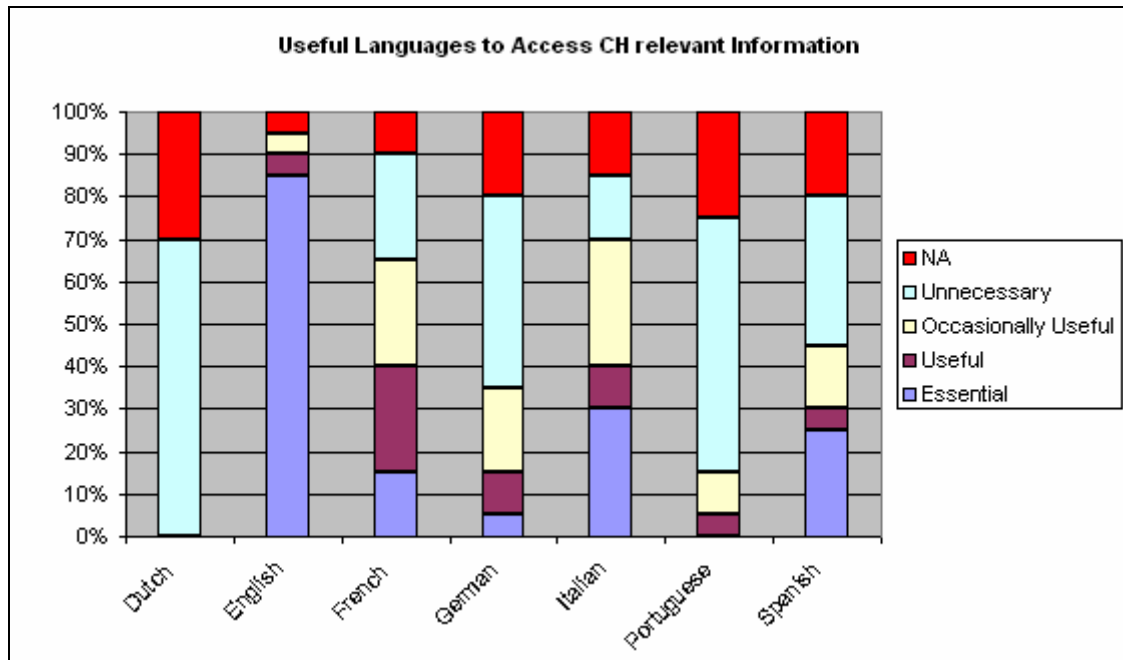


Figure 37: Experts language skills (CH professionals)



**Figure 38: Considerations about which languages are more useful for CH access (CH professionals)**

## E2. Textual Search

Question	Comments	Figure
<b>EQ 2.1</b>	<p>The main sources when searching text cultural information were the following (for daily, weekly or monthly use):</p> <ul style="list-style-type: none"> <li>General web search engines (e.g. Google, Yahoo!, MSN, Ask) (<math>\approx 100\%</math>)</li> <li>Library catalogues and other Textual databases (<math>\approx 80\%</math>)</li> <li>Vertical search engines (domain-specific) (<math>\approx 50\%</math>)</li> <li>Encyclopaedic sources (e.g. Wikipedia) (<math>\approx 80\%</math>)</li> <li>Official web sites from museums, Universities (<math>\approx 85\%</math>)</li> <li>News services (<math>\approx 55\%</math>)</li> <li>Other sources were not considered as useful (<math>&lt; 40\%</math> of daily or weekly use): <ul style="list-style-type: none"> <li>Web directories (e.g. Yahoo!, dmoz.org) (<math>\approx 35\%</math>)</li> <li>Informal sources such as blogs (<math>\approx 20\%</math>)</li> <li>Book search engines (e.g. Google print, Amazon) (<math>\approx 40\%</math>)</li> </ul> </li> </ul>	<b>39</b>
<b>EQ 2.2</b>	<p>The most useful services were:</p> <ol style="list-style-type: none"> <li>General search engines (such as Google, Yahoo! or MSN)</li> <li>Wikipedia, Yahoo!</li> <li>Library catalogues and textual databases</li> <li>Official web sites</li> </ol>	<b>40</b>
<b>EQ 2.3, 2.4</b>	<p>Almost all the text searches proposed in the questionnaire were accepted as useful in daily, weekly or monthly work (except for search for an art period that is <math>&lt; 40\%</math>).</p> <ol style="list-style-type: none"> <li>Search for specific authors (<math>\approx 80\%</math>)</li> <li>Search for specific web sites (<math>\approx 100\%</math>)</li> <li>Search for books (<math>\approx 70\%</math>)</li> </ol>	<b>41</b>

	<p>11. Search for scholar publications (<math>\approx 85\%</math>)</p> <p>12. Search for cultural events (<math>\approx 85\%</math>)</p> <p>13. Search for news (<math>\approx 70\%</math>)</p> <p>14. Search for an art period (<math>\approx 40\%</math>)</p> <p>Textual searches proposed by professionals were the following:</p> <ul style="list-style-type: none"> <li>• Search for quotes</li> <li>• search on technology</li> <li>• search for title of the document- not books only but other kinds of documents such as studies, articles, posts etc.</li> <li>• search for issue (keywords)- it is similar to search for scholar publications but not the same because covers other kinds of documents</li> <li>• search in databases. Specific words or texts of one or more classic authors</li> <li>• Thematic searches</li> <li>• Search for an archaeological monument (e.g. Stonehenge)</li> <li>• Search for a town (e.g. Athens)</li> </ul>	
<b>EQ 2.5, 2.6</b>	<p>With respect to information compilation and aggregation professional users considered essential or normally useful the following compilations:</p> <ul style="list-style-type: none"> <li>▪ Announcements of cultural events (<math>\approx 90\%</math>)</li> <li>▪ Online cultural events (<math>\approx 90\%</math>)</li> <li>▪ Compiled information about authors (<math>\approx 75\%</math>)</li> <li>▪ Compiled information about artistic periods (<math>\approx 75\%</math>)</li> <li>▪ Compiled information about artistic items (paintings, musical compositions, writings) (<math>\approx 85\%</math>)</li> <li>▪ Compiled information about cultural sites (other cultural sites near to visit, tourism and accommodation information) (<math>\approx 85\%</math>)</li> <li>▪ Job posts in the cultural area (<math>\approx 65\%</math>)</li> </ul> <p>The other options presented were also considered normally useful but the percentage was always under 50%.</p> <p>CH group experts made few suggestions about other useful kind of compilations:</p> <ul style="list-style-type: none"> <li>▪ Compiled information about cultural items</li> <li>▪ Compiled information by features such as lang, country or topic</li> </ul>	<b>42</b>
<b>EQ 2.7, 2.8</b>	<p>Regarding information extraction and pattern discovering, professional users considered essential or normally useful the following relations:</p> <ul style="list-style-type: none"> <li>▪ compared popularity of artists/masterpieces/cultural sites (<math>\approx 70\%</math>)</li> <li>▪ places/people/topics most closely associated with an artist. (<math>\approx 85\%</math>)</li> <li>▪ artists/masterpieces/cultural sites most closely associated with a city. (<math>\approx 90\%</math>)</li> <li>▪ opinions/feelings/moods related to an artist/masterpiece/cultural site. (<math>\approx 65\%</math>)</li> <li>▪ artists/masterpieces/CH sites most closely related to a given keyword (<math>\approx 70\%</math>)</li> <li>▪ masterpieces most closely associated to an author (<math>\approx 80\%</math>)</li> <li>▪ Authors most closely associated to a given author (social network of artists) (<math>\approx 65\%</math>)</li> <li>▪ Things in common between two or more authors (<math>\approx 65\%</math>)</li> </ul> <p>CH professionals didn't make suggestions about other useful kind of compilations.</p>	<b>43</b>

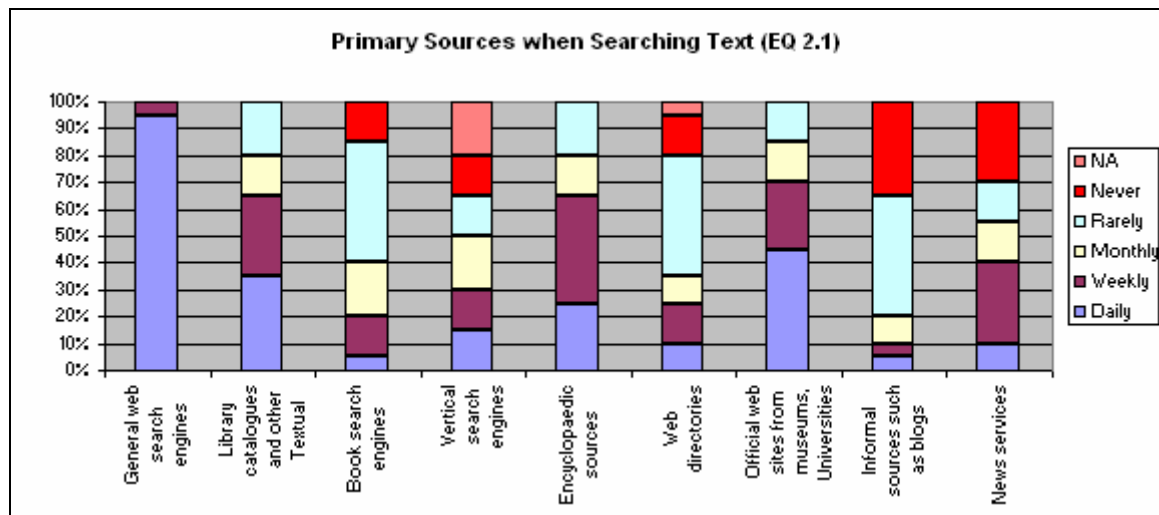


Figure 39: Primary sources for textual cultural searches (CH professionals)

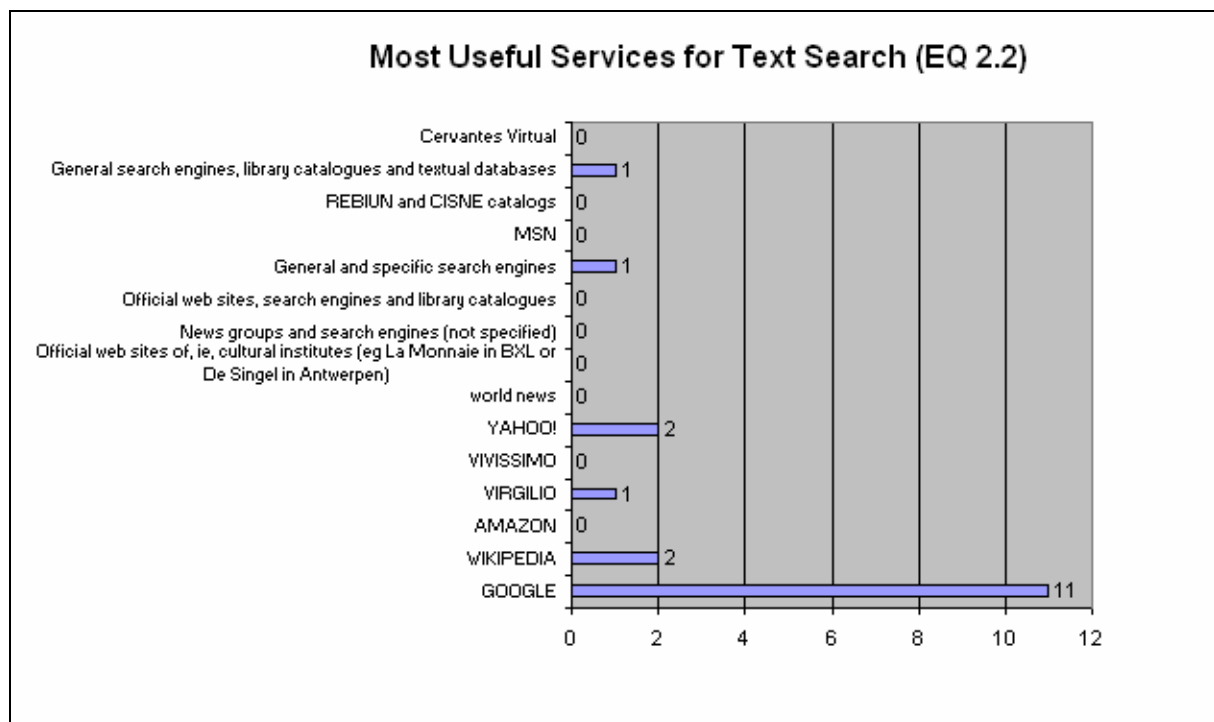


Figure 40: Most useful services for textual cultural searches (CH professionals)

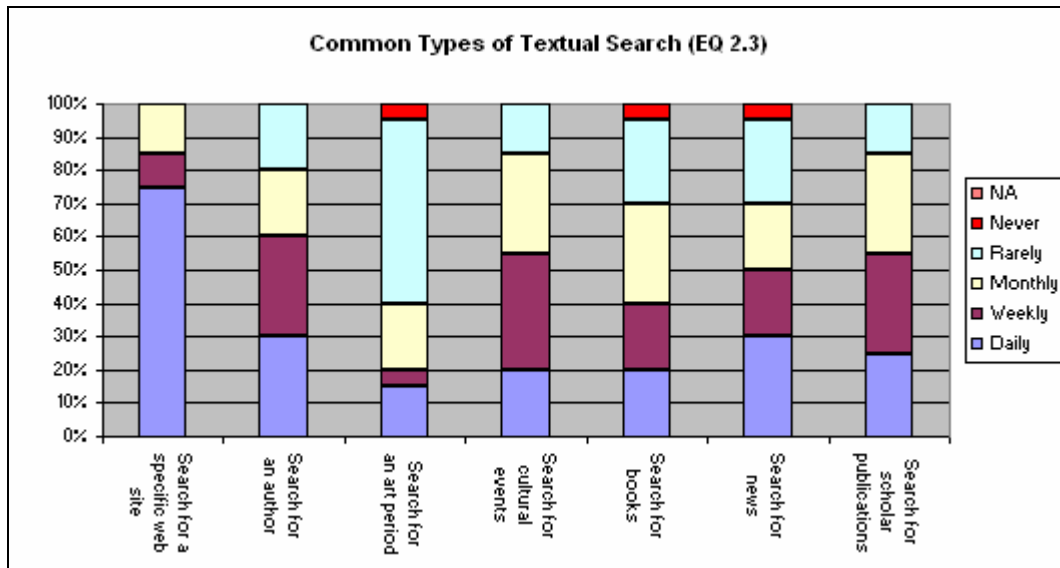


Figure 41: Common types of textual search (CH professionals)

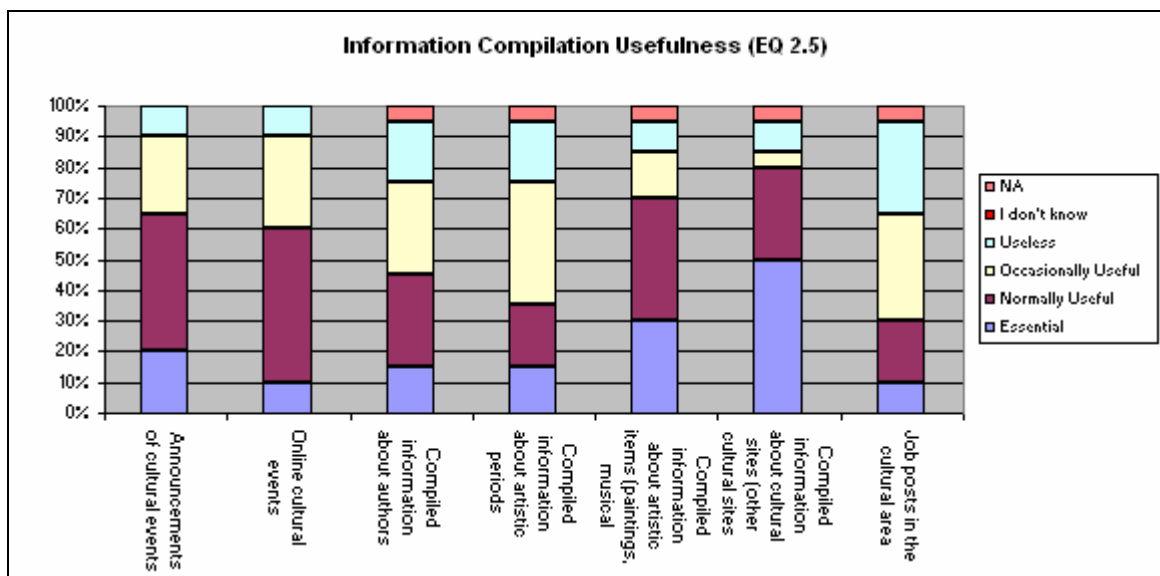
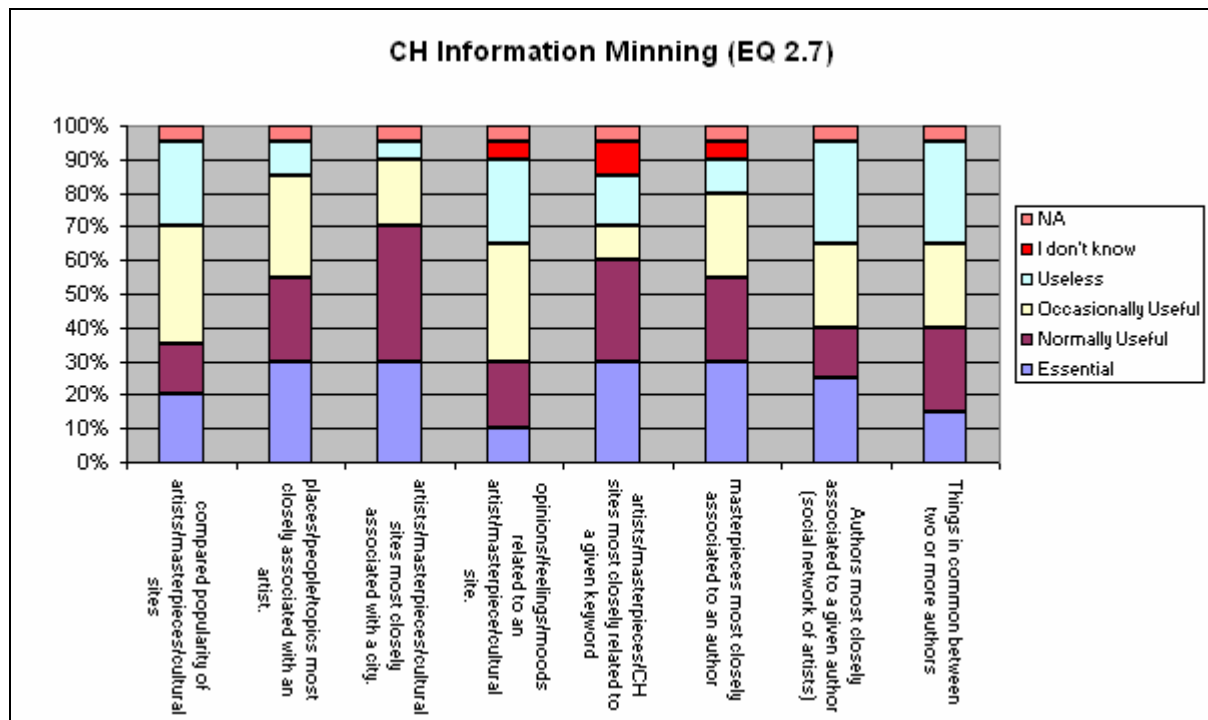


Figure 42: CH Information compilation usefulness (CH professionals)



**Figure 43: CH information mining usefulness (CH professionals)**

### E3. Image Search

Question	Comments	Figure
<b>EQ 3.1</b>	<p>The primary sources when searching for images were the following (for daily, weekly and monthly use):</p> <ul style="list-style-type: none"> <li>general image search engines (e.g. Google, Yahoo!...) (~65%)</li> <li>Cultural web sites (~55%)</li> <li>Proprietary databases (~45%)</li> </ul> <p>With lower frequencies (&lt;40%) are:</p> <ul style="list-style-type: none"> <li>collective web repositories (e.g. Flickr) (~30%)</li> <li>Classified web directories (~30%)</li> </ul>	<b>44</b>
<b>EQ 3.2</b>	The most used services were mainly general image search engines such as Google	<b>No figures</b>
<b>EQ 3.3, 3.4</b>	<p>The most common types of image searches were the following:</p> <ul style="list-style-type: none"> <li>Search for photographs of artworks (explanation: paintings, buildings, sculptures) (~60%)</li> <li>Historic photography (~60%)</li> <li>Illustrative images (for a text under preparation) (~65%)</li> </ul> <p>Other options were also considered useful but with a lower frequency (&lt;50%)</p> <ul style="list-style-type: none"> <li>Search for artistic images (~35%)</li> <li>Portraits (~30%)</li> </ul> <p>Types of image search proposed by expert users were the following:</p> <ul style="list-style-type: none"> <li>images of artists</li> <li>related with an specific research topic</li> <li>Maps, Graphics and film images</li> </ul>	<b>45</b>



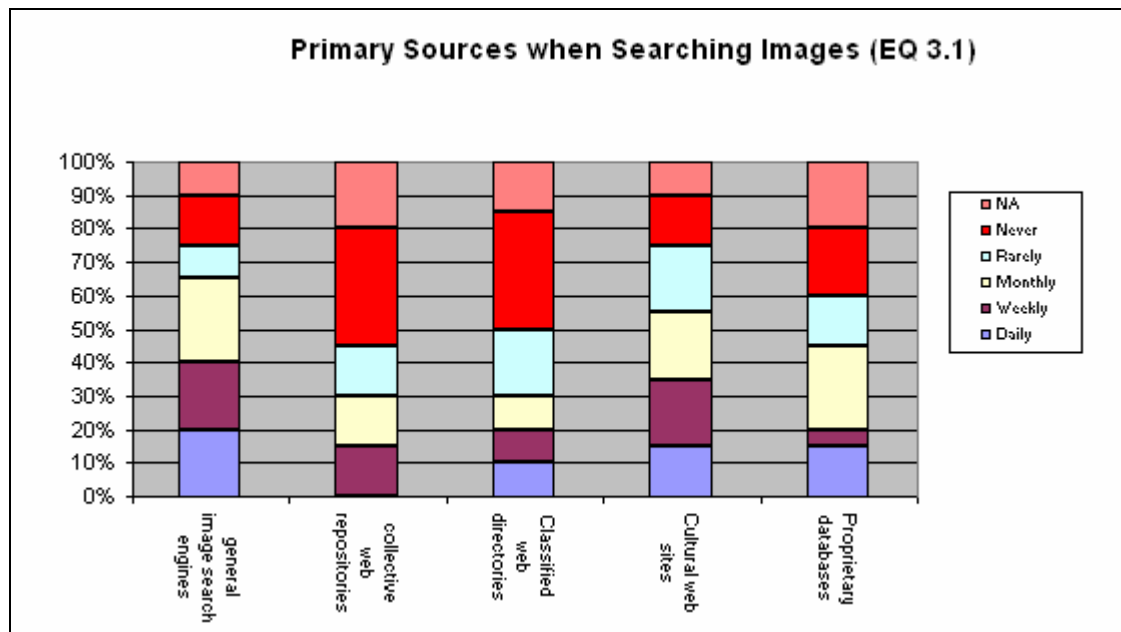


Figure 44: Primary sources for cultural image search (CH professionals)

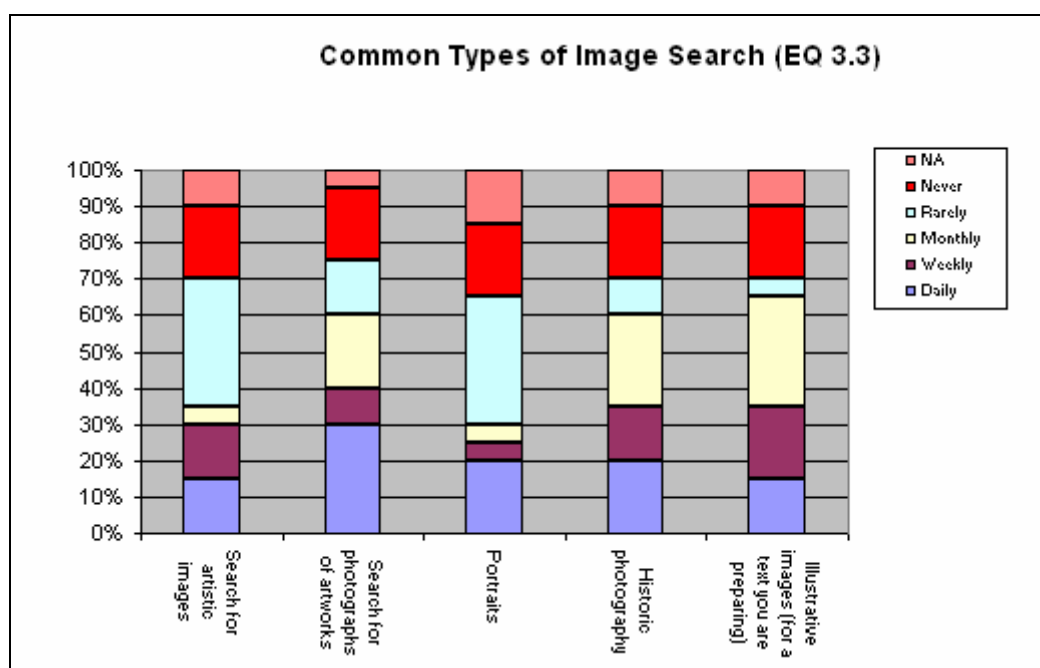
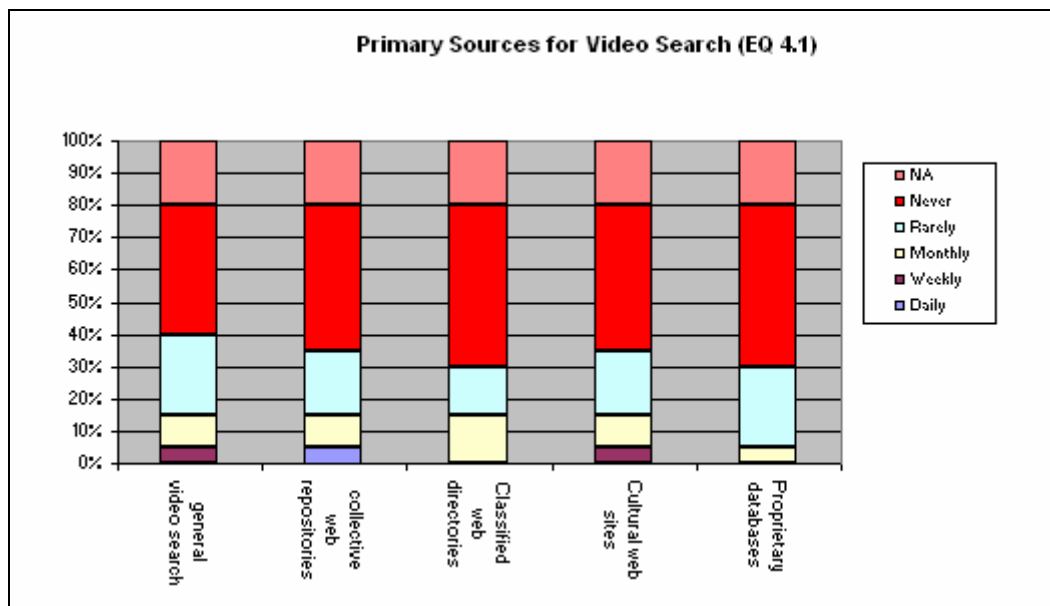


