

# REVEAL THIS : RETRIEVAL OF MULTIMEDIA MULTILINGUAL CONTENT FOR THE HOME USER IN AN INFORMATION SOCIETY

Stelios Piperidis\*, Harris Papageorgiou \*

\*Institute for Language and Speech Processing / IRIS  
Artemidos 6 & Epidavrou, Athens, Greece  
{spip, [xaris](mailto:xaris@ilsp.gr)}@ilsp.gr

**Keywords:** audio-image-text analysis, cross-media linking and indexing, cross-media categorisation, cross-media summarisation, cross-lingual translation.

In order to exploit the opportunities offered by the digital media market, content providers need to apply new business models, content representations and CRMs.

## Abstract

REVEAL THIS addresses a basic need underlying content organisation, filtering, consumption and enjoyment by developing content programming systems that will help European citizens keep up with the explosion of digital content scattered over different platforms (radio, TV, World Wide Web, etc), different media (speech, text, image, video) and different languages. *REVEAL THIS aims at developing content programming technology able to capture, semantically index, categorise and cross-link multiplatform, multimedia and multilingual digital content, as well as provide the system user with semantic search, retrieval, summarisation and translation functionalities.* The innovative aspects of the project spring out of the main scientific and technological challenges: 1) semantic enrichment of multilingual multimedia content with topic, entity and fact information relevant to user profiles; 2) development of suitable cross-language, cross-media representations; and 3) deployment of the above in building search, retrieval, classification and summarization capabilities. We exploit explicit cross-media links and implicit links uncovered by methods such as latent semantic analysis and kernel canonical correlation analysis. Adequate cross-language capabilities (cross-language information retrieval, categorization and machine translation of indicative summaries) will be provided by the latest statistical machine translation technology.

## 1 Introduction

Digital media assets are proliferating and most organizations, large broadcasters as well as SMEs are building networks and technology to exploit them. The cross-media market for broadband media will, according to predictions, develop rapidly within the next 10 years. Traditional broadcasters, publishers and Internet content providers will migrate into increasingly similar roles as multimedia content providers. Fresh cross-media players are also entering the market, aiming to exploit different distribution channels (IP-based, mobile) and to reach personalized and fragmented audiences.

Digital technology today allows the user to manipulate or interact with content in ways not possible in the past. The combination of PCs and networks allows the individual to create, edit, transmit, share, aggregate, personalize and interact with multimedia content in increasingly flexible ways. The same technology allows content to be carried across different platforms. In fact, much of the information that reaches the user nowadays is in digital form: digital radio, music CDs, MP3 files, digital satellite and digital terrestrial TV, personal digital pictures and videos and, last but not least, digital information accessed through the Web. This information is heterogeneous, multimedia and, increasingly, multi-lingual in nature.

The development of methods and tools for content-based organization and filtering of this large amount of multimedia information reaching the user through many and different channels is a key issue for its effective consumption and enjoyment. In fact, despite recent technological progress in the new media and the Internet, the key issue remains “how digital technology will add value to present information channels and systems”.

## 2 Objectives

The main objective of the REVEAL THIS project is the design, development and testing of an integrated infrastructure that allows the user to store, categorize and retrieve multimedia and multi-lingual digital content across different sources (TV, radio, music, Web), with a view to personalize the user experience with these sources.

To achieve the main objective, the REVEAL THIS project has set the following measurable innovative technological & scientific objectives:

1. Augment the content of multimedia documents with semantic information like: characteristic keyframes, identified speakers, faces, entities, topics and fact information.

