MULTIMATCH
A Vertical Search Engine for European Cultural Heritage

Pasquale Savino
Giuseppe Amato
Carol Peters
ISTI – CNR
pasquale.savino@isti.cnr.it

http://www.multimatch.org
MultiMatch: The Initial Idea

Problem:

- The Web contains a wealth of fragmented CH information stored in web pages as well as in Digital Libraries managed by cultural institutions.
- This information is dispersed and fragmented.
- Users are left to discover, interpret, and aggregate it.
A user perspective

- Information needed is likely to be available on the Internet but to access it, language, media, and source boundaries need to be overcome.

- Key ideas underlying the MultiMatch vision
  - Multiplicity
    - Queries and retrieved items in multiple languages
    - Queries and retrieved items in multiple media
    - Items selected from multiple sources
  - Aggregation
    - Presentation of aggregated results
    - Relationships between retrieved items
  - Personalization
What is supported by existing systems?

- **Google-like search**
  - **Pros**
    » Large content-base
    » High recall
  - **Cons**
    » Low precision
    » No multilingual support
    » Single media search
    » No relationship between items

- **Access to Specialized CH sites**
  - Very precise results but limited to a single site
  - Results are presented in an attractive way, but cannot be adapted to specific user requests
MultiMatch Objectives

- Develop a search engine that provides targeted, enriched access to heterogeneous CH objects
  - Across all media types and language boundaries
  - Gathering information from the web and from CH sites and digital libraries
  - Enable the interoperability of CH Digital Libraries based on existing standards
  - Supporting various user classes and offering personalized access
  - with aggregate views on complex task scenarios

- Assist CH institutions to raise visibility and disseminate content
Web Resources:
- Museums
- Libraries
- Archives
- Newspapers
- News agencies
- Personal Pages
- Blogs

Museums Databases

Museums:
- Van Gogh Museum (NL)
- National Gallery (UK)
- Musée d'Orsay (F)

Acquisition

Crawling
What MM intends to achieve

- MultiMatch is developing a **Web search engine** specialized in the **cultural heritage** domain.
- Queries can be expressed in **multiple languages**
- **Multilingual and multimedia retrieval**
- **Access to multiple sources of information**
  - Web sites containing authoritative cultural heritage data (e.g. museums, cultural institutions, CH educational sites, tourist information portals)
  - IPR protected CH material, provided by content providers (e.g. Alinari, Sound and Vision, Biblioteca Virtual de Cervantes)
  - Other OAI compliant resources
Default search functionality
The Project

- VI FP Project
- Technology-enhanced Learning and Access to Cultural Heritage
- Started: May 2006
- First prototype: August 2007
- Second Prototype: July 2008
- Evaluation and field trials: October 2008
MultiMatch Partners

- **Cultural Heritage**
  - Fratelli Alinari Istituto Edizioni Artistiche SpA (Alinari)
  - Netherlands Institute for Sound and Vision (Sound and Vision)
  - University of Alicante – Biblioteca Miguel de Cervantes (UA-BVMC)

- **Industry**
  - OCLC PICA (FDI)
  - WIND Telecomunicazioni S.p.A. (WIND)

- **Academia**
  - Istituto di Scienza e Tecnologie dell’Informazione (ISTI-CNR)
  - University of Sheffield (USFD)
  - Dublin City University (DCU)
  - University of Amsterdam (UvA)
  - University of Geneva (UniGE)
  - Universidad Nacional de Educación a Distancia (UNED)
Main Research Challenges

From document to complex objects retrieval

- **Focused crawling** for acquisition of CH-related information from heterogeneous MM resources
- Semantic indexing of CH **content and relation extraction** using information extraction and text mining techniques
- **Multimedia search** and mixed media search
- **Multilingual management** with support for query formulation, cross-language retrieval and summarization
- Integration and representation of **related objects**
- Presentation of **aggregated search results**
- **User support** (e.g. search history, annotation facilities, personalised presentation of results, etc.) for different communities of users
Metadata modelling and interoperability

- Definition of a MultiMatch conceptual reference framework suitable for Cultural Heritage
  - Metadata schema
  - Thesauri for artists names and descriptions, geospatial information
- Simple to use (interoperability)
- Suitable to the application of automatic population techniques
Indexing and information extraction

- Automatic extraction of indexing features for all media (text, speech, images, video) and crawled data
- Text indexing for four languages (Italian, Spanish, Dutch, and English, plus others)
- Automatic generation of inter-document links
- Development of algorithms for classification and information extraction
  - Creators, type/genre, subject, place/time, art objects/works
- Semantic enrichment of documents
- Linking information sources
  - Thesauri from outside the system
  - Documents within the system
Multilingual support

- Provide system with monolingual and multilingual search functionalities (initially four languages, to be extended)
- Provide effective translation strategies e.g. multilingual dictionaries, machine translation, thesaurus term expansion
- Multilingual query expansion
- Dynamic summarization
Multimedia search

- Similarity search based on visual features (low level and high level – faces, objects, etc.)
- Efficient retrieval
- Support of relevance feedback and interactive search
- Combined text and visual search
User interaction and interface design

- User-centred design process
  - Evaluate and refine the interface based on empirical evaluation and usability testing
- Interface supporting multilingual and multimedia search
- Default search on all types of content
- Specialized search on metadata fields and on different media
- Use of semantic structures for search and browse
Project achievements

- First project phase completed
  - User requirements analysis
  - MultiMatch conceptual reference framework for Cultural Heritage
  - First prototype of the MultiMatch search engine
    » Baseline for experimentation, user feedback and second prototype development
    » Limited content
    » Based on manually associated semantic concepts
    » No relationship between retrieved objects
Research issues for the second project phase