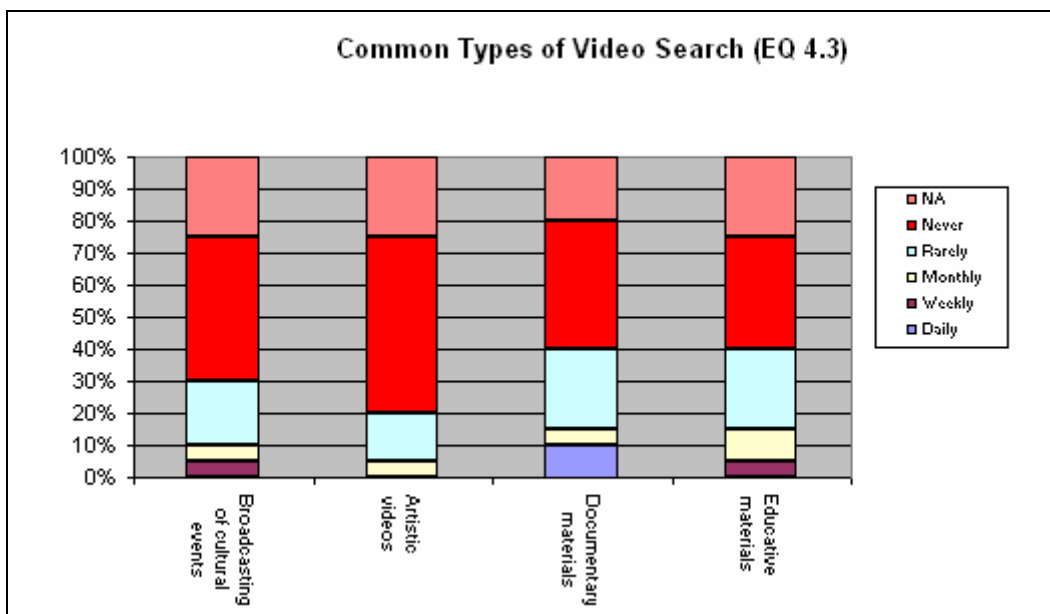
Figure 45: Most common types of cultural image search (CH professionals)

## E4. Video Search

Question	Comments	Figure
EQ 4.1	The use of primary sources for searching videos was always below 15% (daily, weekly and monthly)	46
EQ 4.2	The most useful services were mainly general video search engines such as Google	None
EQ 4.3, 4.4	The most common types of video searches were always below 15% (daily, weekly and monthly use)	47



**Figure 46: Primary sources for cultural video search (CH professionals)**



**Figure 47: Most common types of cultural video search (CH professionals)**

## E5. Desirable Search Facilities

Question	Comments	Figure
<b>EQ 5.1. At query time</b>	<p>The most requested general search facilities were the following (considered as essential or normally useful):</p> <ul style="list-style-type: none"> <li>Relevance feedback(explanation: “more like this image”, “images with this shape”) (≈90%)</li> <li>Spell checking (explanation: “did you mean..?”) (≈95%)</li> <li>Search suggestions displayed as you type in the query (≈100%)</li> <li>Search suggestions displayed along with search results (≈95%)</li> </ul>	<b>48, 49</b>

	<ul style="list-style-type: none"> <li>Filter by language (<math>\approx 90\%</math>)</li> <li>Filter by file type (html, PDF, jpg, gif, avi, mpeg) (<math>\approx 85\%</math>)</li> <li>Filter by date (<math>\approx 100\%</math>)</li> <li>Filter by usage rights (<math>\approx 80\%</math>)</li> <li>Boolean search (explanation: exclude these words, search documents including any of these words, including all of these words) (<math>\approx 95\%</math>)</li> <li>Use of advanced fixed fields (explanation: fields specifically oriented to query for authors, artists, buildings) Similar to query a database (<math>\approx 95\%</math>)</li> <li>Use of authority lists (explanation: lists of authors, artists, buildings) Similar to obtain a complete field database view (<math>\approx 85\%</math>)</li> </ul> <p>The most requested image and video specific search facilities were the following (considered as essential or normally useful by the expert):</p> <ul style="list-style-type: none"> <li>- Visual relevance feedback (explanation: “more like this image”, “images with this shape”) (<math>\approx 60\%</math>)</li> <li>- Filter by image size (<math>\approx 65\%</math>)</li> <li>- Filter by coloration (<math>\approx 50\%</math>)</li> </ul> <p>Visual filtering (such as by length) were also considered useful but with very low frequencies (<math>&lt; 30\%</math>)</p>	
<b>EQ 5.1 (when obtaining and displaying results)</b>	<p>The most requested general search facilities were the following (considered as essential or normally useful by the professionals):</p> <ul style="list-style-type: none"> <li>sort by pertinence (<math>\approx 90\%</math>)</li> <li>Sort by popularity (explanation: input from other users) (<math>\approx 55\%</math>)</li> <li>Sort by date   title   other criteria (<math>\approx 100\%</math>)</li> <li>Cluster results by topic (grouping related documents together) (<math>\approx 95\%</math>)</li> <li>Graphical visualization of clustered results (e.g. kartoo.com) (<math>\approx 60\%</math>)</li> <li>Interface customisation / personalization facilities (<math>\approx 75\%</math>)</li> <li>Search results adjusted to your profile (personalized search) (<math>\approx 90\%</math>)</li> <li>Timeline (explanation: results organized chronologically by periods) (<math>\approx 100\%</math>)</li> <li>Browsing by categories or web directories (<math>\approx 85\%</math>)</li> <li>Faceted search/browse (e.g. constraining the theme and the artistic period of a painting simultaneously) (<math>\approx 85\%</math>)</li> <li>Aggregated results (e.g. displaying results from the web and from online encyclopaedias simultaneously) (<math>\approx 85\%</math>)</li> </ul> <p>The most requested image and video specific search facilities were the following (considered as essential or normally useful by the professionals):</p> <ol style="list-style-type: none"> <li>Cluster results by description (<math>\approx 50\%</math>)</li> </ol> <p>The other options proposed (such as different kind of searches and clustering approaches according to the contents’ visual properties) were considered useful but with lower frequencies (<math>&lt; 50\%</math>).</p>	<b>50, 51</b>
<b>EQ 5.1 (when handling results)</b>	<p>The most requested general search facilities were the following (considered as essential or normally useful by the professionals):</p> <ul style="list-style-type: none"> <li>Basket (explanation: save together items picked up from search results) (<math>\approx 90\%</math>)</li> <li>Search history (explanation: see my previous searches) (<math>\approx 90\%</math>)</li> <li>Annotation (explanation: annotate search results with keywords and comments) (<math>\approx 85\%</math>)</li> <li>Collaborative search (explanation: see findings and annotations made by user communities I trust) (<math>\approx 75\%</math>)</li> </ul>	<b>52</b>

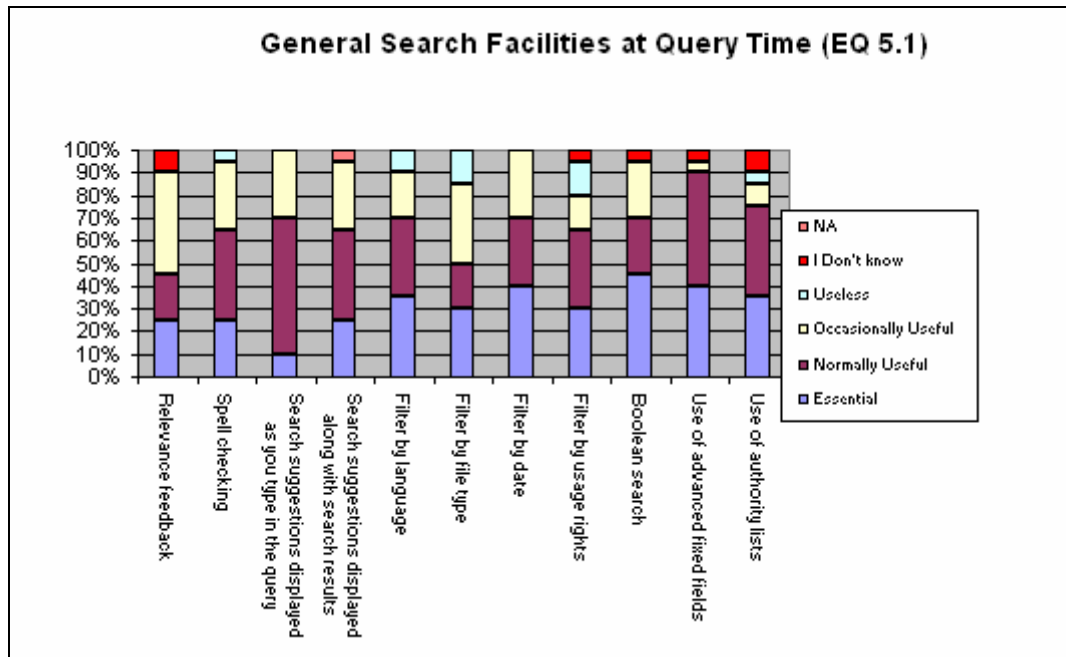


Figure 48: General search facilities at query time (CH professionals)

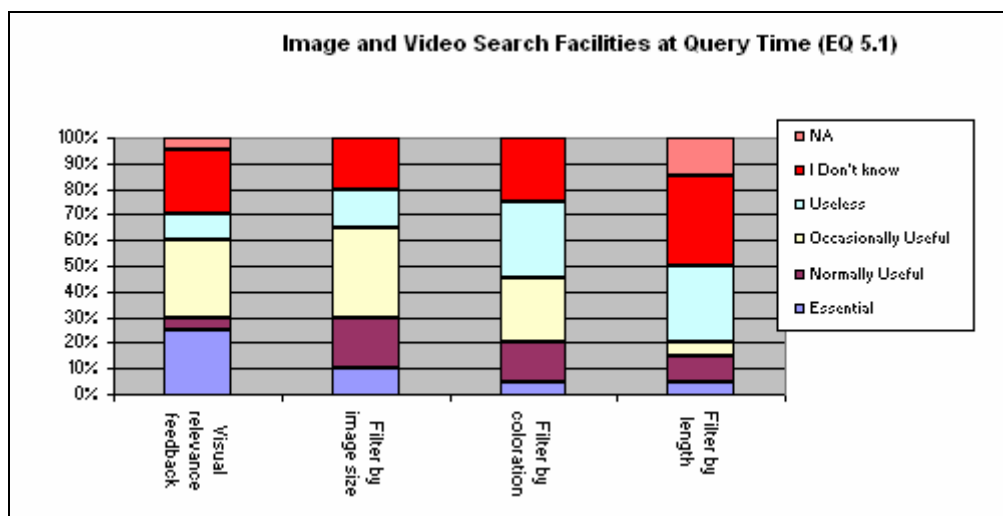


Figure 49: Specific image and video search facilities at query time (CH professionals)

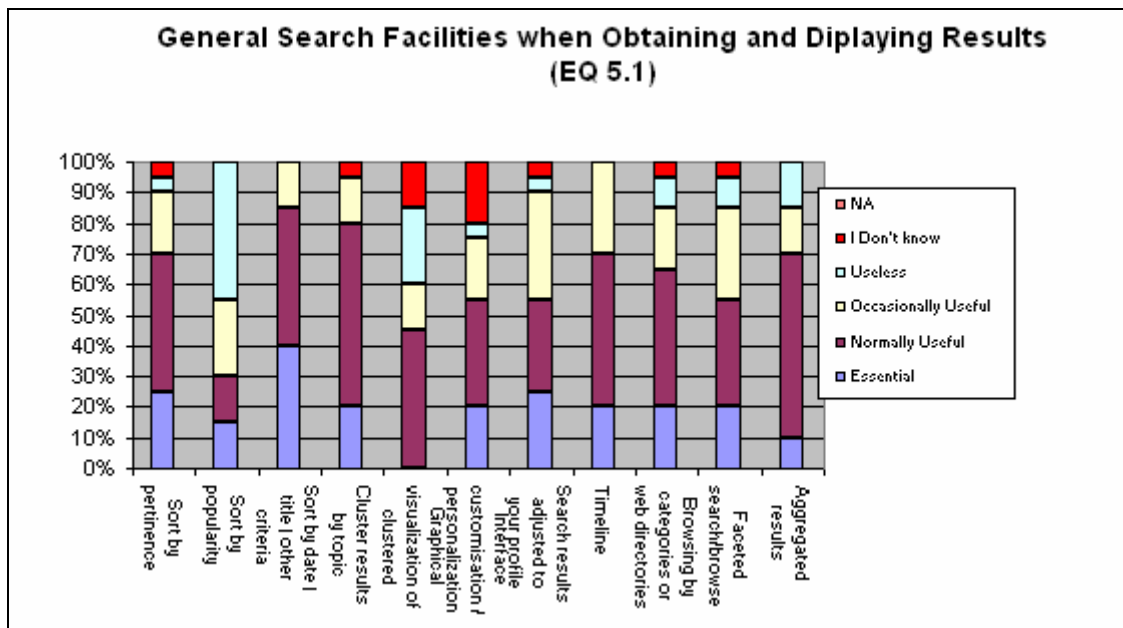


Figure 50: General search facilities when obtaining and displaying results (CH professionals)

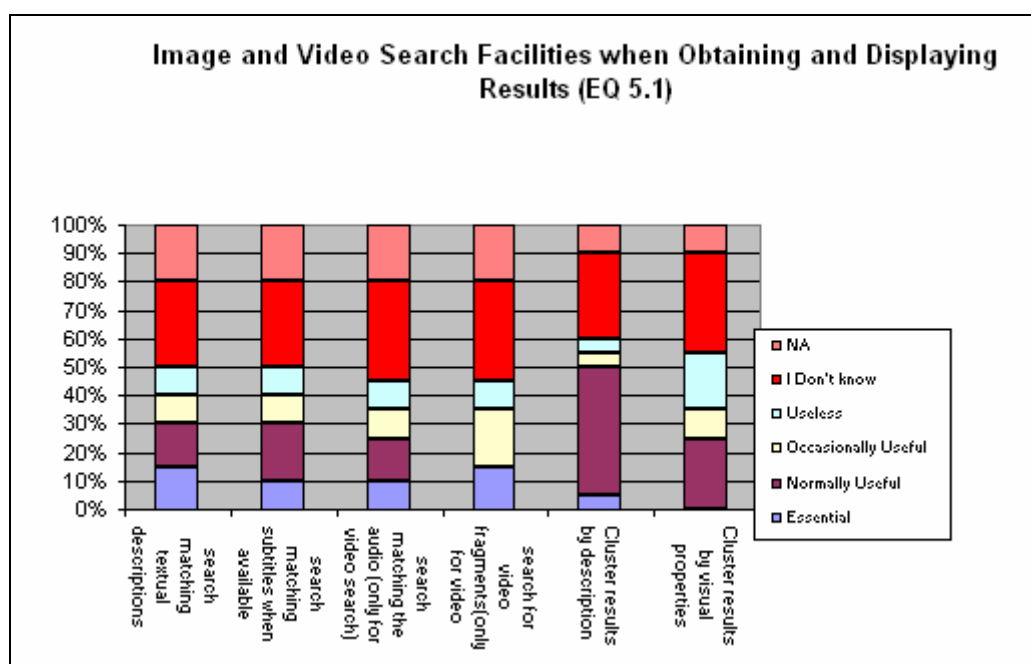
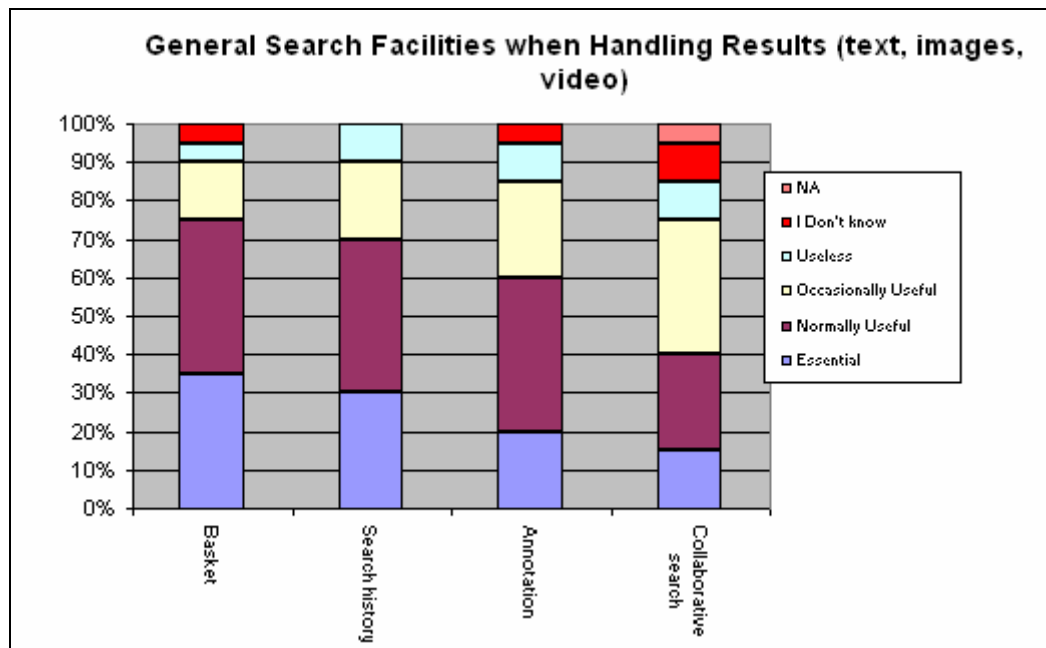


Figure 51: Specific image and video search facilities when obtaining and displaying results (CH professionals)



**Figure 52: General search facilities when handling results (CH professionals)**

## CH Group Conclusions

<b>E1</b>	<p>The most common age range of the CH professional respondents was between 30-45 (50%) followed by 45-60 (40%). All the experts associated their job with the CH area (head of CH departments, researchers, content creators, e-Learning project managers, etc.).</p> <p>55% of experts considered the WWW essential in their daily work and 36% useful .</p> <p>Users regularly visit specialized web sites, general search services, specialized search services, databases and catalogs. While RSS feeds from specialized sites and newsletters subscriptions were not considered very useful .</p> <p>Cultural information search is the most common, mainly performed daily or weekly by these experts. While image searches are performed daily or weekly by 50% of the professionals, video searches are performed only occasionally.</p> <p>The preferred search engines were generalist (for text search: Google, Yahoo!; for image search: professional data bases, Google, Yahoo! and Corbis; For video search: Google, Youtube, Limewire) English and local languages are the most adopted to get updated and comprehensive cultural information from web sources.</p>
<b>E2</b>	<p>When searching for text CH professionals mainly use: general web search engines (e.g. Google, Yahoo!, MSN, Ask), library catalogues and other textual databases; vertical search engines (domain-specific); encyclopaedic sources (e.g. Wikipedia); official web sites from museums and news services. The assumption is that no specialized databases collect as much information as the generalist engines do. However, content brokers mainly work with specialized databases (such as Getty's).</p> <p>Almost all the text searches proposed in the questionnaire were accepted as useful in daily, weekly or monthly work.</p> <p>Text searches proposed by these experts were related to search for citations, search related to technology and meta data search for the title of documents; search for issue (keywords); search for scholar publications; search for specific words or texts of one or more classic authors; thematic searches; search for an archaeological monument (e.g. Stonehenge);search for a town (e.g. Athens).</p> <p>With respect to information compilation and aggregation CH professionals considered essential or normally useful the announcements of cultural events with compilation of information about authors; artistic items (paintings, musical compositions, writings) and cultural sites (other cultural</p>

	<p>sites near to visit, tourism and accommodation information).</p> <p>With respect to information extraction and pattern discovering, expert users considered essential or normally useful in particular sites most closely associated with a city, artist or an author.</p>																
<b>E3</b>	<p>CH professionals mainly use general image search engines and official cultural web sites but 45% of the interviewed experts use proprietary databases (these users are typically those who search for data to be used in applications, the content brokers).</p> <p>The majority of searches are for illustrative images (to improve a text) and photographs of artworks and historic photography or illustrative images.</p>																
<b>E4</b>	The expert users interviewed rarely use video searches.																
<b>E5</b>	<p>CH professionals identify the following features as very useful to improve their searches:</p> <table border="1"> <thead> <tr> <th colspan="2">At query time</th></tr> </thead> <tbody> <tr> <td>most desired general search facilities</td><td> <ul style="list-style-type: none"> <li>• Relevance feedback (explanation: “more like this image”, “images with this shape”) (≈90%)</li> <li>• Spell checking (explanation: “did you mean..?”) (≈95%)</li> <li>• Search suggestions displayed as you type in the query (≈100%)</li> <li>• Search suggestions displayed along with search results (≈95%)</li> <li>• Filter by language (≈90%)</li> <li>• Filter by file type (html, PDF, jpg, gif, avi, mpeg) (≈85%)</li> <li>• Filter by date (≈100%)</li> <li>• Filter by usage rights (≈80%)</li> <li>• Boolean search (explanation: exclude these words, search documents including any of these words, including all of these words) (≈95%)</li> <li>• Use of advanced fixed fields (explanation: fields specifically oriented to query for authors, artists, buildings) Similar to query a database (≈95%)</li> <li>• Use of authority lists (explanation: lists of authors, artists, buildings) Similar to obtain a complete field database view (≈85%)</li> <li>• </li> </ul> </td></tr> <tr> <td>most desired image and video specific search facilities</td><td> <ul style="list-style-type: none"> <li>• Visual relevance feedback (explanation: “more like this image”, “images with this shape”) (≈60%)</li> <li>• Filter by image size (≈65%)</li> <li>• Filter by coloration (≈50%)</li> </ul> </td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">When obtaining and displaying results</th></tr> </thead> <tbody> <tr> <td>Source related</td><td> <ul style="list-style-type: none"> <li>• Aggregated results</li> </ul> </td></tr> <tr> <td>Sorting related</td><td> <ul style="list-style-type: none"> <li>• sort by pertinence</li> <li>• Sort by popularity</li> <li>• Sort by date   title   other criteria</li> </ul> </td></tr> <tr> <td>Browsing related</td><td> <ul style="list-style-type: none"> <li>• Search results adjusted to your profile</li> <li>• Browsing by categories or web directories</li> <li>• Faceted search/browse</li> </ul> </td></tr> <tr> <td>Clustering related</td><td> <ul style="list-style-type: none"> <li>• Results clustered by topic</li> <li>• Results clustered by image/video description</li> <li>• Cluster results by topic</li> </ul> </td></tr> </tbody> </table>	At query time		most desired general search facilities	<ul style="list-style-type: none"> <li>• Relevance feedback (explanation: “more like this image”, “images with this shape”) (≈90%)</li> <li>• Spell checking (explanation: “did you mean..?”) (≈95%)</li> <li>• Search suggestions displayed as you type in the query (≈100%)</li> <li>• Search suggestions displayed along with search results (≈95%)</li> <li>• Filter by language (≈90%)</li> <li>• Filter by file type (html, PDF, jpg, gif, avi, mpeg) (≈85%)</li> <li>• Filter by date (≈100%)</li> <li>• Filter by usage rights (≈80%)</li> <li>• Boolean search (explanation: exclude these words, search documents including any of these words, including all of these words) (≈95%)</li> <li>• Use of advanced fixed fields (explanation: fields specifically oriented to query for authors, artists, buildings) Similar to query a database (≈95%)</li> <li>• Use of authority lists (explanation: lists of authors, artists, buildings) Similar to obtain a complete field database view (≈85%)</li> <li>• </li> </ul>	most desired image and video specific search facilities	<ul style="list-style-type: none"> <li>• Visual relevance feedback (explanation: “more like this image”, “images with this shape”) (≈60%)</li> <li>• Filter by image size (≈65%)</li> <li>• Filter by coloration (≈50%)</li> </ul>	When obtaining and displaying results		Source related	<ul style="list-style-type: none"> <li>• Aggregated results</li> </ul>	Sorting related	<ul style="list-style-type: none"> <li>• sort by pertinence</li> <li>• Sort by popularity</li> <li>• Sort by date   title   other criteria</li> </ul>	Browsing related	<ul style="list-style-type: none"> <li>• Search results adjusted to your profile</li> <li>• Browsing by categories or web directories</li> <li>• Faceted search/browse</li> </ul>	Clustering related	<ul style="list-style-type: none"> <li>• Results clustered by topic</li> <li>• Results clustered by image/video description</li> <li>• Cluster results by topic</li> </ul>
At query time																	
most desired general search facilities	<ul style="list-style-type: none"> <li>• Relevance feedback (explanation: “more like this image”, “images with this shape”) (≈90%)</li> <li>• Spell checking (explanation: “did you mean..?”) (≈95%)</li> <li>• Search suggestions displayed as you type in the query (≈100%)</li> <li>• Search suggestions displayed along with search results (≈95%)</li> <li>• Filter by language (≈90%)</li> <li>• Filter by file type (html, PDF, jpg, gif, avi, mpeg) (≈85%)</li> <li>• Filter by date (≈100%)</li> <li>• Filter by usage rights (≈80%)</li> <li>• Boolean search (explanation: exclude these words, search documents including any of these words, including all of these words) (≈95%)</li> <li>• Use of advanced fixed fields (explanation: fields specifically oriented to query for authors, artists, buildings) Similar to query a database (≈95%)</li> <li>• Use of authority lists (explanation: lists of authors, artists, buildings) Similar to obtain a complete field database view (≈85%)</li> <li>• </li> </ul>																
most desired image and video specific search facilities	<ul style="list-style-type: none"> <li>• Visual relevance feedback (explanation: “more like this image”, “images with this shape”) (≈60%)</li> <li>• Filter by image size (≈65%)</li> <li>• Filter by coloration (≈50%)</li> </ul>																
When obtaining and displaying results																	
Source related	<ul style="list-style-type: none"> <li>• Aggregated results</li> </ul>																
Sorting related	<ul style="list-style-type: none"> <li>• sort by pertinence</li> <li>• Sort by popularity</li> <li>• Sort by date   title   other criteria</li> </ul>																
Browsing related	<ul style="list-style-type: none"> <li>• Search results adjusted to your profile</li> <li>• Browsing by categories or web directories</li> <li>• Faceted search/browse</li> </ul>																
Clustering related	<ul style="list-style-type: none"> <li>• Results clustered by topic</li> <li>• Results clustered by image/video description</li> <li>• Cluster results by topic</li> </ul>																



	Visualization related	<ul style="list-style-type: none"> <li>Graphical visualization of clustered results</li> <li>Timeline</li> </ul>
	<b>When handling results</b>	
	Action related	<ul style="list-style-type: none"> <li>Results annotation</li> <li>Collaborative search</li> </ul>
	History related	<ul style="list-style-type: none"> <li>Search history</li> <li>Basket</li> </ul>

## Tourism Group Results

MultiMatch partners interviewed 16 different experts from the Cultural Heritage Tourism area. The results obtained were the following.