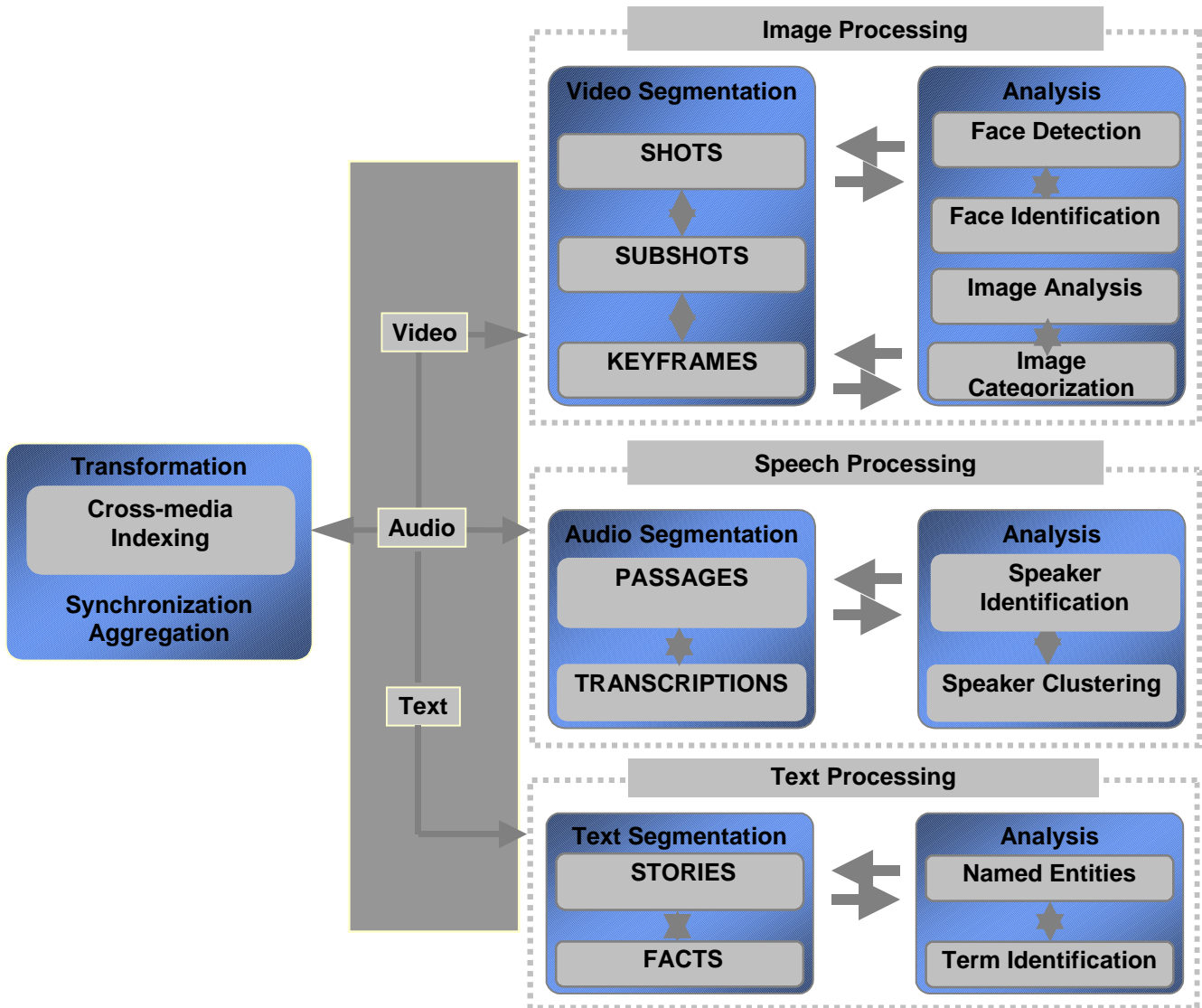
2. Develop cross-media and cross-language representations.
3. Build high-level functionalities, namely categorization, and summarization, from 1 and 2.
4. Provide cross-lingual capabilities (cross-lingual information retrieval, categorization and machine translation of indicative summaries) based on the latest statistical machine translation technology.
5. Integrate the above technologies in a unified platform equipped with semantic search, retrieval engine, personalization and delivery mechanisms to at least two different devices (mobiles and PCs).

