

The MultiMatch Prototype: Multilingual/Multimedia Search for Cultural Heritage Objects*

Giuseppe Amato, Franca Debole, Carol Peters, and Pasquale Savino

ISTI-CNR, Pisa (I)
firstname.lastname@isti.cnr.it

Abstract. MultiMatch is a 30 month targeted research project under the Sixth Framework Programme, supported by the unit for Content, Learning and Cultural Heritage (Digicult) of the Information Society DG. MultiMatch is developing a multimedia/multilingual search engine designed specifically for the access, organization and personalized presentation of cultural heritage information. The demonstration will present the MultiMatch system prototype.

1 Introduction

Europe's vast collections of unique and exciting cultural content are an important asset of our society. On the web, cultural heritage (CH) content is everywhere, in traditional environments such as libraries, museums, galleries and audiovisual archives, but also reviews in popular magazines and newspapers, in multiple languages and multiple media.

What means do users have to access these complex CH objects? How can they explore and interact with CH content? Currently, users are left to discover, interpret, and aggregate material of interest themselves. The MultiMatch Search Engine (<http://www.multimatch.org/>) is a first attempt to provide a complete and integrated to solution to search CH content. It supports retrieval of cultural objects through different modalities:

- The simplest one is a traditional *free text* search. This search mode is similar to that provided by general purpose search engines, such as Google, with the difference that MultiMatch is expected to provide more precise results – since information is acquired from selected sources containing Cultural Heritage data – and with support for multilingual searches. Users can formulate queries in a given language and retrieve results in one or all languages covered by the prototype (English, Italian, Spanish, Dutch, German, and Polish)

* Work partially supported by the European Community under the Information Society Technologies (IST) programme of the 6FP MultiMatch contract IST 033104. The authors are solely responsible for the content of this paper. It does not represent the opinion of the EC and the EC is not responsible for any use that might be made of data appearing therein.

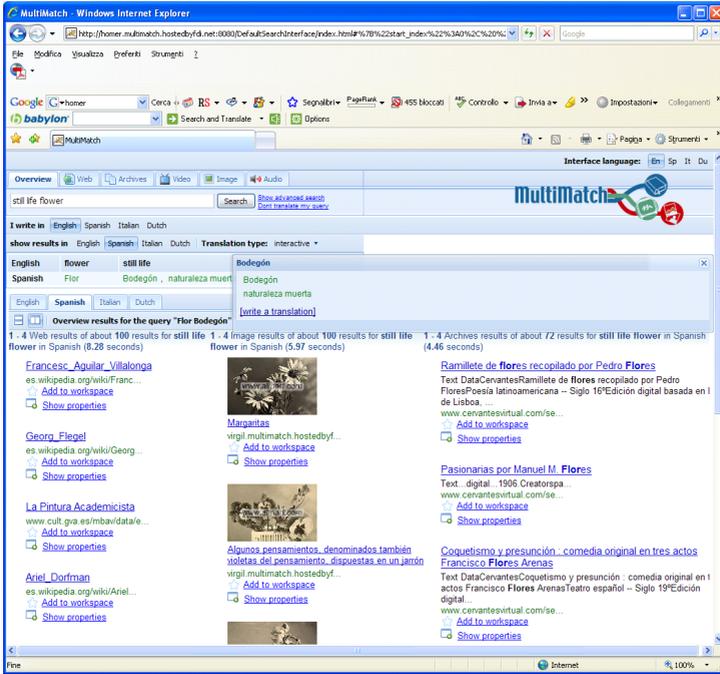


Fig. 1. Example of multilingual search

according to their preferences. Multilingual searches are performed by a combination of machine translation and domain-specific dictionary components. Users can select the source and the target languages as well as the most appropriate translation among those proposed by the system (see Figure 1).

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

The main source of information stored in the MultiMatch prototype system is composed of cultural heritage objects obtained through crawling and indexing of material obtained from cultural heritage sites (e.g. museums), web encyclopedias (e.g. Wikipedia), digital libraries of specific cultural heritage organizations, OAI compliant digital resources, and RSS feeds from cultural web sites.

MultiMatch searches can be made in three modes: (i) Default search mode, (ii) Specialized search mode, and (iii) Composite search mode.

The simplest search mode is the *default* MultiMatch search level. This is provided for generic users, with a limited knowledge of MultiMatch system capabilities, or with very general search needs. In this case, no assumption is

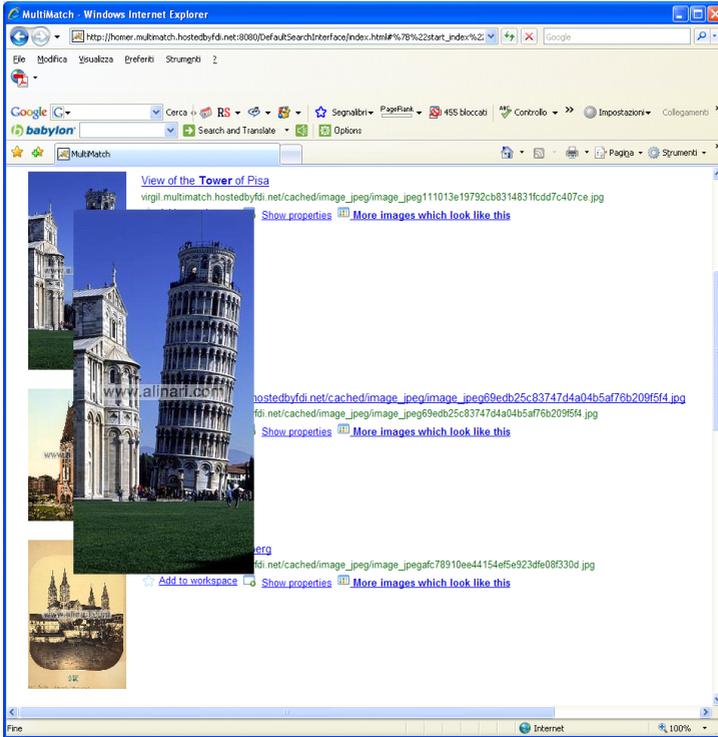


Fig. 2. Example of image similarity search

made on the user query, and MultiMatch will retrieve all the cultural objects, web pages and multimedia content that best suit the query. Merging, ranking and classification of these results will be also performed by the system.

Users with a more precise knowledge of MultiMatch system functionality, and with specific search needs, may use one of the *specialized* interaction levels available. These allow the user to query MultiMatch in-depth search services (e.g., video search, image search, etc.) and retrieve all information available via the selected search service. In this way, MultiMatch will include standalone image, video and web-based searches, each with its own search fields, display and refinement options. It also includes a set of browsing capabilities to explore MultiMatch content. Figure 2 gives an example of image similarity search.

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

The demonstration will illustrate all the search functionality of the MultiMatch prototype system, from simple text searches (mono and multilingual), to image, audio and video search.