### E1. User Profile

Question	Comments	Figure
EQ 1.1	The most common age range of the expert respondents was between 18-45 (49%)	53
EQ 1.2, 1.3	All experts were identified as being part of the tourism group. We included: <ul style="list-style-type: none"> <li>• Employees of travel agencies</li> <li>• Travellers</li> <li>• Publishers</li> </ul>	
EQ 1.4	70% of experts considered essential or useful (49% vs 30%) the internet in their daily work	54
EQ 1.5	Most popular sources of cultural information (considered essential or useful) were the following: <ol style="list-style-type: none"> <li>1. Regular visits to specialized web sites (<math>\approx 90\%</math>)</li> <li>2. Use of databases and catalogs (<math>\approx 90\%</math>)</li> <li>3. Specialized search services (<math>\approx 80\%</math>)</li> <li>4. General search services (<math>\approx 70\%</math>)</li> </ol> <p>While RSS feeds from specialized sites and newsletters subscriptions were not as useful (<math>&lt;40\%</math>)</p>	55
EQ 1.6	Cultural information search frequency was as follows: <ul style="list-style-type: none"> <li>• Cultural text searches are the most common, mainly performed daily or weekly by experts (<math>\approx 70\%</math>)</li> <li>• Image and video searches are performed occasionally, video searches being less frequent than image searches (30% vs 10% for daily and weekly searches respectively)</li> </ul>	56
EQ 1.7	<ul style="list-style-type: none"> <li>• Google was the preferred search engine.</li> </ul>	none
EQ 1.8, 1.9	Although most experts interviewed were native Spanish or Dutch speakers, results of question 1.9 show that English is the most useful language to get updated and comprehensive cultural information from web sources.	57

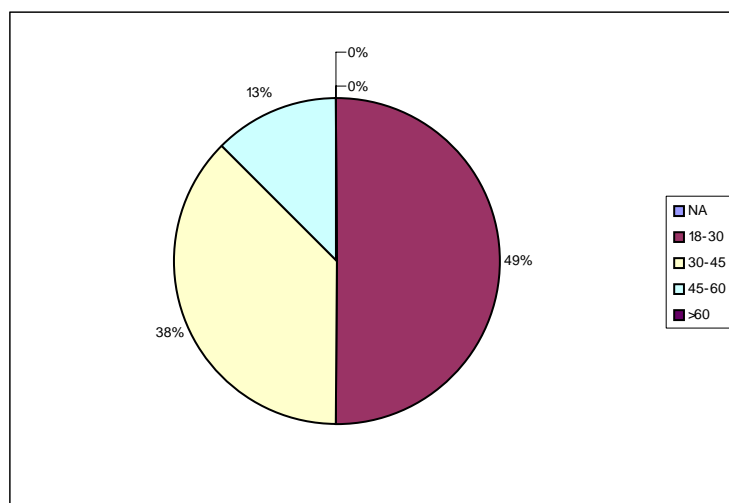
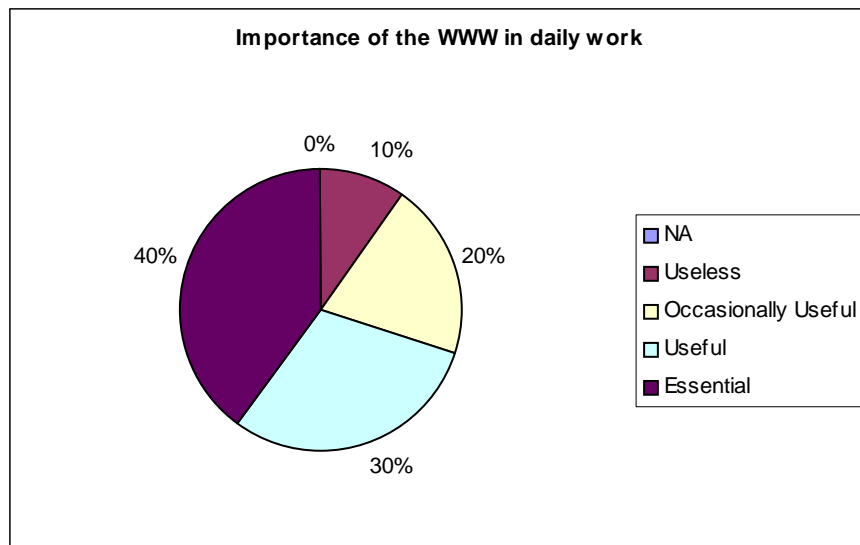
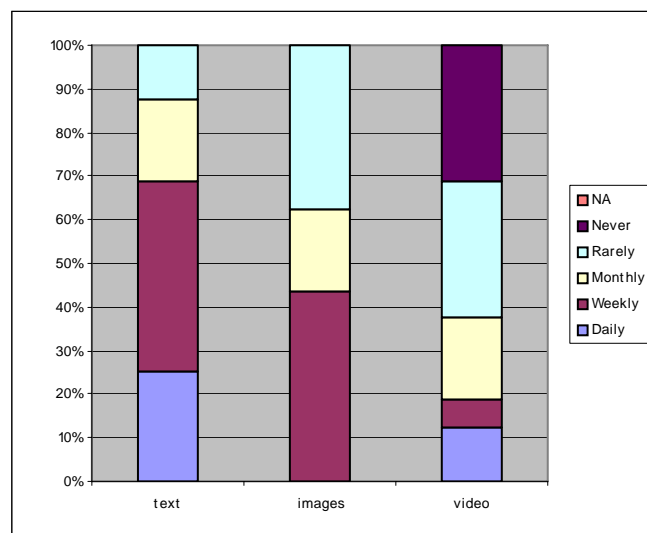


Figure 53: respondents age distribution (Tourism Group)



**Figure 54: Importance of the WWW in daily work (Tourism Group)**



**Figure 55: Text, images and video search frequencies (Tourism Group)**

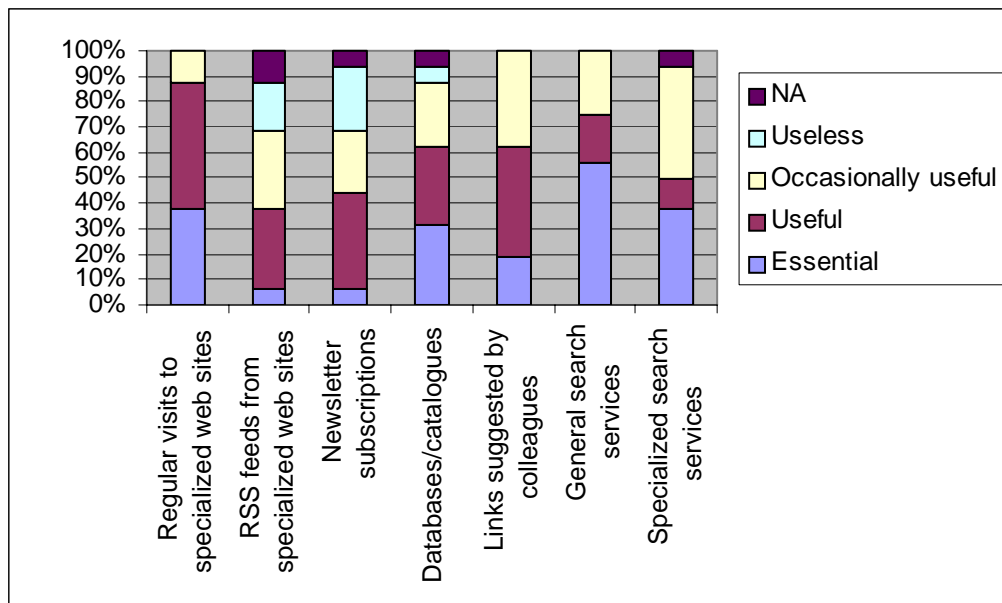


Figure 56: Main sources of cultural information (Tourism Group)

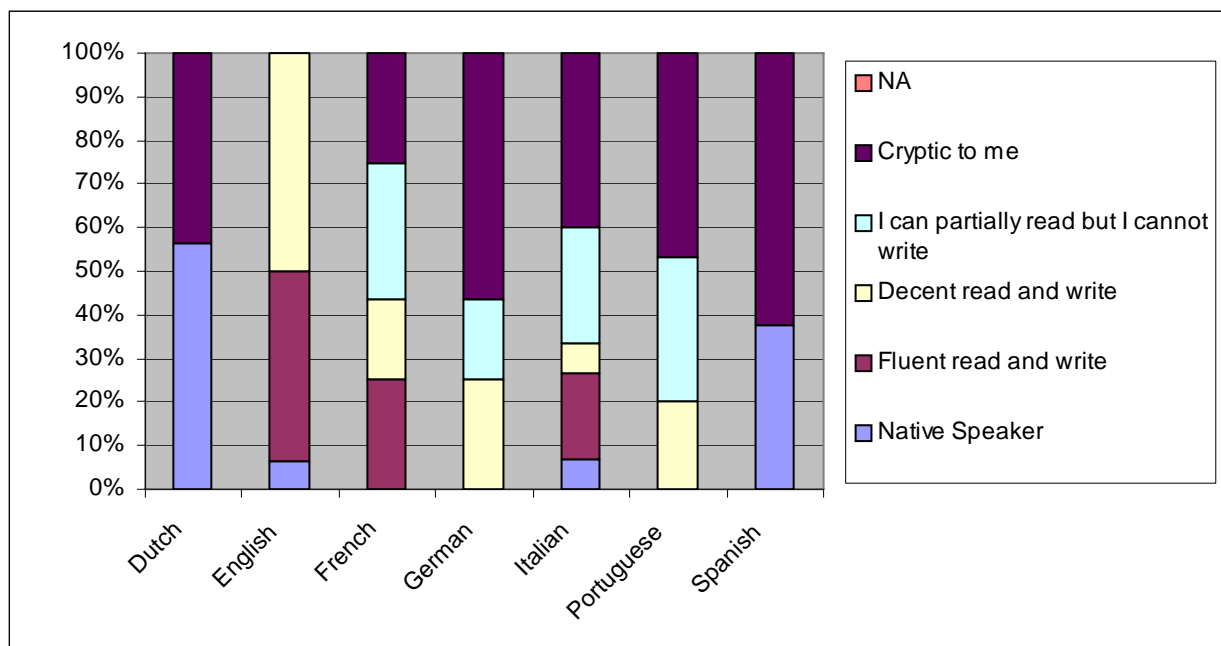


Figure 57: Experts language skills (Tourism Group)

## E2. Textual Search

Question	Comments	Figure
<b>EQ 2.1</b>	<p>The main sources when searching cultural information text were the following (for daily, weekly or monthly use):</p> <ol style="list-style-type: none"> <li>1. General web search engines (=90%)</li> <li>2. News services (60%)</li> <li>3. Official web sites of museums or universities (<math>\approx 60\%</math>)</li> <li>4. Library catalogs and other textual databases (<math>\approx 50\%</math>)</li> <li>5. Encyclopaedic sources (<math>\approx 40\%</math>)</li> <li>6. Book search engines (<math>\approx 10\%</math>)</li> </ol>	<b>58</b>
<b>EQ 2.2</b>	<p>The most useful services were:</p> <ol style="list-style-type: none"> <li>1. Google (90%)</li> <li>2. Wikipedia (10%)</li> </ol>	<b>none</b>
<b>EQ 2.3, 2.4</b>	<p>All the text searches proposed in the questionnaire were accepted as useful in daily, weekly or monthly work. However, results can be ranked accordingly the following list:</p> <ol style="list-style-type: none"> <li>1. Search for specific authors (<math>\approx 80\%</math>)</li> <li>2. Search for specific web sites (<math>\approx 70\%</math>)</li> <li>3. Search for scholar publications (<math>\approx 60\%</math>)</li> <li>4. Search for cultural events (<math>\approx 50\%</math>)</li> <li>5. Search for books (<math>\approx 50\%</math>)</li> <li>6. Search for news (<math>\approx 50\%</math>)</li> <li>7. Search for an art period (<math>\approx 40\%</math>)</li> </ol> <p>Text searched proposed by experts in tourism were the following:</p> <ul style="list-style-type: none"> <li>• Thematic searches</li> <li>• Search for citations</li> <li>• Search in databases for specific words or texts related with one or more classic authors</li> </ul>	<b>59</b>
<b>EQ 2.5, 2.6</b>	<p>With respect to information compilation and aggregation expert users in the tourism domain considered essential or normally useful the following compilations:</p> <ol style="list-style-type: none"> <li>1. Announcements (<math>\approx 80\%</math>)</li> <li>2. Compiled information about authors (<math>\approx 20\%</math>)</li> <li>3. Compiled information about cultural sites (<math>\approx 65\%</math>)</li> </ol> <p>The other options presented were also considered normally useful but the percentage was always under 50%.</p>	<b>60</b>
<b>EQ 2.7, 2.8</b>	<p>With respect to information extraction and pattern discovering, expert users in the tourism domain considered essential or normally useful the following relations:</p> <ol style="list-style-type: none"> <li>1. Ideas in common between two or more authors (<math>\approx 90\%</math>)</li> <li>2. Artists/masterpieces/cultural sites most closely associated with a city (<math>\approx 70\%</math>)</li> <li>3. Masterpieces most closely associated with an author (<math>\approx 70\%</math>)</li> <li>4. Artists/masterpieces/CH sites most closely related to a given keyword (<math>\approx 60\%</math>)</li> <li>5. Places/people/topics most closely associated with an artist (<math>\approx 50\%</math>)</li> <li>6. Authors most closely associated with a given author (<math>\approx 30\%</math>)</li> </ol> <p>Options related with authors popularity and opinions were not considered as useful (<math>&lt; 50\%</math>)</p>	<b>18</b>

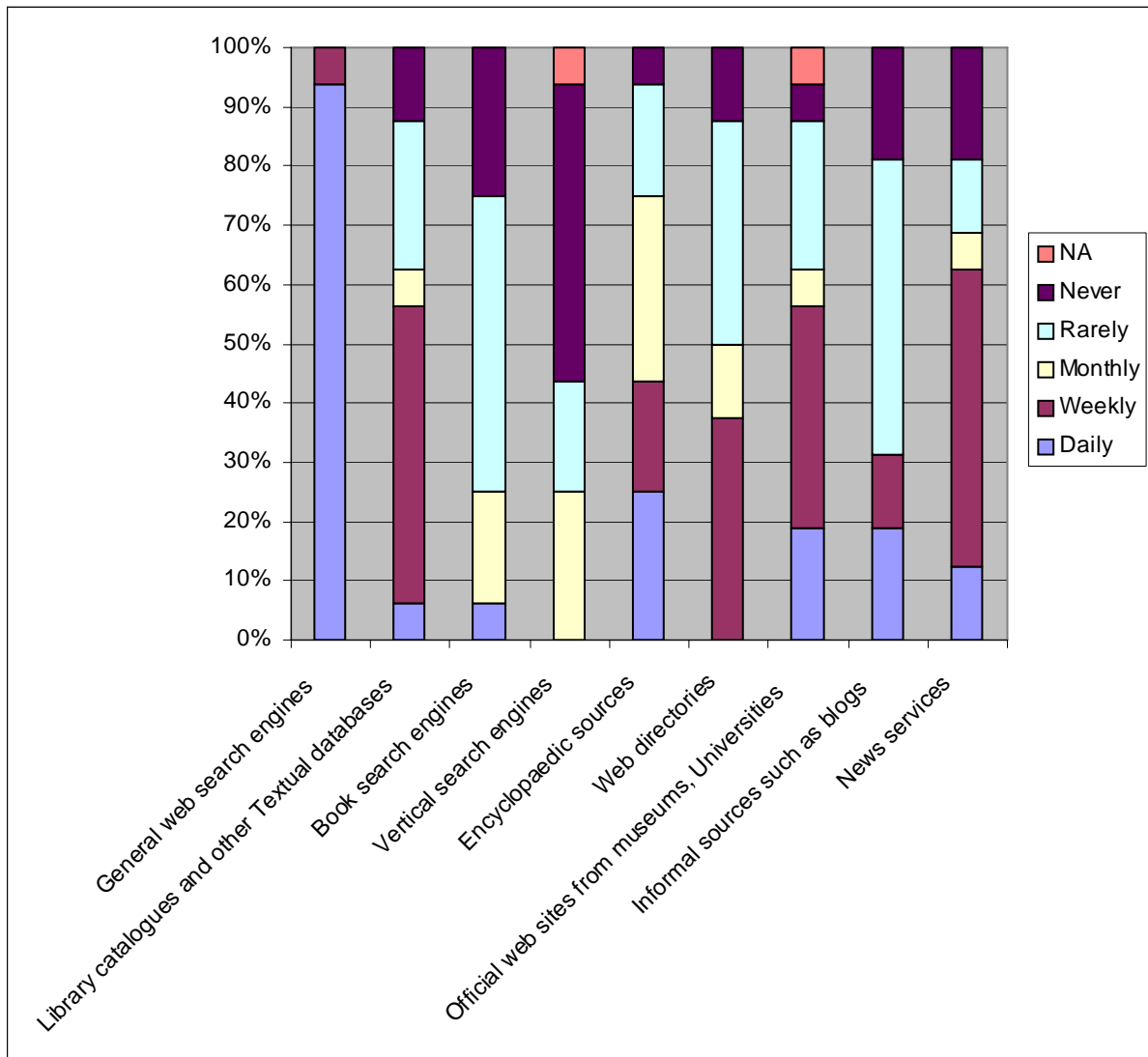


Figure 58: Primary sources for textual cultural searches (Tourism Group)

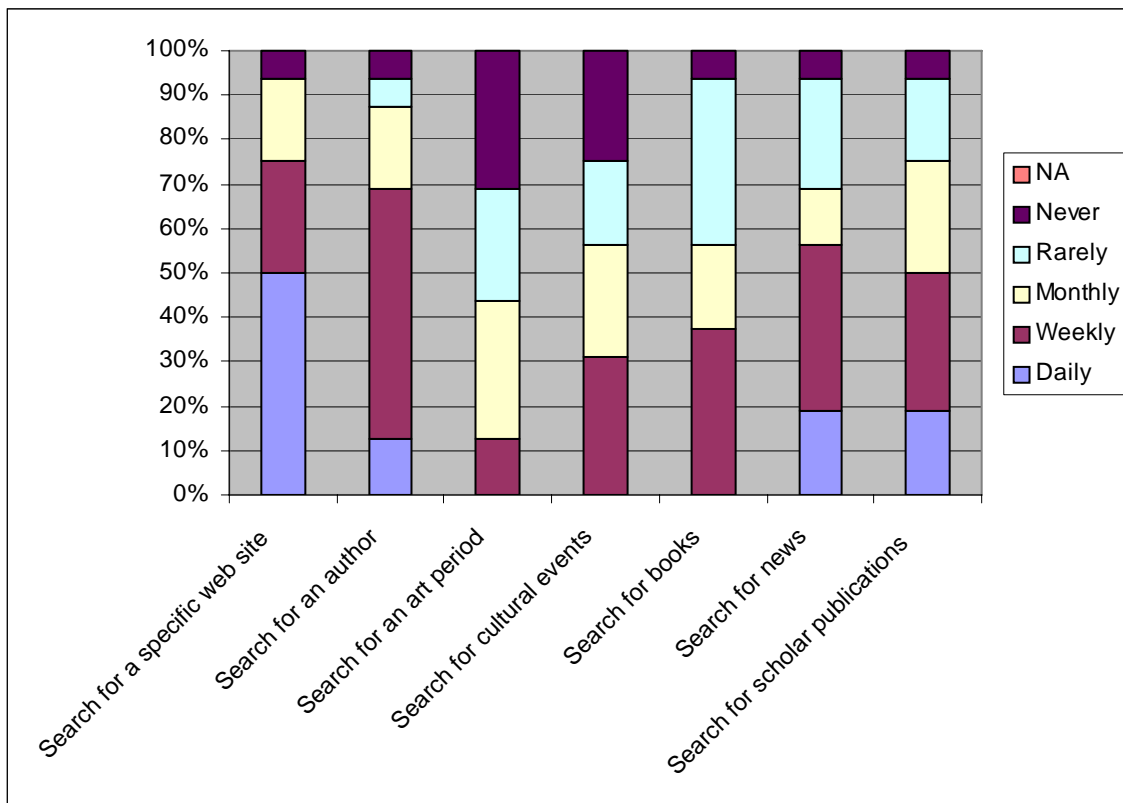


Figure 59: Common types of textual search (Tourism Group)

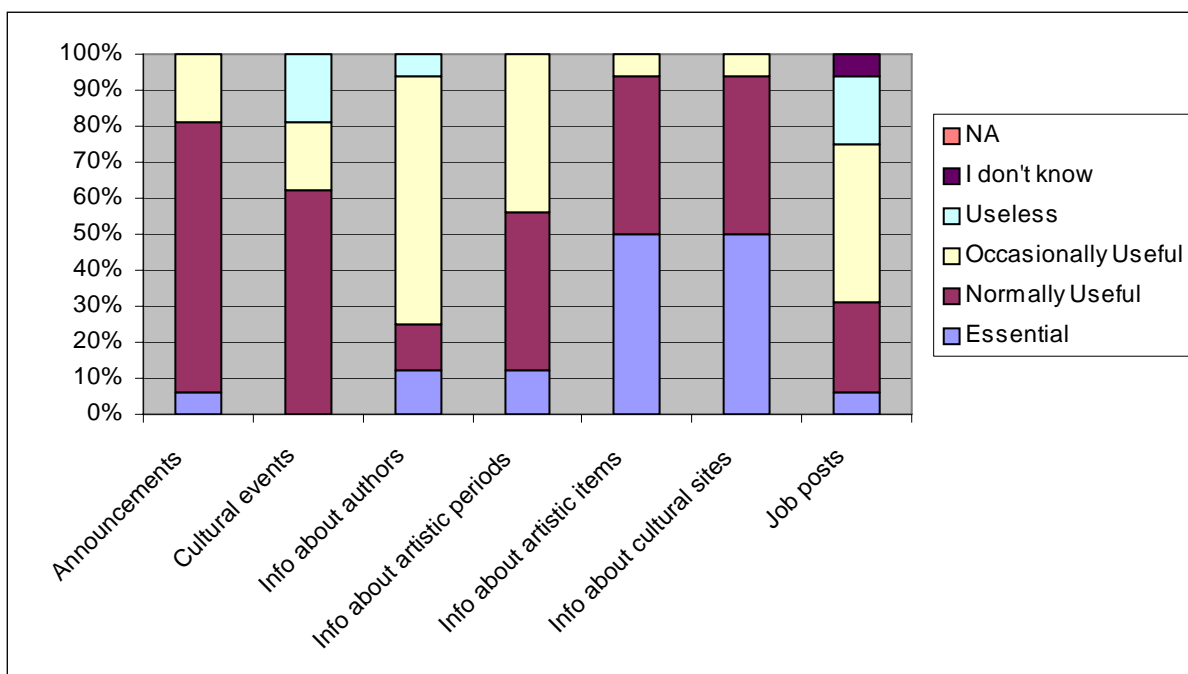


Figure 60: CH Information compilation usefulness (Tourism Group)