These objectives aim at providing a technology suite that will enable the development of a fully operational personalized entertainment system. **Such a system can be used** (a) by **content providers**, to add value to their content, restructure and re-purpose it and offer their subscribers, individual or

corporate users, personalized content, or (b) **directly by end users**, for gathering, filtering and categorizing information collected from a wide variety of sources in accordance with user preferences.

In order to achieve the first scientific objective and effectively render possible the above functionalities, it is necessary to provide additional indices that pertain to:

- **text:** named entities (e.g. names of persons, places, organizations), terms, topics and facts
- **speech:** speakers (e.g. speaker identity), transcriptions and stories
- **video and images:** persons, faces, keyframes and objects.



**Figure 1.** REVEAL THIS Cross-media Content Analysis and Indexing Subsystem.

REVEAL THIS deploys state-of-the-art technologies and components, namely indexing techniques for each media (Figure 1):

- Speech processing component – automatic speech recognition, automatic speaker identification.
- Image analysis component – shots, keyframes, objects, characteristic patches.
- Face analysis component – face recognition, face identification.
- Text processing component – entity, term and fact extraction, topic detection, story segmentation.

The metadata/indices produced by the above components will be aligned, synchronized and encoded in XML. Once documents<sup>1</sup> have been enriched and indexed, a major challenge lies in developing *representations* suitable for crossing media and languages in the processes of *retrieval*, *categorization* and *summarization*. To achieve this second scientific objective, REVEAL THIS exploits the links between different media explicated in each document. For example, images are sometimes associated with captions, from which textual indexes can be extracted. A textual query can then be issued to retrieve images. Similarly, speech recognition directly allows retrieving written and spoken documents from written or spoken queries. A promising approach to alleviate imprecision lies in exploiting similarities between media in each document and deriving semantic representations that are “naturally” linked across media supported by evidence provided by the different processing components. REVEAL THIS investigates two different, promising models based on probabilistic latent semantic analysis and kernel canonical correlation analysis. The envisaged technology is provided as a separate component, namely Cross-media Indexing component – CMIC (links between different multimedia objects across media).

Based on the above indexical information and representations, REVEAL THIS deploys state-of-the-art technologies for categorizing and summarizing documents.

However, the diversity of the documents reaching the citizens of an integrated, interconnected Europe does not only lie on the different media used, but also on the different *languages* through which content is mediated. A content-based solution that really addresses European needs to seriously tackle the language issue and propose components that can not only work on different languages, but also establish bridges across languages. REVEAL THIS develops a cross-lingual translation subsystem equipped with state-of-the-art components capable of automatically extracting multilingual glossaries from specific collections via word/term alignment techniques, so as to complement existing multilingual resources in the provision of domain-tuned retrieval and

categorization technologies. Furthermore, REVEAL THIS will have the capability to present summaries of documents in different languages. To this end, we resort to state-of-the-art statistical machine translation models and investigate their tight coupling with summarization, across both text and speech.

Finally, REVEAL THIS integrates the above technologies in a unified infrastructure equipped with retrieval, semantic search based on RDF representations, personalization and delivery mechanisms to at least two different devices (mobiles and PCs). In this context, state-of-the-art techniques of distributed information retrieval, related to multimedia resource selection, data fusion and presentation of results, are coupled with multi-document summarization and hierarchical categorization to enable the user to effectively search and browse the large amount of multimedia content gathered.

### 3 System Architecture

The REVEAL THIS system architecture is depicted in Figure 2. The main subsystems are (i) the Cross-media Content Analysis & Indexing Subsystem, (ii) the Cross-media Categorisation Subsystem, (iii) the Cross-media Summarisation Subsystem, (iv) the Cross-lingual Translation Subsystem, and (v) the Cross-media Content Access and Retrieval Subsystem.