- Automatic document labeling, enrichment and linking and automatic document classification
- Combination of cross media query results
- Translation and summarization of results
- Innovative User Interface
  - Multimedia search and browse
  - Presentation of video material
  - Multilingual interaction
  - Optimized search and browse based on semantic structures
Conclusions

- Further evolutions of MultiMatch
  - Use of MultiMatch technologies to build large scale Digital Libraries and a large scale search engine specialized for Cultural Heritage
    - Enlarge the content base through the access to a complete set of CH sites and crawling of a significant part of the web
    - Enlarge the number of languages managed (possibly to cover all EU languages)
    - Invest on system efficiency, system scalability, and robustness
Conclusions

- Significant investment in dissemination of project results, in order to attract cultural heritage institutions
- Dissemination of MultiMatch technology into other related application fields
- Further investment on the key research topics addressed in MultiMatch
Further information

- MultiMatch Web site
  - http://www.multimatch.org

- Project coordination: Carol Peters
  - carol@isti.cnr.it

- Technical coordination: Pasquale Savino
  - pasquale.savino@isti.cnr.it

- User group coordination: Sam Minelli
  - sam@alinari.it
Results for "van gogh": 4 biography pages | 4 artworks-related pages | 8 reviews pages | 2 news pages | 2 Not categorized sites

Categorized Results

Profile

Name: Vincent Van Gogh
Birth: 1853
Death: 1890
Country: Holland
Category: Painter

Biographical Data

WebMuseum. Gogh. Vincent van Gogh
His father was a Protestant pastor, and van Gogh first trained for the ministry. ... for the 20th-century visitor to romanti...eological biography. ... www.bibliothek.uni/biologie/van_gogh/ - 13k - Cached - Similar pages

Vincent van Gogh Online

Tate Collection. Vincent van Gogh
Vincent van Gogh 1853-1890. Dutch painter of landscapes, figures and still life. Born at Groot-Zundert, son of a parson. Entered the firm of Goupil, ... www.tate.org.uk/servlet/ArtistsWochtlycatid=6999999999999... - 18k - Cached - Similar pages

ArtLex. Post-Impressionist Vincent van Gogh
Post-Impressionist Vincent van Gogh, with images of his works, great quotations ... for the second half of the 19th century vogue in romanticized psychological biography. ... www.artlex.com/Artists/PostImp/vangogh.html - 41k - Cached - Similar pages

News

Van Gogh dispute is coming to a head
Sydney Morning Herald, Australia. Oct 11, 2008 ...
... of Victoria, Dr Gerard Vaughan, is optimistic that research under way in Amsterdam will confirm the authenticity of the gallery's only Van Gogh painting. Head ...

Van Gogh exhibit focuses on wheat, harvest themes
KXNO, OK. Oct 21, 2008
DALLAS A vibrant painting Vincent van Gogh finished weeks before his death is the centerpiece of a new exhibit at the Dallas Museum of Art ...

Artworks

Vincent van Gogh Online
Vincent van Gogh (Dutch Post-Impressionist Painter, 1853-1890) Guide to pictures ... Kunst Institutes Den Haag - database of artworks in Dutch museums ... www.artyclopedia.com/artists/van_gogh_vincent.html - Similar pages

Artyclopedia. Monthly Spotlight Vincent van Gogh
The featured artist is Vincent van Gogh. Click here for a new complete list of sites where his artwork can be seen on the Internet. Portrait of Dr. Gachet ... www.artyclopedia.com/feature-mar99.html - Similar pages

Famous Artworks by Theme: Mosaic Paintings
Famous Artworks by Theme: Mosaic Paintings ... Vincent van Gogh. Wheat Field with Crows (Probably van Gogh's last painting). Artposters.com ... www.artyclopedia.com/artists/famous-mosaic.html - Similar pages

Artקלאסיקה. Featured Online Exhibits: Vincent van Gogh...
Vincent van Gogh Information Gallery ... presents the artworks in a professional manner: by listing the date of each work and its current location ... www.artlections.com/features-exhibits-mar09.html - Similar pages

Reviews

Your Reviews of Art Shows and Exhibitions
Your reviews critique of art shows and art exhibitions you did or didn't like. ... The myth of Van Gogh - the tormented artist shunned by society. ... www.westcolgallery.co.uk/quick- content/subpage_id=Van+Gogh+Museum&... - 29k - Cached - Similar pages

Van Gogh and Britain - Pioneer Collectors at Compton Verney...
A contemporary review in the Daily Telegraph asked if the artists were "masters or pioneers?" and described Van Gogh as "the fiery spirit of revolt" ...
www.24hoursmuseum.org.uk/birmingham/news/art/358... - 37k - Cached - Similar pages

Result Page: 1 | 2 | 3 | Next

Other

Vincent Van Gogh: Two Sunflowers, 1887

Van Gogh Museum Press
Van Gogh Museum offers an exhibition about that of Paintings. The Van Gogh Museum is offering a substantial reward of up to €10,000 for ... www.vangoghmuseum.nl/1996- nieuws/video/nieuws_view&d=111&cat_id=4 - 9k - Cached - Similar pages

Copyright MultiMatch 2006

Author search: Classified Web page results
Relations between Author and works of art
Author and related CH sites

[Map of Europe with locations marked]

Related Sites
- Detroit Institute of Arts
- Michigan
- Fine Arts Museums of San Francisco
- San Francisco
- Fitzwilliam Museum at the University of Cambridge
- United Kingdom
- Guggenheim Museum
- New York
- Hermitage Museum
- Saint Petersburg
- Vincent van Gogh at the Rijksmuseum
- Amsterdam
- Neue Pinakothek
- Munich
- Statale Essen
- Fortius
Specialized image search
Multilingual Search (2)
Multilingual Search (3)
Multilingual Search (4)