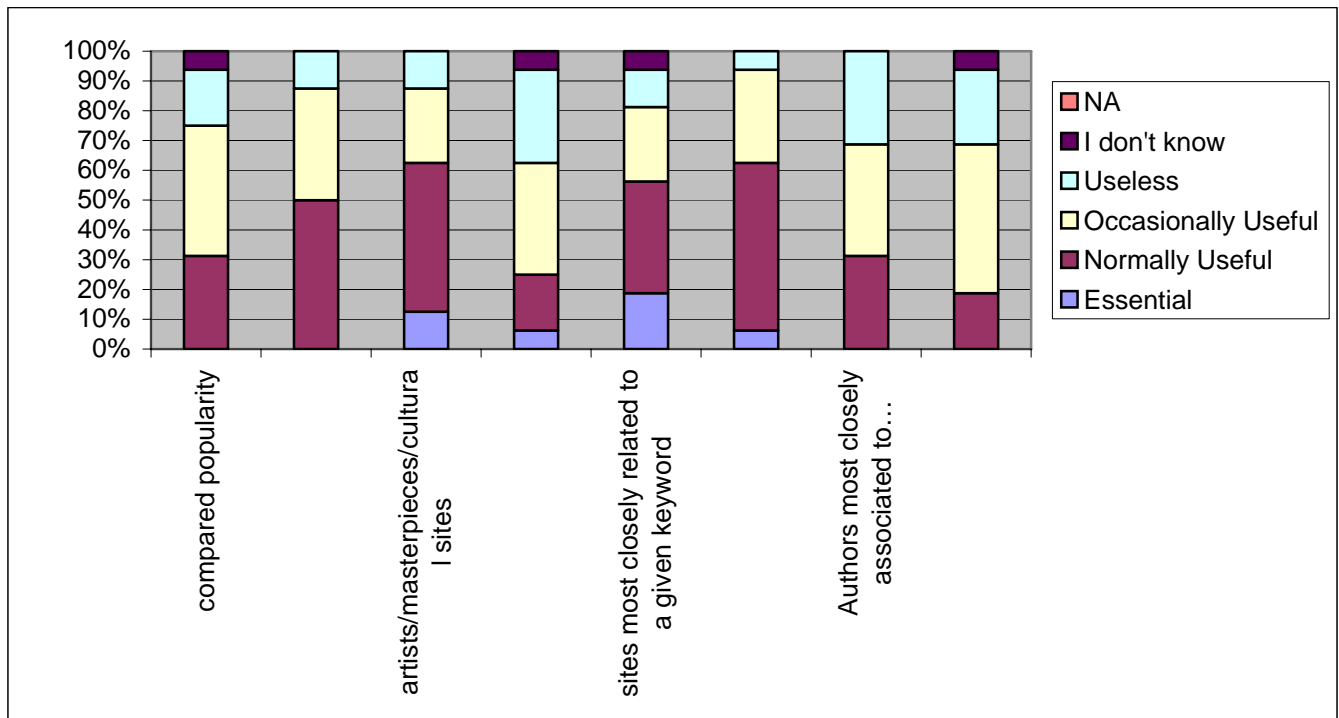
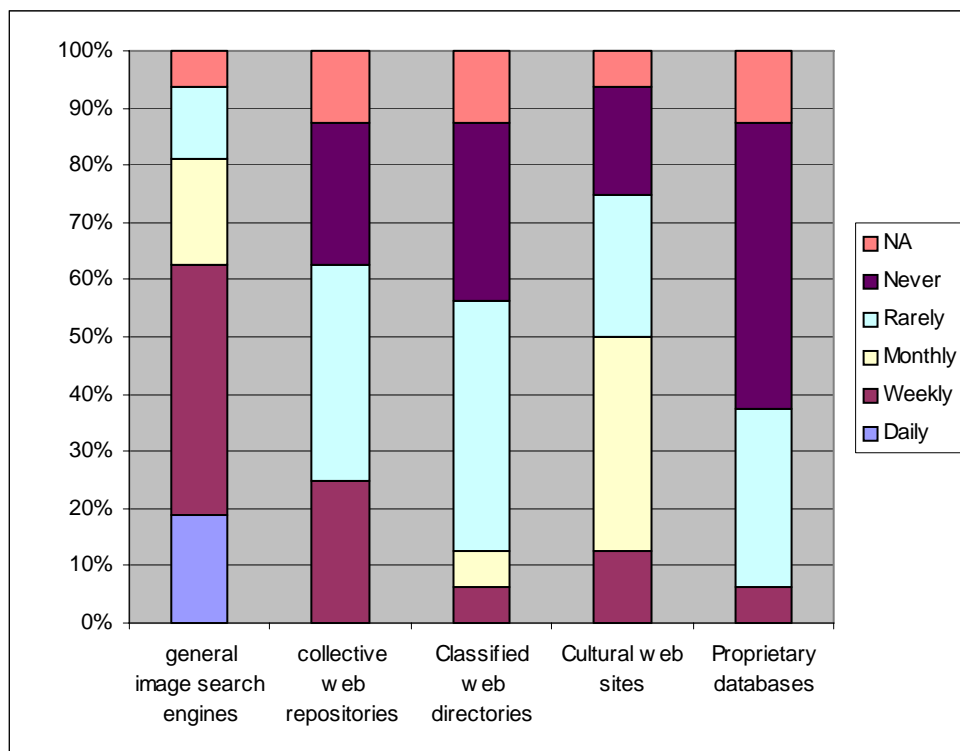


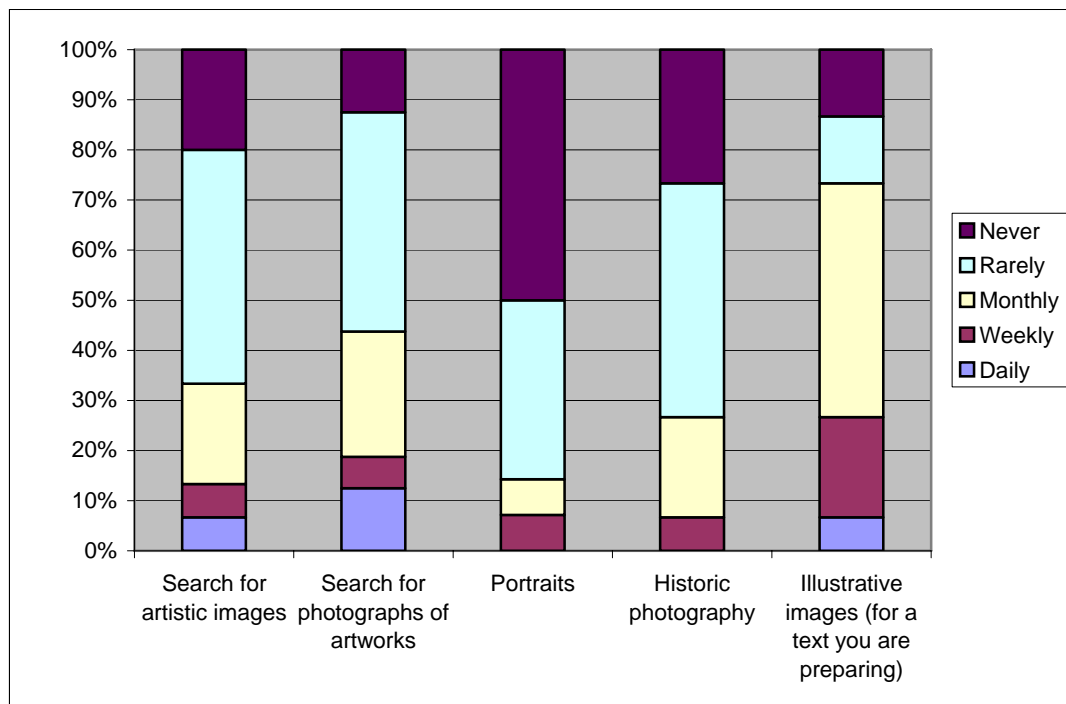
Figure 61: CH information mining usefulness (Tourism Group)

### E3. Image Search

Question	Comments	Figure
<b>EQ 3.1</b>	<p>The primary sources stated by expert users in tourism when searching for images were the following (for daily, weekly and monthly use):</p> <ol style="list-style-type: none"> <li>1. General image search engines (<math>\approx 60\%</math>)</li> </ol> <p>Proprietary databases, classified web directories, Cultural web sites and collective web repositories are also used but with a lower frequency</p>	<b>62</b>
<b>EQ 3.2</b>	The most useful services stated by these users were mainly focused in the use of general image search engines such as Google	<b>none</b>
<b>EQ 3.3, 3.4</b>	<p>The most common types of image searches were the following:</p> <ol style="list-style-type: none"> <li>6. Search for illustrative images (<math>\approx 25\%</math>)</li> <li>7. Search for photographs of artworks (<math>\approx 20\%</math>)</li> <li>8. Search for artistic images (<math>\approx 20\%</math>)</li> </ol> <p>The other options were also considered useful but the expert user but with a lower percentage.</p> <p>Types of image search proposed by expert users were the following:</p> <ul style="list-style-type: none"> <li>• Search for images of hotel rooms</li> <li>• Search for maps</li> <li>• Search for images of artists</li> </ul>	<b>63</b>



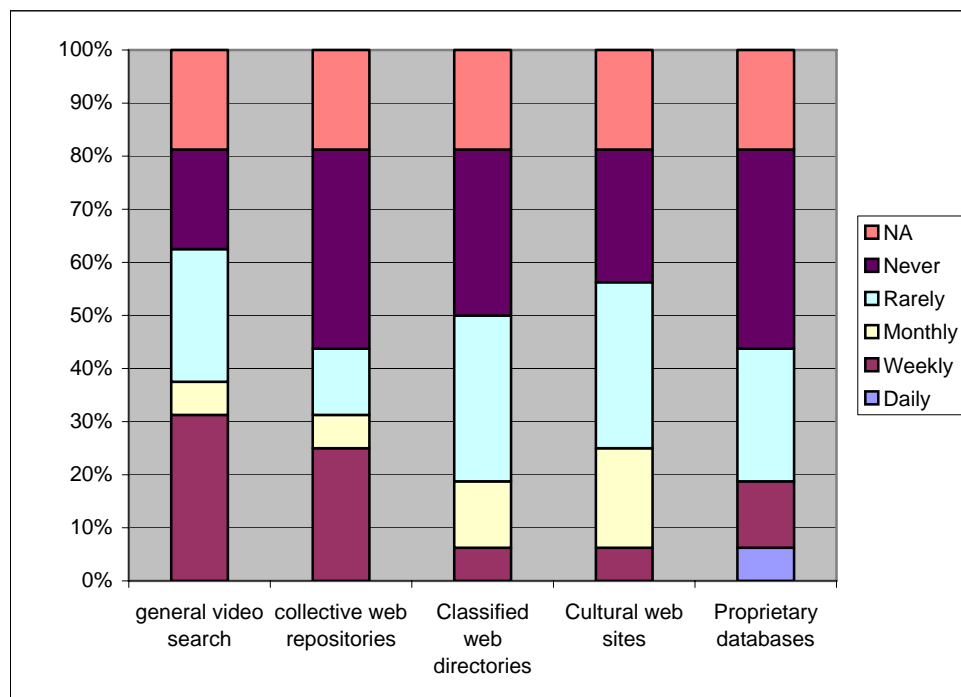
**Figure 62: Primary sources for cultural image search (Tourism Group)**



**Figure 63: Most common types of cultural image search (Tourism Group)**

## E4. Video Search

Question	Comments	Figure
<b>EQ 4.1</b>	The primary sources stated by expert users when searching for videos were the following (for daily, weekly and monthly use): 1. General video search engines ( $\approx 30\%$ )  Proprietary databases, cultural web sites, classified web directories and collective web repositories are also used but with a very low frequency.	<b>64</b>
<b>EQ 4.2</b>	The most useful services stated by expert users were mainly focused in the use of general video search engines such as Google	<b>None</b>
<b>EQ 4.3, 4.4</b>	The most common types of video searches were the following (for daily, weekly and monthly use): 1. Search for documentary materials ( $\approx 35\%$ )  The other options were ranked with a percentage lower than 35 percent ( $<50\%$ )  No relevant feedback by the user was given to enlarge the types of video searches	<b>65</b>



**Figure 64: Primary sources for cultural video search (Tourism Group)**

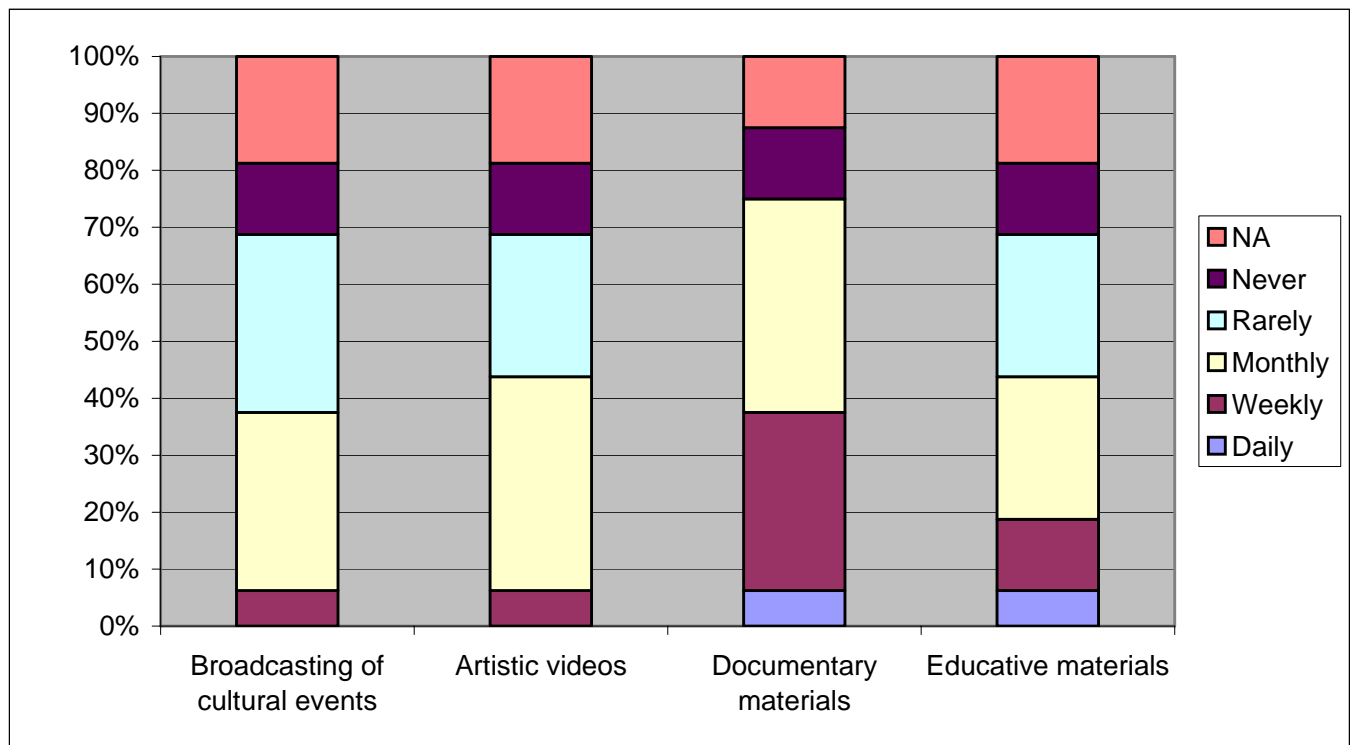


Figure 65: Most common types of cultural video search (Tourism Group)

## E5. Desirable Search Facilities

Question	Comments	Figure
<b>EQ 5.1. At query time</b>	<p>The most desired general search facilities were the following (considered as essential or normally useful by the expert in the tourism domain):</p> <ol style="list-style-type: none"> <li>1. Boolean search (<math>\approx 60\%</math>)</li> <li>2. Spell checking (<math>\approx 60\%</math>)</li> <li>3. Use of advanced fixed fields (<math>\approx 50\%</math>)</li> <li>4. Relevance feedback (<math>\approx 50\%</math>)</li> <li>5. Filtering by language (<math>\approx 45\%</math>)</li> <li>6. Filtering by date (<math>\approx 30\%</math>)</li> <li>7. Use of authority list (<math>\approx 20\%</math>)</li> </ol> <p>Other options such as search suggestions while typing the query or retrieving results, and filtering by type or usage rights were considered useful by very few people</p> <p>The most desired image and video specific search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Visual relevance feedback (<math>\approx 70\%</math>)</li> </ol> <p>Visual filtering (such as by coloration, image size or length) was also considered useful but with low frequency (<math>&lt; 30\%</math>)</p>	<b>66</b>
<b>EQ 5.1 (when obtaining and displaying results)</b>	<p>The most desired general search facilities were the following (considered as essential or normally useful by the expert in the tourism domain):</p> <ol style="list-style-type: none"> <li>1. Sort by date, title or any other criteria (<math>\approx 80\%</math>)</li> <li>2. Clustering of the results by topic (<math>\approx 75\%</math>)</li> <li>3. Faceted search and browsing (<math>\approx 70\%</math>)</li> <li>4. Graphical visualization of clustered results (<math>\approx 65\%</math>)</li> <li>5. Timeline (<math>\approx 60\%</math>)</li> <li>6. Sort by pertinence (<math>\approx 55\%</math>)</li> <li>7. Aggregated results (<math>\approx 55\%</math>)</li> <li>8. Browsing by categories or web directories (<math>\approx 50\%</math>)</li> </ol> <p>Options related with user profile and search customisation and sorting by popularity were considered useful but with a very low frequency (<math>&lt; 40\%</math>)</p> <p>The most desired image and video specific search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Search for video fragments</li> <li>2. Cluster results by description</li> </ol> <p>The other options proposed (such as different kind of searches and clustering approaches accordingly the contents' visual properties) were considered useful but with lower frequencies (<math>&lt; 50\%</math>).</p>	<b>67, 68</b>
<b>EQ 5.1 (when handling results)</b>	<p>The most desired general search facilities were the following (considered as essential or normally useful by the expert):</p> <ol style="list-style-type: none"> <li>1. Annotation (<math>\approx 60\%</math>)</li> <li>2. Search history (<math>\approx 75\%</math>)</li> <li>3. Basket (<math>\approx 60\%</math>)</li> </ol> <p>The collaborative search option was considered useful but with a 50%</p>	<b>69</b>

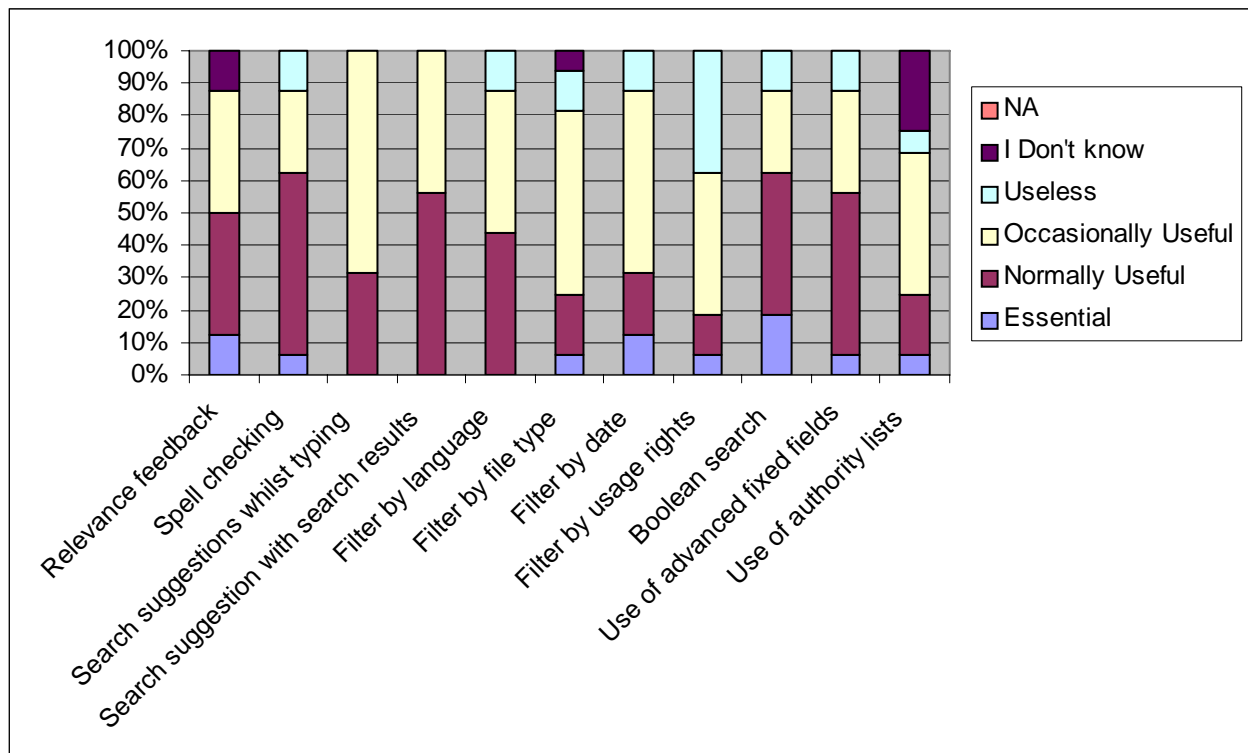


Figure 66: General search facilities at query time (Tourism Group)

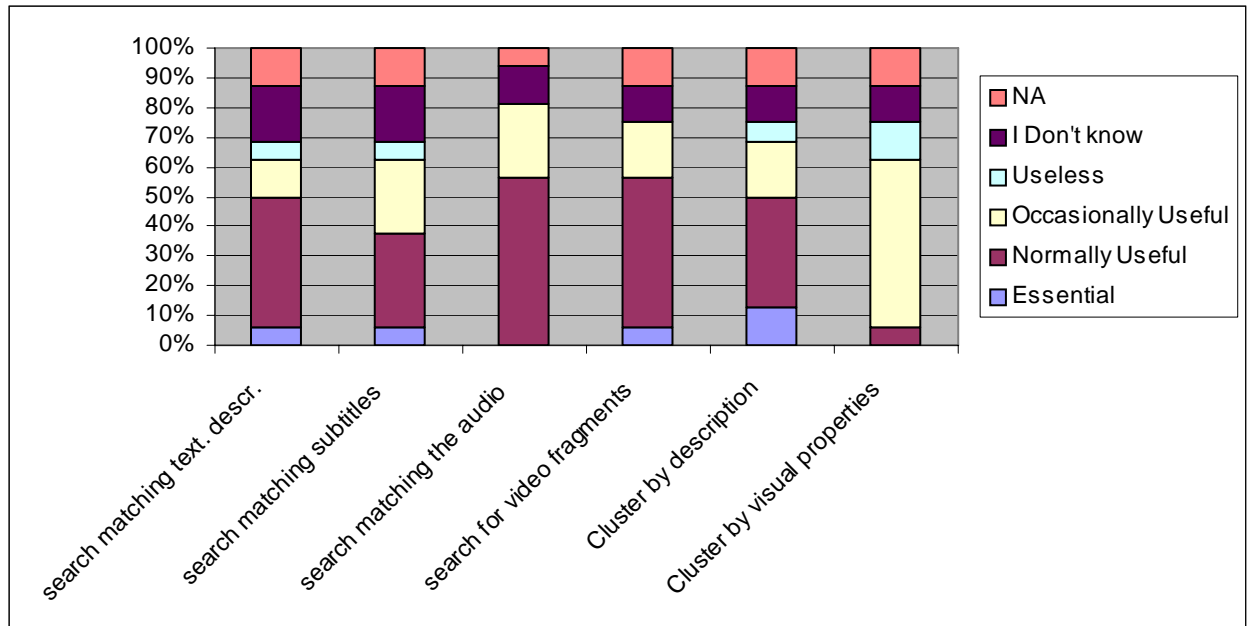
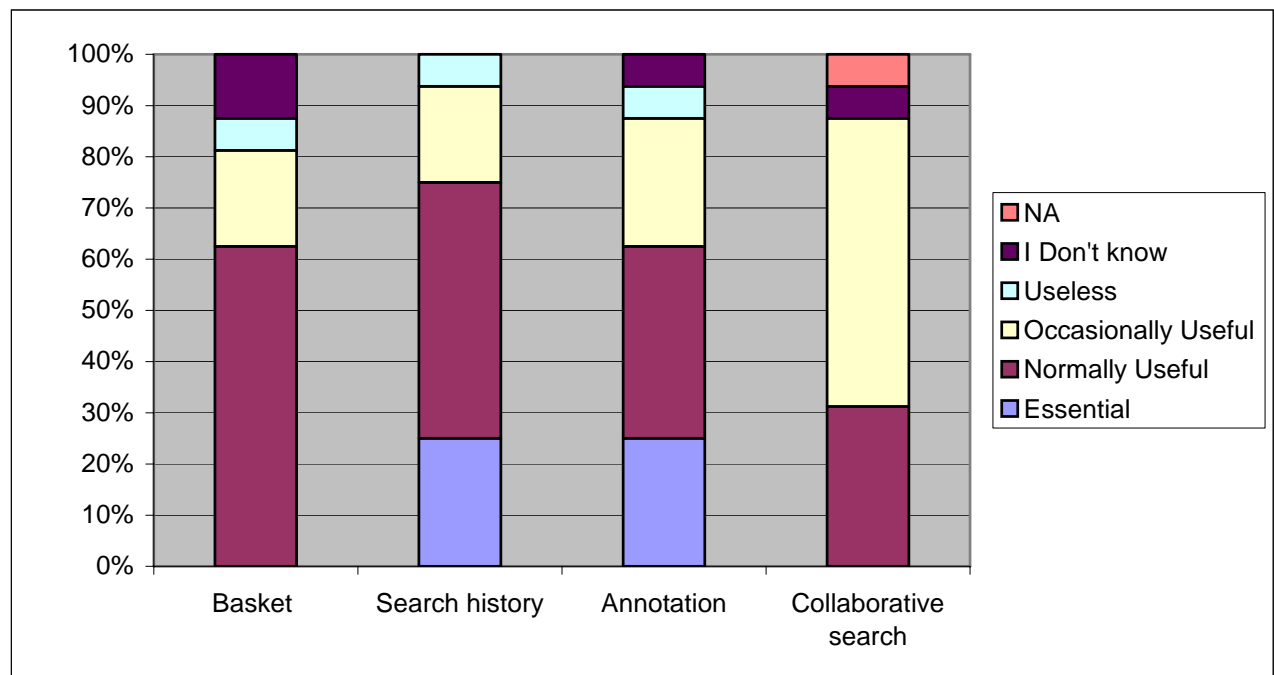