In summary, the REVEAL THIS archival system has the following general functionality:

- Provides robust audiovisual content analysis including spatio-temporal video segmentation, face detection and identification, automatic generation of spoken transcriptions and speaker identification, automatic extraction of textual metadata like named entities, terms, stories, topics, facts, as well as links between different media. (*Cross-media content analysis* components).
- Supports cross-media summarisation by exploiting cross-media information and fusing video, audio and textual metadata to determine the most salient parts that might be of interest to the user (*Cross-media content summarisation subsystem*).
- Supports cross-media categorisation of multimedia objects and when applicable of multilingual content (*Cross-media content categorisation subsystem*). This subsystem incorporates three uni-modal categorizers: (a) face-based categorization, (b) image categorization, and (c) text-based categorization.
- Provides cross-lingual functionalities such as querying multimedia objects written in different languages, categorizing multilingual and multimedia content as also producing summaries in the user’s language (*Cross-lingual translation subsystem*).

---

<sup>1</sup> When we use the term document, we implicitly refer to a multimedia document, written in, spoken in and/or associated with one or more languages.

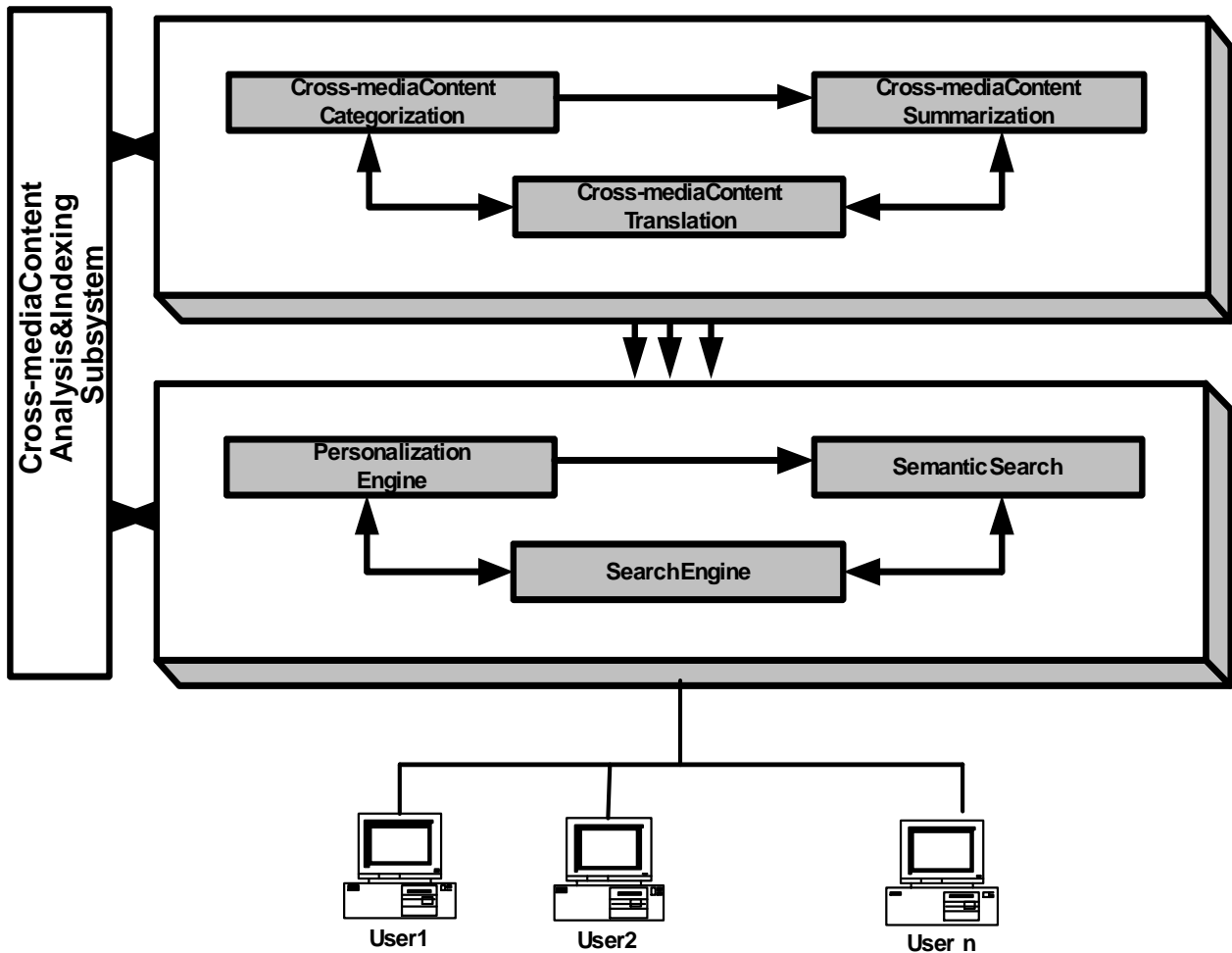


Figure 2. REVEAL THIS system architecture.

- Provides an integrated system permitting collection, storage and management of audiovisual content and related metadata, as well as user management, streaming and distribution of content depending on user profiles and preferred devices (*Reveal This System*).
- Supports content access and retrieval services to home users or operators, including content browsing, semantic search based on RDF representations built by the cross-media content indexing, personalized response, preview of retrieved video segments and real-time delivery of the audiovisual content and related metadata (*Content Access and Retrieval subsystem*).