Figure 67: Specific image and video search facilities when obtaining and displaying results (Tourism Group)





**Figure 68: General search facilities when obtaining and displaying results (Tourism Group)**

## Tourism Group Conclusions

<b>E1</b>	<ul style="list-style-type: none"> <li>The Tourism group consists of tour operators, leisure travellers and publishers of tourist guides. They use the internet often. Some Internet specific services such as newsletters or RSS feeds are not very widely used amongst interviewed experts.</li> <li>Most of the respondents do not use the internet extensively, and do not make use of the latest features. They are definitely not early adopters. The starting point for their searches is almost exclusively Google.</li> <li>Usual searches are text searches but also image and video. However these last are currently not as frequent. The questionnaire did not capture these expectations over longer periods.</li> <li>The most useful languages were English and the native language of the respondents. These results justify multilingualism in MM. Note: almost all interviewed experts were either Dutch or Spanish native speakers.</li> </ul>
<b>E2</b>	<ul style="list-style-type: none"> <li>Experts in the Tourism Group mainly use general search engines and official sites of museums for textual information. Other options (library catalogues and textual databases, book search engines, encyclopaedic sources and news services) are not used widely. Specific services such as web directories, blogs or vertical search engines are not very well known. Only Google and Wikipedia are mentioned.</li> <li>The most common text searches look for specific cultural web sites and cultural events; i.e. the places where tourist information can be gathered. This was very much in line with expectations.</li> <li>With respect to information compilation issues, expert users focused their preferred compilations on artistic items and –less- on cultural sites. This last one from the tourism point of view (e.g. information about near cultural sites, hotels, transports).</li> <li>With respect to information mining and pattern discovering, expert users focused their preferred associations with authors (70%) and cities (65%). So, tourists would like to find information about cities.</li> </ul>
<b>E3</b>	<ul style="list-style-type: none"> <li>Experts in the tourism group mainly use general image search engines. Again, Google is mentioned as a primary source.</li> <li>The majority of searches are focused on illustrative images (to improve a text) and</li> </ul>

	photographs of artworks.																						
<b>E4</b>	<ul style="list-style-type: none"> <li>Experts in the tourism group mainly use general image search engines and official cultural web sites for video searches. Specific and general web repositories and proprietary databases are also used but with very low frequency.</li> <li>The majority of searches are focused on documentary materials.</li> </ul>																						
<b>E5</b>	<ul style="list-style-type: none"> <li>Experts in the Tourist user group have identified the following features as very useful to improve their searches:</li> </ul> <table border="1"> <thead> <tr> <th colspan="2">At query time</th></tr> </thead> <tbody> <tr> <td>Query building related</td><td> <ul style="list-style-type: none"> <li>Boolean search (they are familiar with this as it is very common in almost all general search engines)</li> <li>Other options:(i.e. use of authority lists, Relevance feedback, Visual relevance feedback) are not used. Most probably because this group is not familiar with using these features.</li> </ul> </td></tr> <tr> <td>Filtering related</td><td> <ul style="list-style-type: none"> <li>Spell checking</li> </ul> </td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">When obtaining and displaying results</th></tr> </thead> <tbody> <tr> <td>Sorting related</td><td> <ul style="list-style-type: none"> <li>Common sorting options such as date or title sorting</li> <li>More complex sorting options such as relevance ranking is not required. Again, probably because this group is not familiar with using these features</li> </ul> </td></tr> <tr> <td>Browsing related</td><td> <ul style="list-style-type: none"> <li>The advanced options mentioned: (i.e., Browsing by categories or web directories, Search for video fragments) were not requested often by this group.</li> </ul> </td></tr> <tr> <td>Clustering related</td><td> <ul style="list-style-type: none"> <li>Results clustered by topic (55%).</li> </ul> </td></tr> <tr> <td>Visualization related</td><td> <ul style="list-style-type: none"> <li>Graphical visualization of clustered results was valued highly.</li> <li>Also searches for related multimedia (audio, video) were valued highly.</li> <li>Timeline view was only valued as important by 20%</li> </ul> </td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">When handling results</th></tr> </thead> <tbody> <tr> <td>Action related</td><td> <ul style="list-style-type: none"> <li>Results annotation</li> </ul> </td></tr> <tr> <td>History related</td><td> <ul style="list-style-type: none"> <li>Search history</li> <li>Basket</li> </ul> </td></tr> </tbody> </table>	At query time		Query building related	<ul style="list-style-type: none"> <li>Boolean search (they are familiar with this as it is very common in almost all general search engines)</li> <li>Other options:(i.e. use of authority lists, Relevance feedback, Visual relevance feedback) are not used. Most probably because this group is not familiar with using these features.</li> </ul>	Filtering related	<ul style="list-style-type: none"> <li>Spell checking</li> </ul>	When obtaining and displaying results		Sorting related	<ul style="list-style-type: none"> <li>Common sorting options such as date or title sorting</li> <li>More complex sorting options such as relevance ranking is not required. Again, probably because this group is not familiar with using these features</li> </ul>	Browsing related	<ul style="list-style-type: none"> <li>The advanced options mentioned: (i.e., Browsing by categories or web directories, Search for video fragments) were not requested often by this group.</li> </ul>	Clustering related	<ul style="list-style-type: none"> <li>Results clustered by topic (55%).</li> </ul>	Visualization related	<ul style="list-style-type: none"> <li>Graphical visualization of clustered results was valued highly.</li> <li>Also searches for related multimedia (audio, video) were valued highly.</li> <li>Timeline view was only valued as important by 20%</li> </ul>	When handling results		Action related	<ul style="list-style-type: none"> <li>Results annotation</li> </ul>	History related	<ul style="list-style-type: none"> <li>Search history</li> <li>Basket</li> </ul>
At query time																							
Query building related	<ul style="list-style-type: none"> <li>Boolean search (they are familiar with this as it is very common in almost all general search engines)</li> <li>Other options:(i.e. use of authority lists, Relevance feedback, Visual relevance feedback) are not used. Most probably because this group is not familiar with using these features.</li> </ul>																						
Filtering related	<ul style="list-style-type: none"> <li>Spell checking</li> </ul>																						
When obtaining and displaying results																							
Sorting related	<ul style="list-style-type: none"> <li>Common sorting options such as date or title sorting</li> <li>More complex sorting options such as relevance ranking is not required. Again, probably because this group is not familiar with using these features</li> </ul>																						
Browsing related	<ul style="list-style-type: none"> <li>The advanced options mentioned: (i.e., Browsing by categories or web directories, Search for video fragments) were not requested often by this group.</li> </ul>																						
Clustering related	<ul style="list-style-type: none"> <li>Results clustered by topic (55%).</li> </ul>																						
Visualization related	<ul style="list-style-type: none"> <li>Graphical visualization of clustered results was valued highly.</li> <li>Also searches for related multimedia (audio, video) were valued highly.</li> <li>Timeline view was only valued as important by 20%</li> </ul>																						
When handling results																							
Action related	<ul style="list-style-type: none"> <li>Results annotation</li> </ul>																						
History related	<ul style="list-style-type: none"> <li>Search history</li> <li>Basket</li> </ul>																						

## 7 Past Experience and Query Log Analysis

This section is dedicated to past experience and knowledge of the user groups' needs. Alinari has recently (August 2006) finalized a large interviewing activity with the support of some European universities and other archives. Some aspects of these interviews are strongly related to MultiMatch and consequently we included an extrapolation from the available documentation.

Finally, we also provide an analysis of log data related to WIND and Alinari's user tracking.

### 7.1 Recent Studies of the Educational and CH Professional Groups

Alinari has recently conducted interviews in the educational context for collecting knowledge on the needs of the educational users referring CH contents, search services and user interface.

The method for collecting this information is as follows:

- Observation, group interviews with students on suitability of image use and existing Alinari service for university courses, structured interviews with students about the web portal, structured questionnaire to establish the level of information literacy of students.
- The data collection was conducted during special sessions when students were working on the course tasks in the computer classrooms. Before the sessions started the students completed the questionnaires on computer literacy level. The group interviews lasted from 20 to 45 minutes.
- The interview data were keyed into Excel spreadsheets and analysed using this package. The observation and group interview data records were analysed qualitatively looking for emerging topics related to the assessment questions.

The interview activity was performed during 2006 and finished on September 2006.

Past interviews locations	User groups	Courses attended	ICT Confidence and Skills (average on a scale of 0 to 5)	Main characteristics
University of Borås	23 students	International and intercultural communication	3,85 - The highest confidence is felt in using communication programmes like e-mail (4,54), using the Internet (4,35), in general computer management (4,34), and word processing (4,24). The lowest confidence and skill level concerns usage of database creation software (2,74).	Most of the students (16) fell into the age category from 21 to 25 years. Three were 26-30 and three were more than 30 years old. Six students were in the second year of the studies, ten in the third, four in the fourth year, and two were mature students in the fifth and sixth year of the study.

Past interviews locations	User groups	Courses attended	ICT Confidence and Skills (average on a scale of 0 to 5)	Main characteristics
University of Krakow	140 students	Information sources, Electronic journals, Multimedia, Data formats, Content characteristics of documents	3.95 - High confidence in general using the computer (4.55), for processing texts (4.52), for taking the Internet (4.33) and email (4.29). Less experience in the providing software for creating databases (2.88).	Most of the students (76) fell into the age category from 21 to 25 years. 36 students were in the age of 17-20, 13 students were 26-30 and 15 were more than 30 years old. Students were in the first, second year of the studies, fourth year, and in the fifth year of the study (SUMIK).
University of Leuven	20 students	Notes on Photo Analysis. How to read a photographic image?	All students are trained in using office software and web technologies. They can be considered to be proficient ICT users.	16 students were post-graduate students, in their 3rd to 4th year of study. 8 of them are studying Modern History. 6 were studying for a master after master degree at the Institute for Cultural Studies, this means they are in their 5th year of university studies.
University of Vienna	34 students	Seminar: "Wissensrepräsentation" Knowledge representation	High as a result of the studied programmes	Mainly PhD students but also MA students with a real interest in issues of knowledge representation, eLearning, science communication, information visualization, usability engineering, ontology engineering, pedagogy, history of art and science, theory of art, philosophy of science, information science.
Virtual University of Barcelona	8 Students (virtual community)	Internet Search Course	4,3 in general computer competence (4,6 in communication programmes, 4,4 in using Internet and lowest averages for spreadsheet use and databases)	All students of the Barcelona Virtual University, part of a course of distance learning with the Barcelona Virtual University to complete their public university studies. The 8 participants are enrolled in superior university studies. The average age of the sample is 27 (26.9), with the typical deviation of 6.2, which represents that the sample encompasses ages from 20 to 32.
E-ISOTIS Association	10 students	Intercultural communication	Various usage of assistive technology	people with vision impairments

Past interviews locations	User groups	Courses attended	ICT Confidence and Skills (average on a scale of 0 to 5)	Main characteristics
Alinari	10 students	Seminar on International and intercultural communication	3,9 - The highest confidence is felt in using communication programmes like e-mail (4,5), using the Internet (4,3), in general computer management (4,25), and word processing (4,2). The lowest confidence and skill level concerns usage of database creation software (2,8).	Students from the faculty of architecture and Technology and Management of Cultural Heritage.
ICIMSS International community	12 students	Types of archival documents	Not given	Tests of the image database, search engine, thesaurus, technical compatibility, etc

On the whole the characteristics of the interviewed students were rather similar. Most of them were taking undergraduate and graduate programmes and were between 20-30 years of age. The younger students tended to display higher confidence in ICT use than the older ones. The general level of the ability to use computer technologies (ICT) was on the average quite high. The calculated average is 3.90 (average on a scale of 0 to 5). The lowest confidence and skill levels are noticed in usage of database creation software, and the highest in using e-mail and the Internet. This may be a general trend for the students in humanities of social sciences and humanities in general.

Here we will provide some results of the performed interviews, where the outcome are of interest for MultiMatch user requirements definition.

The area of study of respondents roughly fell into the following categories:

- International and intercultural communication
- Intercultural communication
- Notes on photo analysis
- Electronic journals
- Types of archival documents
- Information sources
- Multimedia,
- Data formats
- Knowledge representation
- Content characteristics of documents
- Internet search course

In general most of the students (about 90%) reported that they could find images useful for their task by using the Alinari “Search” tool (only about 10% gave a negative answer referring to a database of 10.000 CH images). Observation of the students performing tasks showed that the students were

spending rather a long time looking for suitable images and performed several searches before selecting the pictures for their task.

The scope and the contents of the database can be seen from the courses that students named in response to the question “Which courses can benefit most from the service?” The list of courses that according to the students could benefit from CH (archival) database is quite long, and incorporates a variety of disciplines:

Courses using the content	Courses on production of info resources (using the service as an example)
<ul style="list-style-type: none"> <li>• <i>Art and design</i></li> <li>• <i>Audiovisual documents</i></li> <li>• <i>Introduction to information and library science</i></li> <li>• <i>Introduction to painting and art</i></li> <li>• <i>Biology</i></li> <li>• <i>Cultural studies</i></li> <li>• <i>History (of culture, libraries, art, book)</i></li> <li>• <i>Industrial design</i></li> <li>• <i>Intercultural Communication</i></li> <li>• <i>Language</i></li> <li>• <i>Literature</i></li> <li>• <i>Marketing</i></li> <li>• <i>Organization theory</i></li> <li>• <i>Photography</i></li> <li>• <i>Public Relations</i></li> <li>• <i>Social communication</i></li> <li>• <i>Social life documents</i></li> <li>• <i>Sources of information</i></li> <li>• <i>Textile design</i></li> <li>• <i>User studies</i></li> <li>• <i>Industrial Design</i></li> <li>• <i>History of Architecture</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Characteristics of documents content</i></li> <li>• <i>Data formats</i></li> <li>• <i>Databases</i></li> <li>• <i>Digitisation of images</i></li> <li>• <i>Electronic publishing</i></li> <li>• <i>Information retrieval</i></li> <li>• <i>Multimedia</i></li> <li>• <i>Search in the Internet</i></li> <li>• <i>Use of the Internet</i></li> <li>• <i>Virtual libraries</i></li> <li>• <i>Website creation</i></li> </ul>

**Table 69 Courses that could benefit from CH source services (students’ suggestions)**

The students also suggested that a CH search service may be useful for their project and research work, like writing a Master’s thesis. In general the number of suggested courses for using the content shows the potential of the services.

Here are some responses to the question how a controlled list of keywords would be useful:

- *It is possible to find more pictures to the topic interesting for us;*
- *it is possible to broaden the area of searching;*
- *for supplementing the presentation for the text;*
- *it is possible to find out in the description details of the picture, e.g., origin;*
- *they show what search words one should apply for further searching;*
- *they give crucial words: a title of the picture, a format and when it came into existence, etc.*

About 46% of the respondents preferred searching the pictures; 44% chose browsing, and 10% were in favour of combination of both techniques. The question about the reasons of the preferred mode of retrieval were answered by the students as follows:

**Prefer searching because:**

- *I am finding concrete entry interesting to me;*
- *This way is more easy and pleasant;*
- *It gives required results immediately;*
- *It is narrowing closely to the topic that interests me;*
- *I get the sought information to the concrete question;*
- *I can get to the information faster;*
- *It is saving the time and gives a result to the concrete inquiry,*
- *I can use a concrete keyword;*
- *It is easier to reach needed sources;*
- *The right strategy is faster leading to the target*
- *It's more effective to use search function to find relevant pictures*
- *Don't want to browse with poor connection - 10 images download slowly;*
- *I know about searching, databases are made for searching.*

**Prefer browsing because:**

- *I don't always know what to search;*
- *It is easier to happen upon something new;*
- *Easier to find pictures if you don't know the exact wording*
- *Because I can work with many materials and many different topics if necessary;*
- *I do not have to write the right command down;*
- *When I don't know the topic I cannot use searching;*
- *Because searching not always is showing everything;*
- *Since searching as a rule is less intuitive.*
- *Browsing helps to get right pictures and words;*
- *Choosing broad areas give more freedom to get keywords and descriptions;*
- *Broad search is better than narrow one in images.*

**Prefer a combination of browsing and searching because:**

- *Combine browsing allows me to narrow search to a specific category and then search;*
- *Depends on what a person is after - both options serve well;*
- *Search then browse - allows one to get many images and by looking select;*
- *General browsing supports later keyword search.*

## **7.2 CH traffic and Log analysis**

This section is focused on the analysis of the queries are made from end-users on Libero Web Search environment (<http://www.iol.it>) in the time frame March-April 2006. The data collected clearly evidences the user interest in cultural objects. The analysis is limited to the Italian language: we imagine how MultiMatch multilinguality would be multiplying the numbers reported here.

We proceeded with the following methodology:



- We took all the queries made by end-users on Libero Web Search environment in the timeframe March-April 2006;
- We built at the same time a list of domains which were supposed to be relevant to MultiMatch (i.e. to cultural heritage issues).
- We tried was to automatically associate each query with one of the domains in (b).

This process was done automatically and adhered to the following constraints:

- The query must be a "well-formed" one : by this we mean that every lemma (which in turn can be made by more than one keyword) must represent a precise concept. This constraint roughly discards about 50% of the queries
- Each one of the "well-formed" queries are automatically analyzed to recognize if it can be associated with one of the domains in (b).

We generated for each domain, a list of the queries associated with that domain ordered by rank obtaining the most frequently used lemma for each CH domains ordered by rank.

The application in MultiMatch as possible requirement: if an end-user query (in a general search engine) matches one of the lemmas, a proposal for searching also in MultiMatch engine can be submitted to the user (like "one-box search" of Google).

### The CH domains: an example set

In the timeframe of March-April 2006 the users searched ~ 35.000.000 of "well formed" queries. Every query has been classified, and from this classification we selected the following CH-related domains and their queries and related distribution:

Domain	Sub-domains	Unique queries	Total queries	% total
Fine arts	fine arts, exhibitions, museums, churches, castles, culture, libraries, etc.	9.800	84.726	
Painting	painting	2.901	28.896	0.09
Sculpture	sculpture	627	4.301	0.01
History	history, archaeology, etc.	3.276	29.146	0.09
Religion	religion	31.988	381.613	1.10
Philosophy	philosophy	3.840	28.836	0.09
Literature	literature	20.330	336.814	0.96
Theatre	theatre	30.943	303.261	0.87
Music	music	22.178	315.343	0.90
Cinema	cinema	14.540	258.461	0.74
<b>Total queries CH</b>	<sup>(6)</sup>	<b>119.000</b>	<b>1.437.000</b>	<b>~ 4.1%</b>

**Table 70: CH queries and distribution**

<sup>6</sup> the totals are not the number of queries for each domain, because some queries can belong to more than one domain



It is intended to further investigate the needs of the “casual” or non-professional consumer of CH information for the purposes of personal interest, entertainment or travel via additional log analyses not only in Italy, plus an online questionnaire which is now being included in the WIND Libero portal (<http://www.iol.it/>). The results will be analysed to provide feedback for the functional specifications for the second prototype. The organization of the outcomes will follow categories established in the literature on web search, see [Rose and Levinson 2004].

### 7.3 User Requirements

This section presents the results of an analysis of the interviews, the scenarios, and the literature plus the log data. The rough data is organized into the four aspects: functionality, data, user, usability.

This initial long list will be further reduced to reach the core MultiMatch requirements.

Requirement classification (Functional; Data; Environment; User; Usability)	subgroupin g of the classification n	Referen ce number	User group rele- vance <sup>7</sup>			Requirement	Rationale
			E D	C T	C H		
data	annotation	UR-001	x	o	x	use standard descriptions with meaningful fields.	the user needs standardised metadata presentation
data	file formats	UR-002	x	x	x	support various formats for preview on internet, final edition and streaming;	the users use different platforms and different tools and need compatibility with the data formats
data	file resolution	UR-003	x	x	x	provide low resolution files for the works of an artist.	the user needs various data resolution files
data	file size	UR-004	o	x	x	user wants to have high definition 2K files after editing.	professional users need image size options: different for viewing and using
data	IPR	UR-005	x	x	x	include metadata about IPR	the user needs to be able to add personal metadata when he/she is the owner of part of the content
data	IPR	UR-006	x	x	x	inform about IPR (copyright photo, Museum property, copyright author, restrictions)for each image retrieved with relative reproduction fees	users need to have any information available on the copyright of the content
data	IPR	UR-007	x	x	x	user wants to know the technical costs, the right-of-use costs, the delivery date and wants to pay by credit card to the copyright holders	the user needs operatively to pay and have copyright licensing to use the content

<sup>7</sup> Convention: x= of interest for; o= not interested to.

data	layout	UR-008	x	x	x	content summarization	One or more keyframes should represent one shot (or very short video clip).
data	layout	UR-009	o	o	x	overview should visualize if shots belong to a video.	Videos can be opened to display thumbnail overview of all shots.
data	metadata	UR-010	x	o	o	descriptions of images in more than one language	the students need to be able to compare texts in different languages
data	metadata	UR-011	x	o	x	context and definitions of terms in subject, title, etc.	the user needs to be aware of the context and sources that generate some contents.
data	metadata	UR-012	x	x	x	display the date when an image/video/text was taken	the user needs metadata on all aspects from the contents retrieved
data	metadata	UR-013	x	x	x	show the author of an image/video/text	author is an important metadata for private content
data	metadata	UR-014	x	o	x	display comments or descriptions	annotations are important metadata for private content;
data	metadata	UR-015	o	o	x	provide additional automatically generated metadata	generic CH content cannot be expected to fill all the existing fields, nor to fill them for all if they have been put in a repository and not yet annotated
data	metadata	UR-016	x	x	x	addition of annotations anytime on executed searches	user would like to annotate whenever s/he feels like it, e.g. when remembering context of the searches executed: e.g. 'my trip to Thessaloniki, the hotel where I stayed'
data	metadata	UR-017	x	o	o	image should have references to authors or content owners	sometimes the content owner provides correlated connection to other content owners
data	multilinguality, summarization	UR-018	x	x	x	translate retrieved pages from one language to another from a list	the user needs a function to translate a page from one language into another
data	summarization	UR-019	x	o	x	provide brief and extended information about contents of the database	some user need long descriptions, others need short description
data	tools	UR-020	o	o	x	videos should be represented by keyframe summaries (which need to be meaningful.)	summaries help users quickly find relevant part of video
data	update	UR-021	x	o	x	include high volume of contents	the users require much content
data	volume	UR-022	x	x	x	diversify content	the users need all sorts of cultural data (images, documents, articles, etc.)

data	volume	UR-023	x	o	o	make all descriptions exhaustive and text-like	educational users evidenced the need for long descriptions and texts
functionality	aggregation of results	UR-024	x	o	x	organize pictures into thematic categories,	the user users want to organise results into collections
functionality	aggregation of results	UR-025	x	x	x	sort by relevance	sorting the results of a query
functionality	aggregation of results	UR-026	x	x	x	sort by popularity	sorting the results of a query
functionality	aggregation of results	UR-027	x	x	x	sort by upload date (search for recent contents)	sorting the results of a query
functionality	aggregation of results	UR-028	x	o	x	sort by title	sorting the results of a query
functionality	aggregation of results	UR-029	x	x	x	cluster results by topic	sorting the results of a query
functionality	aggregation of results	UR-030	x	o	x	Cluster results by image/video description	sorting the results of a query
functionality	aggregation of results	UR-031	x	o	o	cluster results by dominant features	the system analyses the dominant features of the results and clusters the results accordingly
functionality	correlation	UR-032	x	x	x	indicator for correlated sources automatically generated by the system according to user profile	who is linked to the content, to the site, to the source etc.
functionality	correlation	UR-033	x	o	x	detect relations between contents,(authors, keywords, IPR owners, period)	This way the user would get more detailed information about the possible search refinements he / she can perform
functionality	layout	UR-034	x	x	o	provide "hot" keys	the user wants to visualize a list of the top used keys. The hot keys should refer to the day of interest or another period (if the reference day is Halloween the top terms probably will refer to it, if there are other events happening, these too could be hot topics)
functionality	layout	UR-035	x	x	x	record all the information about the query results in one single report	after the search the actions performed are lost and some users need to record not only the history of the searches performed but also a summary of the results to look at offline
functionality	metadata	UR-036	x	o	x	search for type of content	sorting the results of a query

functionality	metadata	UR-037	x	o	x	search field for the date or period in which the photograph was taken or the painting created.	sorting the results of a query
functionality	metadata	UR-038	x	o	x	searching according to the author	sorting the results of a query
functionality	metadata	UR-039	x	o	x	search according to crucial words	users could assign different importance to the list of keys they input
functionality	metadata	UR-040	x	o	x	search using long queries	the users could cut and paste an abstract into the system
functionality	metadata	UR-041	o	o	x	search according to file size	sorting the results of a query
functionality	metadata	UR-042	x	o	x	search according to resolution	sorting the results of a query
functionality	metadata	UR-043	x	o	x	search according to format	sorting the results of a query
functionality	metadata	UR-044	x	x	x	search according to content description	sorting the results of a query
functionality	metadata	UR-045	x	x	x	search according to time period	sorting the results of a query
functionality	metadata	UR-046	x	x	x	search in other languages	sorting the results of a query
functionality	multilinguality, summarization	UR-047	x	x	x	provide summarization from source language to target language	the users need to handle other language contents
functionality	multilinguality, summarization	UR-048	x	x	x	provide automatic query completion	while user is entering a query, the system takes over and provides a suggested query
functionality	multilinguality, summarization	UR-049	x	o	x	synchronise results of search in different languages.	the user needs to obtain a translation of the results in one or more languages
functionality	multilinguality, summarization	UR-050	x	x	x	create possibility to search in other languages	the user executes a query in one language and the query is also executed in other languages into other languages
functionality	multilinking	UR-051	o	o	x	provide direct links to related images from the images found.	correlate retrieved results to similar ones by linking them (e.g. : Van Gogh painting of flowers and other flower paintings by other artists')
functionality	ordering results	UR-052	o	o	x	order images on the basis of a key provided by the user	the user needs to order the results give a textual relevance feedback

functionality	personalization	UR-053	x	x	x	provide "search history" feature ("store the old search").	the history of keys searched should be recorded and managed by the user
functionality	personalization	UR-054	x	o	x	allow interaction with the translation component for polysemous terms	the user wants to be able to interact with proposed terms translations when there is ambiguity (example: 'merlo': bird?, architectonic decoration?, wine?)
functionality	personalization	UR-055	x	o	o	adapt search results to user profile	the user needs to set search boxes for his / her typical search activities (ex: author box, related artists box, etc.)
functionality	personalization	UR-056	x	x	x	the user needs to create a personalized list of keywords	provide portion of the ontology for import in the user profile and adapt for the user needs. The user would like to create personalized thesauri connected to the system's. Example: an user creates lists of visited museums and to-be-visited museums with categories (arts, sculpture, painting, etc.)
functionality	personalization	UR-057	x	o	o	send shared results to friends	personalization feature
functionality	query format	UR-058	x	o	o	execute of long queries	input of long sentences similar to Yahoo!-context. The user would be enabled to use as query search an description of the author's works he/she is interested to. This is different from using just some keywords: here the keywords have a syntactic correlation, and not simply logical.
functionality	results aggregation	UR-059	x	o	o	alphabetical ordering of objects and authors.	the user requires a glossary where have term explanations and dictionary answers
functionality	results aggregation	UR-060	x	o	o	filter by language	the user requires ordering functionalities on the basis of language
functionality	results aggregation	UR-061	x	o	o	filter by file type	html, PDF, jpg, gif, avi, mpeg
functionality	results aggregation	UR-062	x	x	x	filter by date	ordering the results
functionality	results aggregation	UR-063	x	o	x	filter by usage rights	ordering the results
functionality	results aggregation	UR-065	o	o	x	filter by image size	ordering the results

functionality	results aggregation	UR-066	o	o	x	filter by coloration	ordering the results
functionality	search and retrieval	UR-067	x	x	x	retrieve material by subject;	ordering the results
functionality	search and retrieval	UR-068	x	x	o	evidence events	the user requests the results to evidence if there is any coming event related to the key
functionality	search and retrieval	UR-069	x	o	o	Retrieval capabilities in languages other than the mother tongue	the user needs to be enabled to have two language results to compare owns language understanding to another
functionality	search and retrieval	UR-070	x	x	x	limitation of the number of results	the user needs to decide how many results to visualize per page
functionality	search and retrieval	UR-071	x	x	x	system guide ('perhaps you are looking for...')	the user needs a suggesting tools that corrects eventual misspellings
functionality	search and retrieval	UR-072	x	x	x	advanced automated functionalities	(directory summarization, context analysis, biography etc.).
functionality	search and retrieval	UR-073	x	x	x	to save the process and later find a result on personal MultiMatch profile	the user requires to save the process that conducted to the results he / she obtained and later to access and modify it
functionality	search and retrieval	UR-074	x	x	x	access to directories: biography / video	the user needs to access pre-organized directories
functionality	search and retrieval	UR-075	o	o	x	user needs paging service.	optimisation of retrieval data collecting
functionality	search and retrieval	UR-076	x	x	o	results classified by country and by place;	possibility to have geo spatial result visualization / ordering
functionality	search and retrieval	UR-077	x	o	x	date ranging / timeline	The list must include only events between a selected date range (time line) and countries (country selection)
functionality	search and retrieval	UR-078	x	x	x	search by visual similarity.	the user needs to search by visual similarity eventually being enabled to upload the seed-image (video)
functionality	search and retrieval	UR-079	x	o	x	retrieval by copyright owner	the user requests to find contents that are freely available or available at low cost
functionality	search and retrieval	UR-080	x	x	x	retrieval of related topics by personal text metadata.	the user requests to be enabled the system to use his/her local list of keywords as a personalized list
functionality	search and retrieval	UR-081	x	x	x	user needs the system to propose controlled dictionaries and thesauri to organize the text information	specialized users (history / arts, etc.) require dedicated list of terms to their topics.

functionality	search and retrieval	UR-082	x	o	x	user needs to collect contents from different authoritative sources using a single search interface	similar to federated archives or open archive initiatives
functionality	search and retrieval	UR-083	x	x	x	user needs to search by similar images or concepts looking for semantically related descriptors in the metadata.	the user selects in an image an area or region of interest as seed for search
functionality	search and retrieval	UR-084	x	x	x	user needs to search by examples.	the user would provide textual /image example of what he/se needs
functionality	search and retrieval	UR-085	x	x	x	hot places search	hot places search: most searched places
functionality	search and retrieval	UR-086	x	x	x	The user needs to find well-known sites (such as Niagara falls or skyline NY) with a certain atmosphere about them: the client needs cold / calm colours and layout.	emotion-based searching
functionality	search and retrieval	UR-087	x	o	o	search by visual similarity to one or several examples,	an initial set of seeds for search referring to initiate the search by visual similarity is needed
functionality	search and retrieval	UR-088	x	o	x	learning of concepts	as the languages are evolving, new terms could be asked for being included
functionality	search and retrieval	UR-089	x	x	x	thesaurus (with translations)	user requests visualizing the thesaurus in more than one language: the thesauri are needed to be linked together: key in one language linked to translation keys in other languages
functionality	search and retrieval	UR-090	x	o	x	index of pictures/authors	users need to have a list of pictures/authors
functionality	search and retrieval	UR-091	x	x	x	Boolean search	users need Boolean operators
functionality	search and retrieval	UR-092	x	x	x	truncation	user need truncated terms to be valid as full terms
functionality	search and retrieval	UR-093	x	x	x	provide truncation in multilingual context	the user decides in which language the truncation has effect
functionality	search and retrieval	UR-094	x	o	x	limitations of the database	the user needs the results coming from specific databases, authoring databases
functionality	search and retrieval	UR-095	x	x	x	provide search operators, multilinguality	this means not simply Boolean but also other operations coming from combination of search tools ('search in English inclusively and in German exclusively')



functionality	search and retrieval	UR-096	x	o	o	"origin of the picture",	the user requires to know all the main subjects that are related to a picture / video: author; owner; who modified the image (if any)
functionality	search and retrieval	UR-097	x	o	x	Search by historical period	timeline
functionality	search and retrieval	UR-098	x	o	o	search by precise location,	geographical references
functionality	search and retrieval	UR-099	o	o	x	search by black and white as a key-word,	Professional users need specialized search options
functionality	search and retrieval	UR-100	x	o	x	availability of advanced fixed fields	fields specifically oriented to query for authors, artists, buildings. Similar to query a database
functionality	search and retrieval	UR-101	x	x	x	availability of authority lists	lists of authors, artists, buildings. Similar to obtain a complete field database view
functionality	search and retrieval	UR-102	x	o	x	textual relevance feedback	the user asks for use a set of results as input for a new search (textual relevance feedback)
functionality	search and retrieval	UR-103	x	x	x	browsing by categories or web directories	user need some pre-collected and organized search areas. White lists
functionality	search and retrieval	UR-104	x	x	x	option to select and acquire items from a past result	the user requires to use part of past queries results
functionality	search and retrieval	UR-105	o	o	x	option to select and acquire items from a visualization aspect	selecting some lists of visual aspects, the user executes queries
functionality	search and retrieval	UR-106	x	x	x	For textual search, user wants to be flexible in specifying the textual query, e.g., single terms or a set of terms, fragmented sentences, keywords from an existing list.	Users have different habits of text input.
functionality	search and retrieval	UR-107	x	x	x	Give a rate to each image with the text	similar to relevance feedback but with differentiation of the evaluation or relevance
functionality	search and retrieval	UR-108				Travel, hotel fares ordering by price	ordering the results
functionality	search and retrieval	UR-109	x	x	x	information about long term search status: found/still searching/all viewed	intelligent search agent management
functionality	services	UR-110	x	x	x	persistent search agents	Important to save a query, if it is complex. The search agent repeats only important queries if repository is changing often or if request is a long-term interest of



							a user. Repeated search is a requirement for certain businesses, not in general.
functionality	services	UR-111	o	x	o	To organize a travel path and actions in a sequential and optimised way	the user needs tools to organize the cultural travel
functionality	suggestion search	UR-112	x	x	x	Search suggestions displayed as you type in the query	the user needs textual suggestion while writing the query
functionality	suggestion search	UR-113	x	x	x	Search suggestions displayed along with search results	the user needs a list of suggestions after the query has been executed
functionality	tools	UR-114	x	x	x	Use the auto supplementation of expressions written down in other languages	the user can select the language for query suggestions
functionality	tools	UR-115	x	x	x	Relevance feedback	"more like this image", "images with this shape"
functionality	tools	UR-116	x	x	x	Visual relevance feedback combined with keywords	possibility to use a relevance feedback referring annotation keys
usability	clearness	UR-117	x	o	o	Clear navigation instructions (including alternative text)	the user needs alternative texts and explanations
usability	communication to the user	UR-118	x	o	x	explanation of search criteria in order to avoid fruitless searching,	the user needs to know the search mechanism
usability	easiness	UR-119	x	o	x	Easy overview of the functions of the databases	the user needs clear commands
usability	help	UR-120	o	o	x	To allow the use of the browser's reverse/back button.	integrated functions that help the results browsing
usability	help	UR-121	x	x	x	Users require the option to get a suggested list of keywords for search as soon as he/she types in the key for search. Fill up.	A push list of existing keywords from a thesaurus makes sure user will not spoil the query by spelling errors or choosing terms that are unknown to the system.
usability	help	UR-122	x	x	x	Search examples	user requires examples to improve the search ability
usability	help	UR-123	x	x	x	Provide guidance and help feature for the advanced search.	user requires guides (video) to improve the search ability
usability	IPR	UR-124	x	x	x	Indicator for detected IPR holder	the user need clear IPR references
usability	IPR	UR-125	x	o	x	Use discreet watermark on the pictures	the users require to see full images
usability	IPR	UR-126	x	x	x	user needs an indicator of authoritative of sources	the users need to know which authoritative repository is related to some contents and the IPR to

							which party are to be paid
usability	layout	UR-127	x	x	x	Standard graphic layout	the layout must respect w3c standard, in particular for accessibility
usability	layout	UR-128	x	x	x	Slide show once many items have been found	so the user has not to select one item after the other to view the results
usability	layout	UR-129	x	o	x	Thumb images on one page	possibility to reduce / enlarge the images viewed in the result list
usability	layout	UR-130	x	o	x	Description of the image when using a magnifier	the user requests to have the description as 'alt' text
usability	layout	UR-131	x	x	x	the user needs the interface to evidence with different colours the keywords of the query on the retrieved pages	Use colours to highlight keywords on the results
usability	layout	UR-132	x	x	x	once the keywords have been colour coded the user needs to be able to choose which keywords to examine	the output page has different coloured keywords: the user needs to navigate links based on one of the keywords and not all of them
usability	multilinguality, summarization	UR-133	x	x	x	translation service for the input keyword checking suggestion	do you need to translate the keyword and execute the search?
usability	navigation	UR-134	x	o	x	To allow users to navigate between pages, without returning to the thumbnails	differentiation of the navigation bars
usability	optimization	UR-135	x	o	x	An option to put the same content in several collections (without keeping several copies of the content): i.e: landscapes, landscapes of Tuscany	For topical collections, the same item of content often fits several categories, but should not be copied and stored redundantly in several collections.
usability	personalization	UR-136	x	x	x	Personal folders and shared folders	the user needs to manage a personal area where results can be stored
usability	personalization	UR-137	x	x	x	interface defining search options	the interface should provide definition of the search options
usability	personalization	UR-138	x	x	o	If the system is going to share results and folders, users require to be informed/manage about this before sharing it with others.	Users want manage content organization and sharing.
usability	personalization	UR-139	x	x	x	user needs to manage the long term searches and the notification	user require to manage the search agents

usability	search and retrieval	UR-140	x	x	x	Browsing support,	the user needs support on how to browse the results and organize and save them
usability	search and retrieval	UR-141	x	x	x	Explanations of choices in advanced search menu	user need information in 'alt' format
usability	services	UR-142	x	x	x	user wants to be informed about the services available	help and manuals
usability	services	UR-143	x	x	x	Provide possibility to go back to a performed search,	so that one does not have to right it all over again
usability	services	UR-144	x	o	x	Users require an option to review when the system plans to delete content	Users want to control important processing of personal content.
usability	tools	UR-145	x	x	x	Easy way to view images	the user needs tools that help the visualization (zooming)
usability	tools	UR-146	x	x	x	Create possibility of enlarging (zoom) of text	some user need interface zooming for vision problems
usability	wizard	UR-147	x	x	x	wizards to organize the results and layouts	wizards help the user to manage their interface and the existing templates help immediate personalization of personal interface
user	aggregation of results	UR-148	x	x	x	automatic organization of results	Occasionally, users want to create a folder of images taken over a longer period of time.
user	annotation	UR-149	x	o	x	Give an annotation relative to the digital quality of the images, colour and orientation	the user need to annotate some results and personalize them
user	data management	UR-150	x	x	x	import of video material in editing software;	the user needs to collect and modify video content
user	layout	UR-151	x	x	x	MultiMatch supports various formats for preview on Internet, final edition and streaming;	the user will be enabled to view all file formats available on the web
user	layout	UR-152	x	x	o	Organize trips and travel mapping	by combining the query results, the user organizes partially a travel trip
user	layout	UR-153	x	x	x	prints the full report as a diary with some blank pages for notes and photos to be added as an album	the user needs to organize the results in a nice way and to print out the report
user	layout	UR-154	x	x	o	To print the full report in a MultiMatch layout	the results of a query are printable with a nice standardised layout
user	layout	UR-155	x	x	x	to print a summarized page that MultiMatch has generated: the page contains artist 's life, some	full data presentation

						top important articles on his masterpieces, the complete list of the movies he made, the film festivals he participate, the actors (for movies) who worked for him and some short stories and pictures	
user	layout	UR-156	x	x	x	The user needs the possibility to input comments on the page generated by the MultiMatch interface.	the users require to annotate contents
user	layout	UR-157	x	x	x	Help function for personalizing the interface	textual and video presentations that help personalization of the interface
user	layout	UR-158	x	x	x	Browsing options	nice and clear browsing options
user	layout	UR-159	x	x	x	Timeline	a line that allows date visualization and ranging
user	layout	UR-160	x	x	x	timeline, flagging	the user needs to put flags and events on the timeline
user	layout	UR-161	x	x	x	Collaborative search	the users need to share a repository of searches
user	layout	UR-162	x	x	x	Search history	the user needs to keep stored past queries
user	layout	UR-163	x	x	x	An option to specify the search result presentation (new results overwrite old results vs. Results are aggregated)	it depends on the task and the context if old results are valuable or can be dismissed.
user	layout	UR-164	x	x	x	option to represent figuratively the search query as an avatar or personage (Jeeves)	the user needs visual representations of the queries as objects
user	layout	UR-165	x	x	o	User requires to see the image/keyframe as a thumbnail or small size display when actually viewing a collection, downloading or listening to audio notes.	layout of the results
user	layout	UR-166	x	o	x	Users require automatic identification of clusters of contents (video/images/audio)	results organization and classification
user	layout	UR-167	x	x	x	option to share a slide show of the retrieved results	Time interval between two slides should be set by user. Also required stopping, switching to manual navigation, interruption, etc.

user	multilingual ity, summarizati on	UR-168	x	x	x	Browsing possibilities in many languages in parallel	the user needs a browsing capability that allows the navigation of the content in more than one language
user	multilingual ity, summarizati on	UR-169	x	x	x	Spelling check to correct possible mistakes.	the user needs the system check the correctness of the query vocabulary by providing term definitions
user	personalizat ion	UR-170	x	x	x	access ones personal area	authentication and privacy
user	personalizat ion	UR-171	x	x	x	upload personal pictures	the user adds content to the system
user	personalizat ion	UR-172	x	x	x	to generate sharable files to provide to friends or customers	basis for communicating among users
user	personalizat ion	UR-173	o	x	x	The user needs to use the MultiMatch page as a brochure to be distributed	layout of the results in a personalized and printable format
user	personalizat ion	UR-174	o	x	x	The user needs to share a page that has designed with the exhibition brochure (this page will be indexed for those who will search events in the fixed event period).	layout of the results in a personalized and printable format
user	personalizat ion	UR-175	o	x	x	The user needs to help clients to find special contents in BigSearchImages' image base.	services to professional users
user	personalizat ion	UR-176	x	o	x	automatic description of the contents	the user needs the contents and results to be described according to his/her profile automatically
user	personalizat ion	UR-177	x	o	x	Graphical visualization of clustered results	representation of results and classification
user	personalizat ion	UR-178	x	x	x	Option to name/rename the search query before/after storing it	query management
user	personalizat ion	UR-179	x	o	x	the user requires third party to offer results and execute searches for him/her	the user allows third parties to provide him/her services such as query results
user	personalizat ion	UR-180	x	x	x	system should offer search terms from user history, speeding up the typing	This would ease text input in case of repeated search by similar queries.
user	personalizat ion	UR-181	x	o	x	option to view video content, i.e. to play the video inside the page provided by the system	video tools and options
user	personalizat ion	UR-182	o	x	o	option to manually prepare a sequence which does not follow any	arbitrary sequence allow user to create effects or cluster images that are not in a formally

						formal sorting criterion (CT users)	specified sequence.
user	refinement of searches	UR-183	x	x	x	The user needs to browse the resulting subset of images or further limit it by adding query expressions.	the user requires to be enabled to modify the query string at any time and in the meantime, keeping the collected results in a temporary area
user	security	UR-184	x	x	x	protect me from undesired sites	the user should be enabled to block results from non desired sites
user	services	UR-185	x	x	x	exporting the interface as a template	after personalizing ones interface, the user needs to export it as a template to be shared or offered to third parties or friends
user	services	UR-186	x	o	x	The user needs saving all the results on third party folder in MultiMatch system.	collaborative search
user	services	UR-187	o	o	x	The user needs to advertise the exhibition somehow on MultiMatch to her address list and to a wider group of users.	communication tools
user	services	UR-188	x	x	x	retrieval support,	the user could ask for professional dedicated search services
user	services	UR-189	x	x	x	Basket	personal folder for contents that need IPR payment
user	services	UR-190	o	x	x	When providing results, ordering by file sizes optimised	Some users filter content and store it separately if it is of a larger size (i.e. for backup reasons)
user	services	UR-191	x	x	x	option to send the topic-search or share it with other users	the user shares the topic of interest
user	services	UR-192	x	o	o	integration of RSS/Blogs services	the user accepts RSS/Blogs feeds services on the interface
user	services	UR-193	x	x	x	see if other researchers are working on the same topic	some users share their data, see community and self advertising

Notice: the requirements collected do not address the environment as the users will access MultiMatch through the Web.

## 8 User Requirements Selection for MultiMatch

This section focuses on the most relevant user requirements that will characterize the MultiMatch system. The extensive set of requirements introduced in section 7.3 is filtered in order to select those with high priority for implementation in MultiMatch; such requirements must clearly also be in line with the MultiMatch vision (as expressed in Appendix I). They will be taken as input by Deliverable 1.3 on Functional Specifications.

The main reasons for the final choice of requirements are the following:

1. Respondents were mainly interested in searches for creators, their works and sites where information could be found, which strongly suggests that MultiMatch should offer this kind of specialised search. Obviously, this requirement implies not only the use of full text indexing for cultural resources but also the use of metadata to improve the retrieval and information managing process.
2. Respondents were also very interested in information aggregation functionalities. Authors, artists, etc. and information on their works of art represented the most searched items, so MultiMatch should clearly classify content in this respect and provide the relevant search functionality. The provision and use of metadata should improve the information aggregation process.
3. Respondents were interested in image and video retrieval although they do not use these facilities as often as they search text.. This suggests the possibility of integrating image and video results with text search results when possible, and also to allow the MultiMatch user to search for images and video independently.
4. Respondents trust in the results provided by general search engines, which suggest the use of aggregated results from other sources (i.e. such as Google, Yahoo!, etc.) to cover those web pages not indexed by MultiMatch. This will allow the user to complement their searches with “well known” search engines.
5. Respondents see clustering as a useful technique for grouping related and relevant search results, which suggest providing MultiMatch with clustering capabilities. MultiMatch should apply clustering from two different perspectives: a) using a pre-defined ontology for cultural heritage context which could be visualized as a web directory (i.e. familiar to the user), and; b) using free text based clustering which builds clusters using relevant keywords extracted directly from the retrieved information.
6. Non-native English speakers ranked English as the most relevant language to obtain updated and authoritative information about cultural heritage (i.e. apart of their native language). This suggests that providing MultiMatch with multilingual functionalities (i.e. for querying and results presentation) should be helpful for the users.
7. Natural language querying, Boolean operators and relevance feedback are retrieval techniques well known for the user and should be integrated in MultiMatch.
8. Respondents were interested in sorting and filtering capabilities for search results, which suggest MultiMatch should provide these kinds of functionalities.
9. Respondents are familiar with spell checking and search suggestion tools which suggest that these should be included in MultiMatch.
10. Respondents were interested in managing their search results. Results annotation and search history were ranked as very useful for the user. This suggests providing MultiMatch with these functionalities based on a registered user profile.



The set of requirements considered most relevant for MultiMatch and that will constitute input for the functional specifications is listed below.

Req ID	Category	Requirement	Rationale
<b>Data Requirements</b>			
MMR-001	Metadata	Categorization of web pages according to content and source	All the web pages indexed by MultiMatch should be categorized as follows: <ol style="list-style-type: none"> <li>1. Author web page</li> <li>2. Artwork web page</li> <li>3. Cultural heritage site web page</li> </ol> Information about source should also be provided: <ol style="list-style-type: none"> <li>1. Cultural heritage site.</li> <li>2. Educational site</li> <li>3. Tourism site</li> </ol>
MMR-002	Metadata	Provide metadata for authors	Authors should be indexed in MultiMatch with a set of author-related metadata. This set could be enlarged using MultiMatch ontology (WP2).
MMR-003	Metadata	Provide metadata for works of art	Works of art should be indexed in MultiMatch with a set of related metadata. This set could be enlarged using MultiMatch ontology (WP2).
MMR-004	Metadata	Provide metadata for cultural heritage sites	Cultural heritage sites should be indexed in MultiMatch with a set of related metadata.. This set could be enlarged using MultiMatch ontology (WP2).
<b>Functionality Requirements</b>			
MMR-005	Search and retrieval	accept natural language queries	The user needs to ask the system about any topic related to cultural heritage (e.g. flowers).
MMR-006	Search and retrieval	process specific natural language queries about authors	The user needs to ask the system about a specific author to retrieve as much information as possible about him/her (e.g. Van Gogh).
MMR-007	Search and retrieval	process specific natural language queries about artworks	The user needs to ask the system about a specific work of art to retrieve as much information as possible about it (e.g. The sunflowers).
MMR-008	Search and retrieval	process specific natural language queries about cultural heritage sites	The user needs to ask the system about a specific cultural heritage site to retrieve as much information as possible about it (e.g. Tate Gallery).
MMR-009	Search and retrieval	process specific natural language queries about cultural events and news.	The user needs to ask the system about news and events related with the cultural heritage context (e.g. Reopening of Liceo in Barcelona) <sup>8</sup> .

<sup>8</sup> This suggests a non-trivial extension to the original proposal, with very substantial implications. It is to be considered as an optional.