- Adopts the general features and description for object-based media coding and multimedia content description related to the MPEG-7 and XML standardisation activities. In particular, it conforms to the normative MPEG-7 description schemes (DSs).

### Acknowledgements

The REVEAL THIS project is funded by the FP6-IST programme of the European Commission, contract number FP6-IST-511689 and is designed and implemented by the REVEAL THIS consortium comprising Institute for Speech and Language Processing / IRIS (Co-ordinator), SAIL LABS Technology AG, Xerox - The Document Company S.A.S, Katholieke Universiteit Leuven R&D, University of Strathclyde, BeTV SA and TVEyes UK Ltd.

### Bibliography

Hauptmann, A., Smith, M. (1995): "Text, Speech, and Vision for Video Segmentation: The Informedia Project", In AAAI Fall 1995 Symposium on Computational Models for Integrating Language and Vision, 1995

T. S. Huang, S. Mehrotra, and K. Ramchandran, (1996) "Multimedia Analysis and Retrieval System (MARS) Project," in Proc of 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval

J.R. Smith and S.-F. Chang (1997) "An image and video search engine for the world-wide web", Proc. of SPIE, vol. 3022, pp 85-95

"MAESTRO: conductor of multimedia analysis technologies" (2000) Communications of the ACM, 43(2): 57 - 63

IEEE Multimedia (2002), Special issue: Content-Based Multimedia Indexing and Retrieval, Vol 9, No 2, pp. 18-60.

Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, Ibrahim Sezan, Thomas Huang, Avidesh Zakhor (2002) "Applications of Video-Content Analysis and Retrieval", In IEEE MULTIMEDIA, Vol. 9, No. 3; July-September, pp. 42-55

Papageorgiou. H., P. Prokopidis, I. Demiros, N. Hatzigeorgiou, G.Carayannis (2004), "CIMWOS: A Multimedia Retrieval system based on combined Text, Speech & Image Processing" at RIAO 2004 "Coupling approaches, coupling media and coupling languages for information retrieval", 26-28 April, Avignon, France.

M. Wallace, Y. Avrithis, G. Stamou and S. Kollias, "Knowledge-based Multimedia Content Indexing and Retrieval", in Multimedia Content and Semantic Web: Methods, Standards and Tools, G. Stamou and S. Kollias (editors), Wiley, 2004.