MMR-010	Search and retrieval	offer search results aggregation	The user needs to retrieve results not only from MultiMatch indexes but also from external sources such as general search engines (e.g. Google or Yahoo!) and encyclopaedic sources (e.g. Wikipedia).
MMR-011	Search and retrieval	offer cross-language retrieval	The user needs to retrieve information from different sources in different languages supported by MultiMatch.
MMR-012	Search and retrieval	offer image retrieval	The user needs to retrieve only images in MM (e.g. like Google Images)
MMR-013	Search and retrieval	offer video retrieval	The user needs to retrieve only video in MM (e.g. like Google Video or YouTube)
MMR-014	Query	accept natural language combined with Boolean operators	The user needs to query the system including or excluding specific keywords using Boolean operators (e.g. flowers AND painting NOT sculpture).
MMR-015	Query	offer relevance feedback (textual or image based)	The user needs to launch new queries using results of previous searches.
MMR-016	Browsing	offer browsing among MultiMatch contents using a web directory like structure	The user needs to browse MM contents using a web directory-like structure which organizes the contents of MM according to cultural heritage related categories (i.e. like dmoz.org but with categories defined by MM ontology).
MMR-017	Browsing	offer browsing across results using a clustering structure.	The user needs to browse across retrieved results using automatically built clustering of relevant and query related keywords.
MMR-018	Information compilation	offer author search results compilations	For specific author searches, MM should be able to compile and classify the search results accordingly basic subcategories related to the author
MMR-019	Information compilation	offer works of art search results compilations	For specific work of art searches MM should be able to compile and classify the search results accordingl to basic subcategories related to the work of art
MMR-020	Information compilation	offer cultural heritage sites search results compilations	For specific sites searches MM should be able to compile and classify the search results according to basic subcategories related the site searched
MMR-021	Information mining	offer author search results mining	For author searches, MM should be able to mine its indexes to find information related to the author searched. This capability will allow the user to improve his knowledge about the author
MMR-022	Information mining	offer works of art search results mining.	For work of art searches MM should be able to mine in its indexes to find related information. This capability will allow the user to improve his knowledge about the work of art
MMR-023	Information mining	offer site search results mining	For site searches MM should be able to mine in its indexes to find related information. This capability will allow the user to improve his knowledge about the cultural heritage site context
<b>User Requirements</b>			
MMR-024	Layout	offer graphical representation for clustering	To allow the user to browse easily between automatically created clusters.
MMR-025	Layout	offer graphical representation for relations between items (authors,	To allow the user to take a quick snapshot of the main relations between items.

		works of art and sites)	
MMR-026	Layout	offer schematic representation for specific items (authors, works of art and sites)	To allow the user to take a quick snapshot of the most relevant data related to a specific item
MMR-027	Layout	offer Google like presentation of web and news results	The user feels familiar and comfortable with this visualization paradigm.
MMR-028	Layout	offer Google Maps like representation of geographical positioning of items (authors, artworks and sites)	The user feels familiar and comfortable with this geographical visualization paradigm.
MMR-029	Layout	offer timeline for specific items (authors and artworks)	This helps the user to take a snapshot of the retrieved items in the time.
<b>Usability Requirements</b>			
MMR-030	customization	offer “sign in” capabilities	To register the user and allow search customization
MMR-031	customization	offer search history	For registered users
MMR-032	customization	offer results annotation	For registered users
MMR-033	customization	offer retrieved results sorting	To sort retrieved results accordingly different criteria such as: <ol style="list-style-type: none"> <li>1. Pertinence (i.e. relevance to the query)</li> <li>2. Language</li> <li>3. Type of web site (educational, cultural heritage o tourism site)</li> <li>4. Size of the resource</li> <li>5. Date when the resource was uploaded</li> <li>6. Type of file (image, video, text)</li> <li>7. File format (HTML, PDF, JPG, GIF, AVI, etc)</li> </ol>
MMR-034	customization	offer retrieved results filtering	To filter search results (at query time or when results are retrieved) according to different criteria such as: <ol style="list-style-type: none"> <li>1. Language</li> <li>2. Type of web site (educational, cultural heritage or tourism site)</li> <li>3. Size of the resource</li> <li>4. Date when the resource was uploaded</li> <li>5. Type of file (image, video, text)</li> <li>6. File format (HTML, PDF, JPG, GIF, AVI, etc)</li> </ol>
MMR-035	Search and retrieval	offer search suggestions	When showing search results
MMR-036	Search and retrieval	offer spell checking	When typing the query or over retrieved results.

**Table 71: User requirements list for MultiMatch.**

## 9 Conclusions

The user requirements study has been performed examining data from a number of sources. We have described the user groups that we have investigated and have defined the methodology adopted to analyse their needs. 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. Although this study has mainly focussed on users of cultural heritage information for professional purposes, we have also studied log data from the WIND portal in order to understand the types of CH query formulated by the general user (in this case the Italian user). In a first stage, we identified a very large set of requirements, which is presented in Section 7.3. We then analysed this set in order to identify in the order: (1) the most requested (and thus considered as high-priority); (2) those requirements that best matched the previously declared project objectives and vision. These requirements are listed in Section 8 and are now being transformed into MultiMatch functional specifications in deliverable 1.3.

This deliverable has focussed on the professional user of CH information. However, we also intend to further investigate the needs of the “casual” or non-professional consumer of CH information for the purposes of personal interest, entertainment or travel via additional search engine log analyses in several countries. An online questionnaire is also being included in the WIND Libero portal (<http://www.iol.it/>). The results will be analysed to provide feedback for the functional specifications for the second prototype.

## 10 References

- Agile (2003). <http://www.agilealliance.org/home> (last accessed September 2006).
- Boehm, B.W. (1988). A Spiral Model of Software Development and Enhancement, IEEE Computer, pp. 61-72.
- Goto, K. and Cotler, E. (2005). Web ReDesign 2.0: Workflow that Works, New Riders Publishing.
- Hackos, J. and Redish, J. (1998). User and Task Analysis for Interface Design, Wiley Computer Publishing.
- Hoffer, J.A., Valacich, J.S. and George, J.F. (2005). Modern Systems Analysis and Design, 4th Edition, Prentice Hall.
- Ingwersen, P. and Järvelin, K. (2005). The turn: integration of information seeking and retrieval in context. Dordrecht, The Netherlands: Springer.
- Kleppe, A., Warmer, J. and Bast, W. (2003). MDA explained: the model driven architecture: practice and promise. Object Technology Series. Addison-Wesley.
- Kruchten, P. (2003). The Rational Unified Process: An Introduction, 3rd Edition, Addison-Wesley.
- Kotonya, G. and Sommerville, I. (1998). Requirements Engineering – Processes and Techniques, Wiley.
- Lethbridge, T.C. and Laganière, R. (2001). Object-Oriented Software Engineering: Practical Software Development using UML and Java, Second Edition, McGraw Hill.
- Maciaszek L.A (2005): Requirements Analysis and System Design, 2 ed. Addison Wesley.  
[http://www.comp.mq.edu.au/books/rasd2ed/ReadersArea/LectureSlides/pdfpage.htm?Chapter%202=RASD2ed\\_Ch2.pdf](http://www.comp.mq.edu.au/books/rasd2ed/ReadersArea/LectureSlides/pdfpage.htm?Chapter%202=RASD2ed_Ch2.pdf)
- Nielsen, J., (1993). Usability Engineering, Academic Press, Boston.
- Preece, J., Rogers, Y. and Sharp, H. (2002). Interaction Design: Beyond Human-Computer Interaction, New York, NY: John Wiley & Sons.

- Pressman, R.S. (2001). Software Engineering: A Practioner's Approach. McGraw-Hill.
- Rose, D.E. and Levinson, D. (2004). Understanding user goals in web search. In Proceedings WWW 2004, pages 13–19, New York, NY, USA, 2004. ACM Press.
- Sommerville, I., Sawyer, P. (1997). Viewpoints: Principles, problems and a practical approach to requirements engineering, Annals of Software Engineering. 3, pp.101-30.Appendix I

## Acknowledgements

The authors gratefully acknowledge the internal reviewing of an earlier version of this deliverable by a project partner from UvA whose comments and criciticisms have been invaluable. The reviewers from USFD who commented on a later version are also thanked; their observations have been incorporated as far as possible.

## Appendix I: The Common Vision

The MultiMatch technical annex specified several research lines, which had to be made concrete in WP1, mapping input from partners and users into a functional specification for the MultiMatch search engine. This appendix is the result of an initial project activity in which the partners discussed and agreed on a unified, common vision of important aspects of the proposed search engine. As such, it does not attempt to present a comprehensive view of the work to be done, but rather to concentrate on issues to be agreed upon before collecting the user requirements and presenting a complete functional specification. The result of the activity was a document denominated “The MultiMatch Common Vision”. The document was kept brief and simple, using simulated interface snapshots to clarify the ideas presented.

Note that interface snapshots are just illustrations of the ideas that have been under discussion, and therefore are not binding on how the different functionalities will ultimately be organized, nor should they be considered as representative of the future “look and feel” for the implemented interface.

### Functionality of the MultiMatch Search Engine

The key ideas of the proposed search engine are *multiplicity* and *aggregation*.

**Multiplicity:** MultiMatch will offer results in multiple languages, multiple media, with multiple views (of searching/browsing), and with multiple links between pages and sites.

**Aggregation:** The MultiMatch interface will integrate standard and specialized search results from our index, results from other search engines, and pointers to external material.

Regarding **multilinguality**, users will be able to pose queries in their preferred language(s) and retrieve material in all languages handled in the project. According to the user’s language profile, results in unknown languages will be returned in a way that is interpretable by the user, e.g. with a summary or associated keywords in the user’s preferred language, or optionally with a translation acquired from an on-line machine translation service.

Regarding **multimediality**, users will be able to search text, images, audio and video. Image search facilities will include text and content-oriented matching. Audio and video search will include the capacity to search transcribed speech, at least in English.

Figure 1 shows a possible initial interface for the search engine. The remainder of this section clarifies a number of issues regarding MultiMatch functionality.

#### What is exactly meant by “specialized search” in the context of Multimatch?

Generically, it refers to results that go beyond simple resource matching & ranking, adapted to the search needs of average user profiles for cultural material, and which typically involve Information Extraction and/or Web Mining techniques.

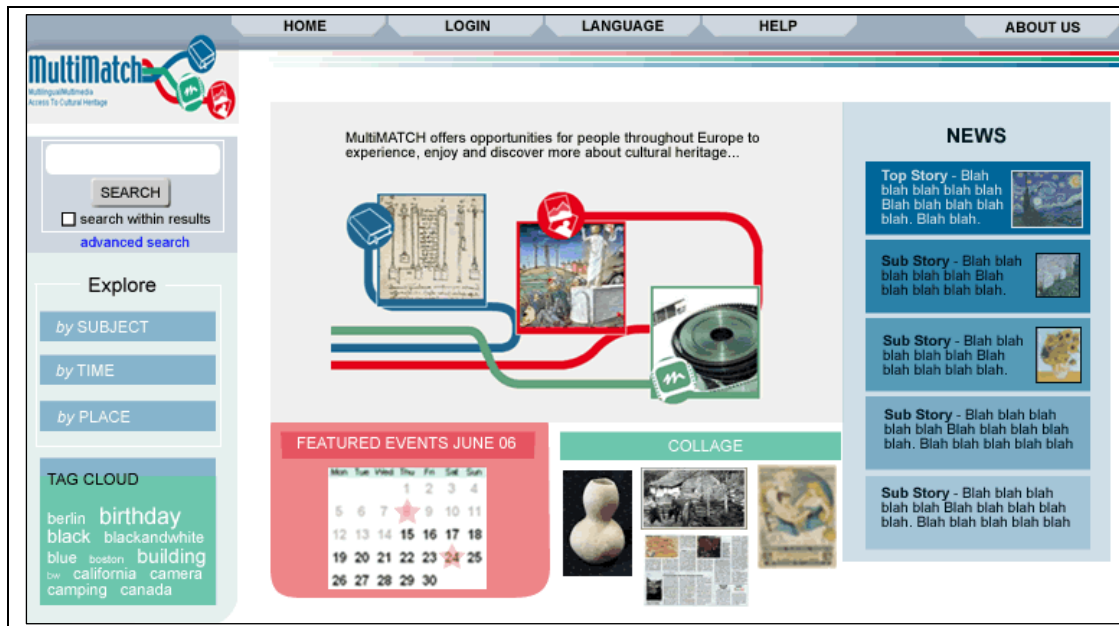


Figure 1: Depiction of a possible homepage for the MultiMatch search engine, including a query box (default query interface) and some browsing possibilities (by ontology, left-hand side, by theme, events by calendar, tag cloud, geographical search, news browsing).

We can think of three kinds of specialized searches:

- 1 **"Information Extraction" example:** Specialized search for online cultural events. By online we mean events that take place on the Internet (e.g. a chat with a famous writer) or that are recorded and uploaded to some web site (e.g. a conference by a famous writer which is videotaped and offered as a video stream by the institution organizing the event). The specialized search interface will allow users to narrow a search according to a number of fields (type of event, topic, date, etc), detect web pages offering cultural events, extract the information for the relevant search fields, and display results as in a database search rather than a ranked list of hits.
- 2 **"Web Mining" example:** Mining the Internet for cultural icons, for instance: which artists are most closely related to X? Which painters are most closely related to writer X? What are the moods most related to paintings/writings by X? What are the cities most closely related to X? What are the icons most related to city Y? Find critical reviews of author A or book B?
- 3 **"Aggregation" example:** summarize everything about author X. Aggregation is slightly different from specialized search, because it uses results from specialized search facilities and also external resources. In this case, MultiMatch will aggregate
  - Multilingual/multimedia results from the MultiMatch index
  - Results from federated database searches (e.g. Alinari image database)
  - Links to relevant external resources such as Wikipedia.
  - Results from MultiMatch specialized search facilities, such as online cultural events related to X and mined trends about X (the two examples above).
  - Faceted search using MultiMatch ontology. Facets will include not only cultural distinctions but also other useful resource attributes, e.g. "official/non-official", "free/pay resource", etc.
  - Live results from major internet search engines, e.g. Alexa, Google or Yahoo APIs.

By default, the system will offer aggregated results, with a number of sources depending on the type of query (see Figure 1). Aggregated results will include, for appropriate searches (e.g. names of artists), the results of one or more MultiMatch specialized searches. Figures 2 and 3 illustrate the aggregated results concept for the query "Van Gogh".





Figure 2: Search for documents on Van Gogh in English, Dutch, French and Spanish

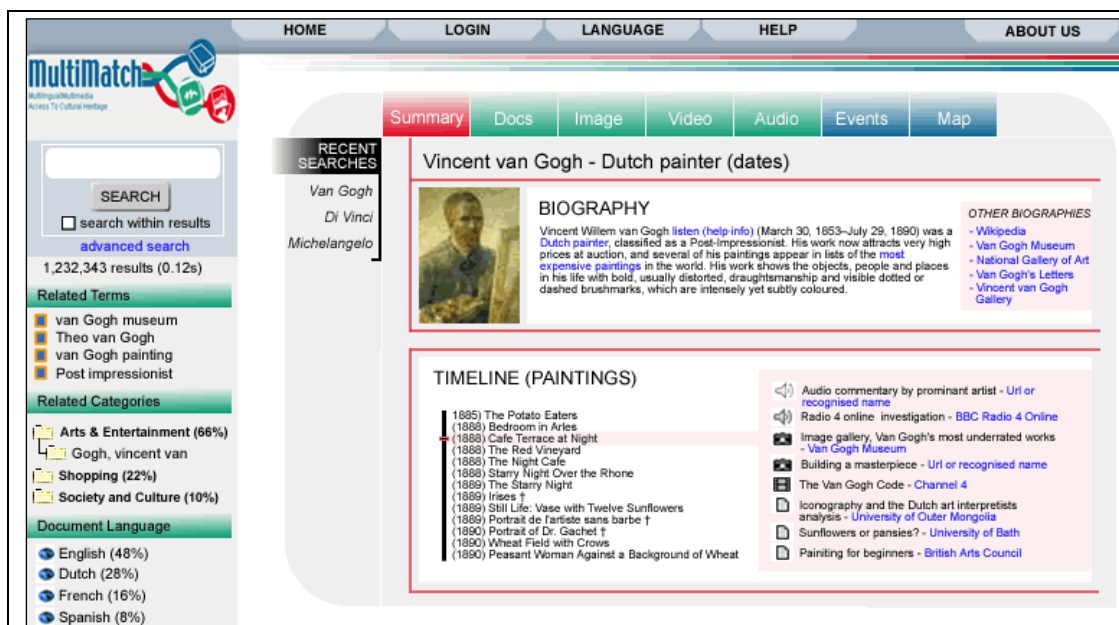


Figure 3: Aggregated results are presented under different tabs, for different media (docs, image, video, audio), languages (lower left corner) and different classes of information (events, biographical info).

It will also be possible to use some of the specialized search services directly, with a specialized search interface designed to interact with the user to retrieve optimised search results. For instance, when looking for online cultural events, the specialized search interface will permit specifying the type of event (seminar, theatrical representation, dance, etc.), the place, date, etc (see Figure 4)

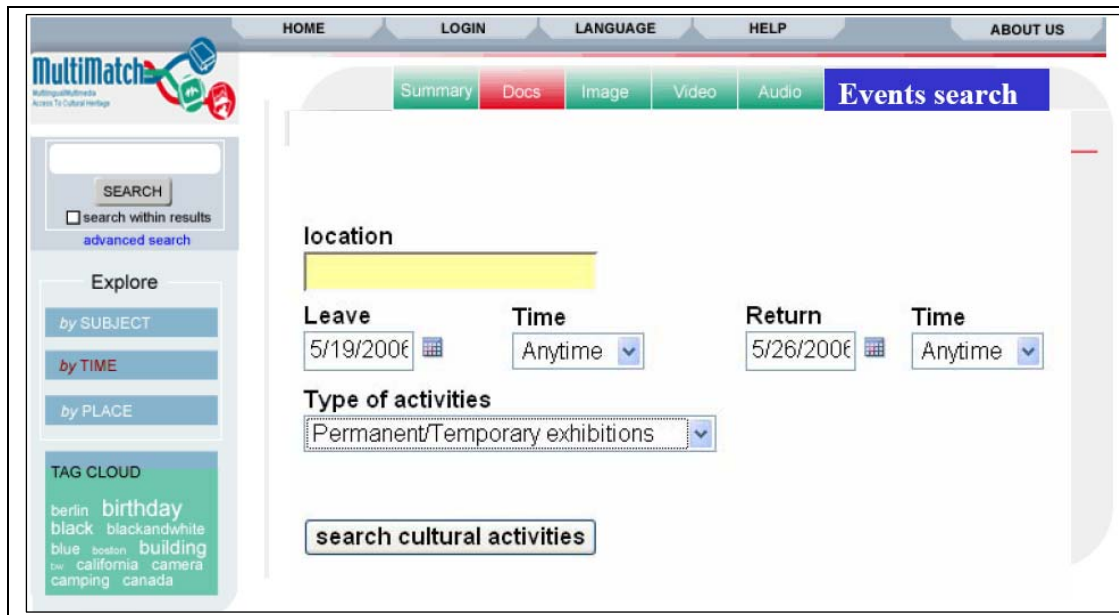


Figure 4: An example of specialized search interface, in this case a search for cultural events in the cultural tourism profile.

Each type of media will also have its own personalized search interface, including search by example (give me more like this) for audiovisual material.

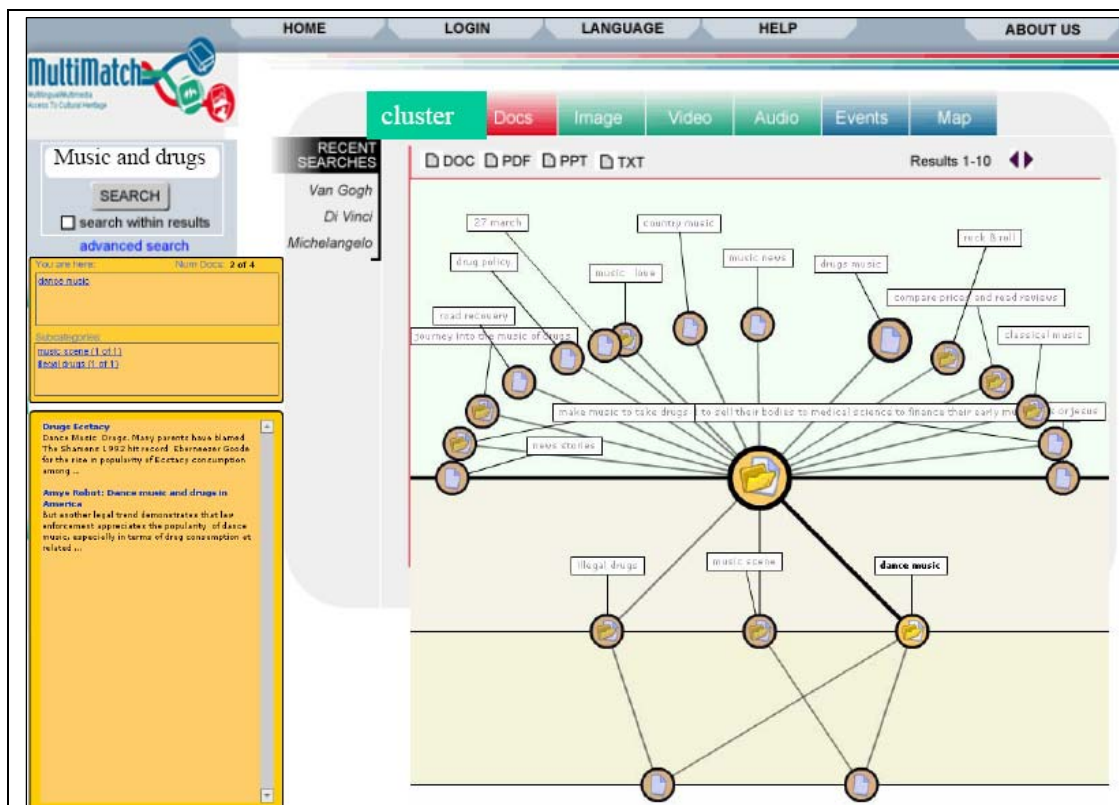


Figure 5: An example of generic clustering facility when the query does not fit a predefined type of MultiMatch object, in this case “music and drugs”. The user selects the “cluster option” and MultiMatch derives a grouping for the search results. In the image, the use has selected the “dance music” cluster, which has an intersection with “illegal drugs” and with “music scene”.



### What does clustering mean in the context of MultiMatch?

For non-structured results (ranks and lists), MultiMatch will have one or more general clustering options (see Figure 5 for an example). For structured results, clustering is implicit in the structure: for instance, the query *Van Gogh* will lead to two sets of aggregated results, one for the film maker and another one for the painter. Inside each cluster, related cultural events will be differentiated from biographical information, and so on.

### And what is meant by browsing?

For each of the examples above, part of the searching involves off-line classification, extraction and mining. All the offline extracted material will be the basis for the "MultiMatch directory", which users will be able to browse with or without posing initial queries. The structure for this directory will be induced (or closely related to) the MultiMatch ontology. See Figures 6 and 7 for examples.

This directory structure will show classified items. In addition, for certain types of data, MultiMatch will extract additional metadata information, which could be explored using faceted search or some similar approach.

### Aren't aggregated results a Wikipedia wannabe?

One substantial difference with Wikipedia is that compilation is automatically built and updated, and you can find entries (with various degrees of richness and complexity) for virtually every topic. As we will provide links to Wikipedia and other encyclopaedic sources, MultiMatch will enhance manually built resources with data mined from the web.

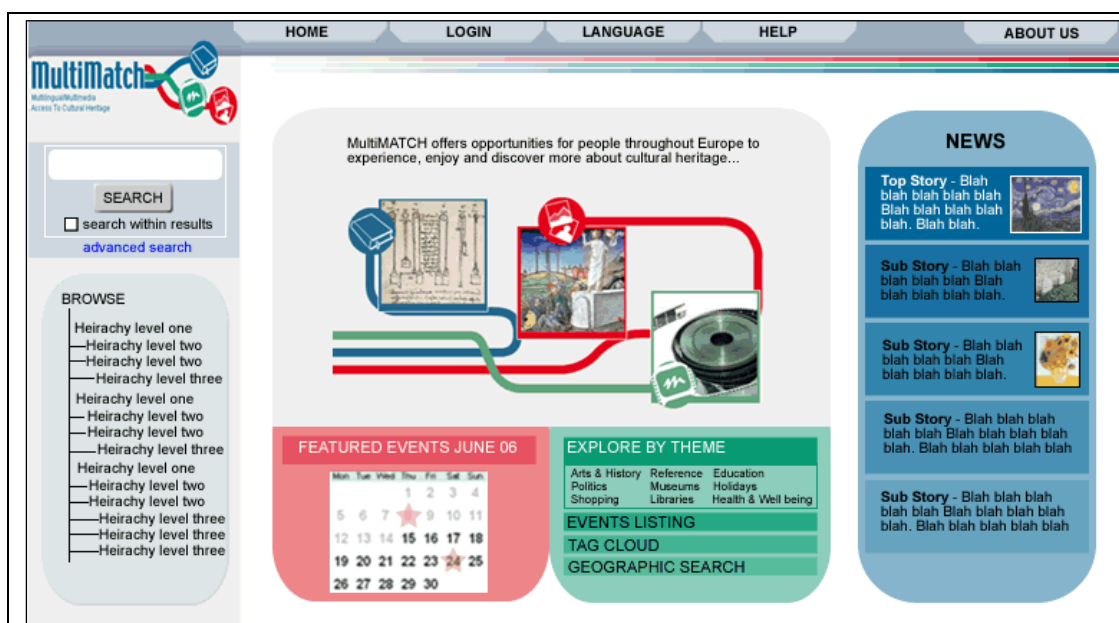


Figure 6: In this representation, browsing can be done by theme (explore by theme), location (geographic search), etc.



Figure 7: Exploration of results can be done in this example by subject, time, place or tag cloud, and the user has chosen the “by place” option.

**OK, these are the types of things we will do. But exactly what we will do? Do we expect to cover all main search scenarios and profiles?**

We will restrict ourselves to a number of specialized search capabilities and user profiles to demonstrate the validity of the approach. Determining exactly which ones is a task for WP1, where functionality will be defined using the results of interviews with experts and users. We will suggest different search facilities for each of the types above, and ask them to rate them and provide additional ones. We will define functionality by getting as close as possible to user needs in the framework of this common vision and the technical annex.

### And what about the Interface?

#### Personalization

There will be three levels of personalization:

- 1 Stereotyped user profiles (general, scholar, teacher, student, etc.)
- 2 Registered users.
- 3 Other web interfaces and programs using MultiMatch services.

For stereotyped user profiles, MultiMatch will provide interface skins.

For registered users, MultiMatch will have at least the following core personalization facilities:

- **Language:** preferred language(s), active languages, passive languages, etc.
- **Stereotyped user profile:** general/student/teacher/scholar/professional, etc
- **Default interface configuration:** where components should be aggregated in the results by default, which types of cultural material, which default media, preferred domains, etc.

For other external programs / web interfaces using MultiMatch services, we will provide an API in a “web 2.0 philosophy” so that other people can build their own mashups.

## Multilinguality and Multimediality

When presenting aggregated results, we will use different tabs for different media and languages. The philosophy is that the interface should be kept as simple as possible even when showing complex or mixed media/language results, and probably tabs are the simplest way of doing this without cluttering the search interface.

## How will the data to be indexed by MultiMatch be collected?

The importance of the metadata is increased when cultural heritage objects are digitized. For example, the metadata can improve the effectiveness and usability of the search system by providing multiple access points and preserving the semantics and context of the objects. The metadata is also critical in linking multiple versions of the same object and objects from the same collections. It can provide detailed description frameworks appropriate for each community as well as more general frameworks for resource discovery across different communities. Information for preservation and rights management can be recorded as metadata.

Metadata regarding Cultural Heritage is found in various places on the web:

1. Online catalogues (using proprietary search engines, and in effect part of the ‘deep web’). These distributed catalogues can be searched using the Z39.50 protocol;
2. Millions of web-pages (museum webpages, academic papers, WikiPedia pages, Blogs and so on);
3. Metadata that is compliant to the Open Archives Initiative Metadata Harvesting Protocol.

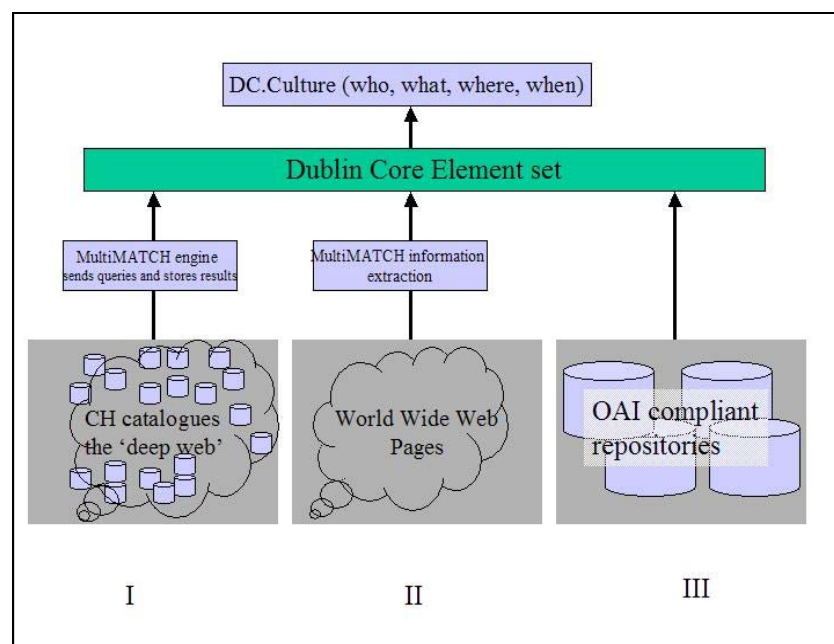


Figure 8: Finding Metadata for MultiMatch

## What MultiMatch will NOT commit to do

It is not possible within the scope of the project to touch every relevant part in the development of a search engine from a research perspective. We have to delimit clearly where we do not expect to make significant contributions to search engine development.

This is our initial list:

- We will not commit to do new research on collaborative search (e.g. search results tagging as in Yahoo's MyWeb). To evaluate this we would need a very large user community, which is unfeasible given the project schedule.
- We will not commit to improve search results using query logs or any other user activity logs. Again, to test and evaluate we would need to have a large user community, which will not be available during the project.
- We will not try to crawl every piece of information in the web related to cultural heritage, and we will not try to keep the index "fresh", i.e., constantly updated, during the project. We will crawl and index enough information to make the service usable so that it can be realistically evaluated with users.
- We will not commit to support complex queries which involve reasoning or simultaneous constraints from multiple media. Take for instance “which painters closely related to X have painted pictures similar to this?” Rather than directly posing this query, the user should, for instance, first do a content-oriented search with the painting, check the names of the artists of the paintings found by the system, then ask for painters related to X, and compare both results.



**Project no. 033104**

## **MultiMatch**

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

### **D1.2 User Requirements Analysis ANNEX I**

#### **Interview recommendations and concrete identification and size of the interview groups vs success indicators**

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

Sam H. Minelli, Alinari IDEA

Version: Final

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

---

## Introduction

This annex collects the list of the users contacted for collecting the user requirements. The interview approaches in the first phase referred to web/mail/direct phone and direct face to face. The interviews activity will also consider focus groups in a later stage and will deeper consider the non-experts.

The user evaluation generated feedback that has been used to model the user requirements.

## List of initial interview process

We identified the user groups to be reached by the partners during the first interview execution, the number of persons to be interviewed, the geographic spreading, the responsible partner and persons for the interview execution. The actual number of respondents may vary in the case where respondents are reached through mailing lists or trough the Web. Partners are expected to be in line with the assumptions written of the DoW (in particular to the page 29, success indicators).

Involved partners: Alinari, UNED, BandG, ISTI-CNR, WIND, USFD

Partner / responsibility	Target user group	Method <sup>1</sup>	Num <sup>2</sup>	Geogr. spreading
<b>UNED<sup>3</sup></b> <b>(RESP: educational)</b>	<b>Cultural Heritage</b>			
	Casa de America : Eva Martinez	Face-to-face, by phone; or e-mail	1	SP
	Instituto Cervantes: Chimo Soler and Raquel Entremozaga	Face-to-face, by phone; or e-mail	2	SP
	Thyssen Museum: Javier Espadas	Face-to-face, by phone; or e-mail	1	SP
	Miguel de Cervantes Digital Library: Rafael and Antonio	Face-to-face, by phone; or e-mail	2	SP
	Real Academia de la Lengua (RAE). María Matesanz	Face-to-face, by phone; or e-mail	1	SP
	<b>Educational</b>			
	Philology University UCM: Jose Joaquín Caerols (lecturers and researchers)	Face-to-face, by phone; or e-mail	1	SP
	Phylology University UCM: Graduate students via private distribution list	Face-to-face, by phone; or e-mail	5	SP
	CH Master at UNED: Estudents and lecturers	Face-to-face, by phone; or e-mail	5	SP

<sup>1</sup> Face to Face / by e-mail / by phone / focus group / web survey / forum / VOIP (Skype, MSN, GoogleTalk)

<sup>2</sup> Expected number of respondents, see also the success indicators table (DoW p.29): Core users are intended to be those we will recall and keep in touch with  
Education: Core:5-10; General >25  
Tourism: Core:5-10; General >10  
CH: Core:5-10; General >10

<sup>3</sup> Juan Manuel Cigarran Recuero and Julio Gonzalo

	UNED: Lecturers of Dept. Lenguajes y Sistemas Informaticos	Face-to-face, by phone; or e-mail	5	SP
	Avicenna Virtual Library. Contact person Covadonga Rodrigo.	Face-to-face, by phone; or e-mail	3	SP
	UCM. Virtual Campus. contact person Ana Fernandez Pampillon (coordinator and manager of virtual campus).	Face-to-face, by phone; or e-mail		SP
	Philology Virtual Campus UCM. Amelia Sanz (coordinator)	Face-to-face, by phone; or e-mail	3	SP
	Geography and History University UCM. Mercedes Guinea and people suggested	Face-to-face, by phone; or e-mail	1	SP
	<b>Tourism</b>			
	None		0	
<b>BandG<sup>4</sup></b> <b>(RESP: Tourism)</b>	<b>Cultural Heritage</b>			
	Sound and Vision Aad van der Valk	Face-to-face, by phone; or e-mail	18	NL
	Nava (Hungarian Audiovisual Archives)			
	ORF - Johannes Kraus			
	Rijksmuseum Peter Gorgels			
	KICH Ronald Wiemer			
	SWR (Sudwestrundfunk) Ruediger Dawideit			
	TV Catalunya			
	Sound and Vision: Aad van der Valk			
	BBC- Andy O'Dwyer			
	Munch Project - Bouke Huurink			
	CIBIT Consultants - Timo Kouwenhove			
	Sound and Vision Dalida van Dessel			
	The European Library Georgina Angelaki			
	ROB Intangible Heritage Institute Hans de Haan			
	INFOMARE Consultancy - Hanneke			
	Dutch National Archives - Jeoren Van Luin Kich			
	Digital Cultural Heritage Netherlands - Marco de Niet			
	Institute Collection Netherlands Marie-France ICN			
	<b>Educational</b>			

<sup>4</sup> Johan Oomen



	Telematica Instituut	Mettina Veenstra	Face-to-face, by phone; or e-mail	8	NL	
	University of Amsterdam	Michie				
	Vrije Universiteit	Veronique Malaise				
	University of Swansea	Alexander Rober				
	Groningen University	Andre				
	Tourism					
	Rink Klaassen Bjorn Witlox Harco Gijsberts Rosemarijn Schulte Melle de Wit Roger Holmes Marleen van den Broek Annelies Schaafsma Nostalgia Net - Judith		Face-to-face, by phone; or e-mail	9	NL	
	Alinari <sup>5</sup> (RESP: Cultural Heritage)			Cultural Heritage		
	Internal staff: Emanuela Sesti, Ilaria del Secco, Rita Scartoni, Gabriella de Polo		Face-to-face	3	IT	
	MINERVA network: Maria Sliwinska		by e-mail	2	IT, PL	
Giunti Labs, Marco Luccini		by e-mail	1	IT		
De Agostini Geografico: DeAgostini: Guido Bucciotti		Face-to-face	1	IT		
Massimiliano Pinucci		Face-to-face	1	IT		
SOTON-Matthew Addis		by e-mail	1	UK		
GettyImages Charles Merullo		Face-to-face	1	UK		
GettyImages Franziska Tayer		VOIP	1	GE		
ONB: Hans Petchar Peter Prokob		by e-mail	1	AT		
UNI TR Gianni Ramponi		by phone	1	IT		
ORF: Michael Grill		by e-mail	1	AT		
Professional users (newsagents, etc.)		Face-to-Face	3	IT		
National Polish Archives: Ewa Rosowska		by phone	1	PL		
National Polish Archives: Piotr Kanisky		e-mail	1	PL		
Educational						
Students of the university of Florence		Face-to-face, by phone	5	IT		

<sup>5</sup> Sam H. Minelli



	IUAV Andrea Ferialdi	by phone	1	IT
	U Torun professor researcher	by e-mail	2	PL
	KU Leuven professor & researchers (Fred Truyen, SaraRoigers)	by e-mail and VOIP	2	NL
	<b>Tourism</b>			
	SaturniaTours Monica T.	By phone	1	IT
DCU <sup>6</sup>	<b>Cultural Heritage</b>			
	None			
	<b>Educational</b>			
	10 reasearchers	Face to face	10	IT
	<b>Tourism</b>			
	None		0	
ISTI-CNR <sup>7</sup>	<b>Cultural Heritage</b>			
	Baldacci	Face-to-Face,	1	IT
	Biagioni	Face-to-Face,	1	IT
	<b>Educational</b>			
	researchers		2	IT
	<b>Tourism</b>			
	None		0	
USFD <sup>8</sup>	<b>Cultural Heritage</b>			
	Vanessa Toulmin, Director National Fairground Archive, Main Library University of Sheffield,	Face-to-Face, by phone	1	UK
	David Dowson	Face-to-Face, by phone	1	UK
	Kate Fernie, Museums Libraries and Archives (MLA), CH expert and educational	Face-to-Face, by phone	1	UK
	Jemima Rellie, Head of Digital Programmes, Tate Online, CH expert questionnaire (and interview).	Face-to-Face, by phone	1	UK
	<b>Educational</b>			
	Professor David Shepherd (Director of Research, Arts and Humanities Division, and Director of the HRI, University of Sheffield	Face-to-Face, by phone	2	UK
	Dr. Vanessa Toulmin, Director National Fairground Archive, Main Library University of Sheffield	Face-to-Face, by phone	1	UK

<sup>6</sup> Camonn Newmann

<sup>7</sup> Carol Peters

<sup>8</sup> Paul Clough

	Tourism			
	None		0	

### Recommendations used for drafting the email message to targeted respondents<sup>9</sup>:

- If this is technically possible: personalization will result in higher response rate
- Keep it short
- Don't mention "survey" in subject line ('research' is better)
- Signed by an authority figure
- Send during the week, best between 10-11h and 14-15h
- If this is technically / practically possible: follow-up email for non-responding targets (not for mailing lists, of course)
- Content of the message:
  - Information about partnership
  - Information on purpose of study
  - Reason why the individual's participation is important
  - Realistic time estimate to complete questionnaires
  - Why and how respondent was chosen
  - Explanation of confidentiality and how data will be handled
  - Name and phone number of contact person

<sup>9</sup> L.B. Bourque & E.P. Fielder, How to conduct self-administered and mail surveys, The survey kit, 2nd edition, part 3, California 2003; Mariet Vierns KU Leuven



**Project no. 033104**

## **MultiMatch**

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

### **D1.2 User Requirements Analysis ANNEX II**

#### **EQ: MultiMatch Experts Questionnaire for acquiring user requirements.**

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

---

## Document Information

Deliverable number:

Deliverable title: MultiMatch Experts Questionnaire for acquiring user requirements

Due date of deliverable: To be included in deliverable D1.2, user requirements specification

Actual date of deliverable: September 11, 2006

Author(s): Johan Oomen, Sam H. Minelli, Julio Gonzalo, Paul Clough, Juan Cigarran

Participant(s): BandG, Alinari, UNED, with contributions from all partners

Workpackage: WP1

Workpackage title: User requirements and functional specification

Workpackage leader: UNED

Est. person months:

Dissemination Level: Restricted

Version: Final

Keywords: functional specification, expert questionnaire, EQ, User Profile, Textual search, Image search, Video search, Desirable search facilities

## History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
1	12/07/2006	draft	Johan Oomen (BandG)	
2,3		draft	Marco Spadoni (WIND), Sam Minelli (Alinari)	Translation and questions selection
4	17/07/2006	draft	Sam Minelli (Alinari)	New adding to all the document, translations, adaptation
5	18/07/2006	draft	Paul Clough, Sam Minelli	Corrections and additions
6,7	19/07/2006	draft	Sam Minelli	updating
8	08/08/2006	Pre-final	Julio Gonzalo	Complete rewriting using input from Geneva meeting
8.1	06/09/2006	Pre-final	Juan Cigarran	Updating Sam's suggestions
8.2	08/09/2006	Pre-final	Juan Cigarran	Updating Paul and Pasquale suggestions. Less repetitive. Includes questions related with tourism promotion.
9	11/09/2006	Pre-Final	Sam Minelli, Juan Cigarran	Final update of the questionnaire and organization
Final	13/09/2006	Approved	Approved by: Emanuela Sesti, Rita Scartoni, Ilaria Del Secco (Alinari), Juan Cigarran, Julio Gonzalo (UNED), Paul Clough (USFD), external approval: Kate Fernie (MLA), Marco Spadoni (WIND), Carol Peters (ISTI-CNR)	Approval of the questionnaires by field experts.

## Abstract

This is the model questionnaire has been used when interviewing experts in the education, cultural heritage and tourism domains. It is also going to be used as guidance when conducting focused user group discussions in coming project phases. The interview methodology has been described in deliverable D1.2 User Requirements Analysis.

## Presentation

Thank you for taking the time to complete this questionnaire. The information that you provide will be used by our team to develop an understanding of how you search for cultural heritage information, and will contribute to improve the design of the MultiMatch vertical search engine for the cultural domain.







The aim of the MultiMatch project is to enable users to explore and interact with online accessible cultural heritage content, across media types and languages boundaries. The project will create a search engine that will retrieve results in multiple languages, multiple media, with multiple views of searching and browsing results, and with aggregated results from other WWW sources.

**Definition of Cultural Heritage:** *(also "national heritage" or just "heritage"):* *The legacy of physical artifacts or and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations*<sup>1</sup>. *Intellectual and artistic activity and the works produced by it*<sup>2</sup>.

**MultiMatch understanding and use of Cultural Heritage:** *Europe "collective memory includes print (books, journals, newspapers), photographs, museum objects, archival documents, audiovisual material (hereinafter 'cultural material')"*<sup>3</sup> *which are digitally stored and accessible through the Web.*

Responses that you provide to this questionnaire are confidential and will not be connected to any other information about you.

The questionnaire is divided in the following sections:

- |   |                               |   |
|---|-------------------------------|---|
|   | 1 User Profile                |  |
|  | 2 Textual search              |   |
|   | 3 Image search                |  |
|  | 4 Video search                |   |
|   | 5 Desirable search facilities |  |
|  | 6 Your comments               |   |

<sup>1</sup> See: [http://en.wikipedia.org/wiki/Cultural\\_heritage](http://en.wikipedia.org/wiki/Cultural_heritage)

<sup>2</sup> The American Heritage College Dictionary, third edition, HOUGHTON MIFFLIN COMPANY, pp. 337 Boston –New York

<sup>3</sup> EC Recommendation on the digitisation and online accessibility of cultural material and digital preservation: [http://europa.eu.int/information\\_society/activities/digital\\_libraries/index\\_en.htm](http://europa.eu.int/information_society/activities/digital_libraries/index_en.htm)

## The Questionnaire



### 1 User Profile

1.1 Your age is:

**18-30 | 30-45 | 45-60 | >60**

1.2 Your job is primarily associated with:

**Cultural Heritage | Education | Tourism**

1.3 Can you describe your job in one sentence?

1.4 How important is the WWW in your daily work?

**useless | occasionally useful | useful | essential**

1.5 What are the WWW sources of cultural information most helpful for you?

	Useless	Occasionally useful	Useful	Essential
Regular visits to specialized web sites				
RSS feeds from specialized web sites				
Newsletter subscriptions				
Databases/catalogues				
Links suggested by colleagues				
General search services				
Specialized search services				

1.6 How often do you search for cultural information?

	Never	Rarely	Monthly	Weekly	Daily
Text					
Images					
Video					



1.7 During your daily search activity which are the most used search engines? (list them by its usefulness)

For text search	For image search	For video search

1.8 What are your language skills?

	Cryptic to me	I can partially read, but I cannot write	Decent read and write	Fluent read and write	Native speaker
Dutch					
English					
French					
German					
Italian					
Portuguese					
Spanish					
Other (specify)					

1.9 What are the languages that you consider useful for getting update, comprehensive information in your daily work?

	Unnecessary	Occasionally useful	Useful	Essential
Dutch				
English				
French				
German				
Italian				
Portuguese				
Spanish				
Other (specify)				



## 2 Textual search

2.1 What are your primary sources when searching text?

	Never	Rarely	Monthly	Weekly	Daily
General web search engines (e.g. Google, Yahoo!, MSN, Ask)					
Library catalogues and other Textual databases					
Book search engines (e.g. Google print, Amazon)					
Vertical search engines (domain-specific)					
Encyclopaedic sources (e.g. Wikipedia)					
Web directories (e.g. Yahoo!, dmoz.org)					
Official web sites from museums, Universities					
Informal sources such as blogs					
News services					

2.2 Please list the two search services most useful to you:

2.3 What are the most common types of textual search for you?

	Never	Rarely	Monthly	Weekly	Daily
Search for a specific web site (e.g. "Tate Gallery")					
Search for an author (e.g. "Joseph Conrad")					
Search for an art period (e.g. "Sturm und drang")					
Search for cultural events (e.g. "Breker retrospective")					
Search for books (e.g. "Don Quixote")					
Search for news (e.g. "re-opening of Liceo in Barcelona")					
Search for scholar publications (e.g. "studies on the use of geometrical shapes in Mondrian")					

2.4 If the options above don't suit your needs, please specify your most common types of search here:





2.5 Imagine a web engine capable of automatically creating a searchable database of classified items from the text contained in cultural web pages. Which of the following databases would you find more useful?

	I don't know	Useless	Occasionally useful	Normally useful	Essential
Announcements of cultural events					
Online cultural events					
Compiled information about authors					
Compiled information about artistic periods					
Compiled information about artistic items (paintings, musical compositions, writings)					
Compiled information about cultural sites (other cultural sites near to visit, tourism and accommodation information)					
Job posts in the cultural area					

2.6 If the examples above do not fit your needs, please add your suggestions here:

2.7 Imagine a web engine capable of combining the information from many sources together and discovering patterns and trends related to cultural information. Which information would you find more useful?

	I don't know	Useless	Occasionally useful	Normally useful	Essential
compared popularity of artists/masterpieces/cultural sites					
places/people/topics most closely associated with an artist.					
artists/masterpieces/cultural sites most closely associated with a city.					
opinions/feelings/moods related to an artist/masterpiece/cultural site.					
artists/masterpieces/CH sites most closely related to a given keyword					
masterpieces most closely associated to an author					
Authors most closely associated to a given author (social network of artists)					
Things in common between two or more authors					

2.8 If the examples above do not fit your needs, please add your suggestions here:

### 3 Image search

3.1 What are your primary sources when searching images?

	Never	Rarely	Monthly	Weekly	Daily
general image search engines (e.g. Google, Yahoo!...)					
collective web repositories (e.g. Flickr)					
Classified web directories					
Cultural web sites					
Proprietary databases					
Other (specify)					

3.2 Please list the two search services most useful to you:

3.3 What are your most common types of image search?

	Never	Rarely	Monthly	Weekly	Daily
Search for artistic images					
Search for photographs of artworks (explanation: paintings, buildings, sculptures)					
Portraits					
Historic photography					
Illustrative images (for a text you are preparing)					
Other (specify)					

### 4 Video search

4.1 What are your primary sources when searching videos?

	Never	Rarely	Monthly	Weekly	Daily
general video search (e.g. Google videos)					
collective web repositories (e.g. youtube)					
Classified web directories					
Cultural web sites					
Proprietary databases					
Other (specify)					

4.2 Please list the two search services most useful to you:

4.3 What are your most common types of video search?


	Never	Rarely	Monthly	Weekly	Daily
Broadcasting of cultural events					
Artistic videos					
Documentary materials					
Educative materials					
Other (specify)					

## 5 Desirable search facilities


### 5.1 How would you rate the following search facilities?

AT QUERY TIME	Don't know	Useless	Occasionally useful	Normally useful	Essential
<b>GENERAL SEARCH (text, images and video)</b>					
Relevance feedback (explanation: "more like this image", "images with this shape")					
Spell checking (explanation: "did you mean..?")					
Search suggestions displayed as you type in the query					
Search suggestions displayed along with search results					
Filter by language					
Filter by file type (html, PDF, jpg, gif, avi, mpeg)					
Filter by date					
Filter by usage rights					
Boolean search (explanation: exclude these words, search documents including any of these words, including all of these words)					
Use of advanced fixed fields (explanation: fields specifically oriented to query for authors, artists, buildings) Similar to query a database					
Use of authority lists (explanation: lists of authors, artists, buildings) Similar to obtain a complete field database view					
<b>IMAGE AND VIDEO SEARCH</b>					
Visual relevance feedback (explanation: "more like this image", "images with this shape")					
Filter by image size					
Filter by coloration (e.g. grey scale, full color)					
Filter by length ( <b>only for video search</b> )					

WHEN OBTAINING AND DISPLAYING RESULTS	Don't know	Useless	Occasionally useful	Normally useful	Essential
<b>GENERAL SEARCH (text, images and video)</b>					
Sort by pertinence					
Sort by popularity (explanation: input from other users)					
Sort by date   title   other criteria					
Cluster results by topic (grouping related documents together)					
Graphical visualization of clustered results (e.g. kartoo.com)					
Interface customisation / personalization facilities					
Search results adjusted to your profile (personalized search)					
Timeline (explanation: results organized chronologically by periods)					
Browsing by categories or web directories					
Faceted search/browse (e.g. constraining the theme and the artistic period of a painting simultaneously)					
Aggregated results (e.g. displaying results from the web and from online encyclopaedias simultaneously)					
<b>IMAGE AND VIDEO SEARCH</b>					
search matching textual descriptions (only for video search)					
search matching subtitles when available (only for video search)					
search matching the audio (explanation: transcribed speech, when subtitles are not available) (only for video search)					
search for video fragments (only for video search)					
Cluster results by description (explanation: grouping images or videos with similar textual descriptions)					
Cluster results by visual properties (explanation: grouping images with similar colors, shapes, textures)					



WHEN HANDLING RESULTS	Don't know	Useless	Occasionally useful	Normally useful	Essential
Basket (explanation: save together items picked up from search results)					
Search history (explanation: see my previous searches)					
Annotation (explanation: annotate search results with keywords and comments)					
Collaborative search (explanation: see findings and annotations made by user communities I trust)					



## 6 Your comments

6.1 Please add any suggestions or comments on this questionnaire here:

**Thank you very much for your time answering the questionnaire.**

**If you are interested to read more about this project, visit: <http://www.multimatch.eu> .  
If would like to receive the project Newsletter please let us know by contacting: [multimatch@alinari.it](mailto:multimatch@alinari.it)**