Enhancing Enterprise Knowledge Processes via Cross-Media Extraction

José Iria
The University of Sheffield
211 Portobello Street
Sheffield, S1 4DP, UK
j.iria@sheffield.ac.uk

Victoria Uren
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA,UK
v.s.uren@open.ac.uk

Categories and Subject Descriptors
H.2.4 [**Systems**]: Multimedia databases; H.3.3 [**Information Search and Retrieval**]; I.2.1 [**Applications and Expert Systems**]: Office automation.

General Terms
Algorithms, Design, Human-Factors

Keywords
Cross-media knowledge extraction, Large-scale datasets, Industrial applications

1. INTRODUCTION
The resources needed to solve problems are typically dispersed over systems within the company, and also in different media. For example, to diagnose the cause of failure of a component, engineers may need to gather together images of similar components, the reports that summarize past solutions, raw data obtained from experiments on the materials, and so on. Considerable effort is spent just to gather such information. In the X-Media project[1] we are investigating the potential of rich semantic metadata to connect up dispersed resources across repositories and media, in order to support knowledge reuse and sharing.

Automatic capture of semantic metadata is available for single medium scenarios. However, there is a need for extraction methods that can capture evidence for a fact from across different media. In many cases the different media must be examined simulaneously to get enough evidence and improve the quality and depth of the extracted knowledge. In this paper, we present initial work on a cross-media knowledge extraction framework specifically designed to handle large volumes of documents composed of three types of media – text, images and raw data – and to enable capturing evidence across media.

[1]http://www.x-media-project.org

1.1 Use Case Scenario
We have collected requirements from our industrial partners using a user centred design process [3]: the car manufacturer FIAT and the aerospace manufacturer Rolls-Royce. We briefly describe one scenario here, defined in cooperation with FIAT, which concerns forecasting the launch of competitors' models. The goal is to collect information about the features of competitors' vehicles from various data sources, so as to produce a calendar that illustrates prospective launches. The required data is to be found scattered throughout the Internet (e.g. in blogs and forums), and covered by international automotive magazines as well by a long tail of automotive national magazines. The collected information is used in the *Set up* stage of new FIAT vehicles (the development stage where a first assessment of the future vehicle's features is carried out). This process is of great value to the company as it contributes to keeping vehicle design up to date with competitors.

We are developing end-user systems able to track knowledge changes and of being proactive in supporting knowledge workers during the *Set up* stage. To support these systems, the underlying knowledge extraction systems need to be able to handle such rapidly evolving multimedia data sources on a large scale. Typically, documents contain complementary information across the media. For example, a document may contain photographs of the front part of the interior of a Toyota Yaris car along with text describing the depicted car components. End-user systems are being built that support issuing queries over the extracted knowledge, e.g. "find competitor car models with ergonomic air ducts". The desired output of such systems for this query would be to present Yaris as a potentially interesting model and provide the worker with a set of images and text snippets, including the ones in the document shown. In order to achieve that, knowledge extraction systems must gather evidence from across the media: on the one hand, identification of the car model depicted in the images can only be done using the text, which explicitly mentions "Yaris"; on the other hand, identification of some of the car model components such as air ducts, steering wheel and gear lever can only be done using the images, since the text only mentions glove box, tray, pockets, bins and cup-holders.

2. KNOWLEDGE EXTRACTION FRAMEWORK

The major requirements for the framework identified from the analysis of the use cases were the ability to exploit evidence for a fact across several media, and the ability to perform the extraction on a large scale. Other requirements that also have implications in design decisions were the ability to exploit background knowledge, portability and the ability to report uncertainty.

The proposed framework consists of three main components:

(i) A multimedia daemon that handles the content related tasks such as dismantling compound documents, enabling fast access of indexed data and making transparent to the rest of framework the data format variety.

(ii) A knowledge extraction processor that operates on the output of the media manager with the aim to provide an interpretation of the content semantics. The employment of low complexity, fast ML algorithms for single medium extraction and concept modeling is mandated by the need to handle large scale datasets.

(iii) A knowledge base that facilitates the storage, retrieval and inference of knowledge.

The knowledge extraction processor receives as input a multimedia document (e.g. a failure report) and produces semantic annotations with a set of inferred concepts. It is divided into the following steps: multimedia document processing, integration of single-medium and cross-media information, background knowledge.

The Multimedia Document Processing step extracts single-medium elements and their relations from the compound document. Document processing literature discusses several approaches to extract layout information from PDF, HTML and other structured documents, see [1] for an overview. Cross-media KA algorithms process both the content from the different modalities and the layout information.

As mentioned, a multimedia document may contain evidence for a fact across different media. However, it is not straightforward to know which media elements refer to the same fact. The document layout and extracted cross-references (e.g. captions) can suggest about how each text paragraph/segment relate to each image/raw data. We extract a set of cross-media features which include: layout structure, distance between segments, cross-references, same type of font, font colour, and background colour/pattern. These feature sets coming from text, images and layout have very different characteristics in terms of sparseness. We follow Magalhães and Rüger [2] and process text and image independently with probabilistic latent semantic indexing to produce a canonical representation of both text feature space and image/raw data feature space. This allows statistical learning algorithms to handle different types of data simultaneously more easily. Once the feature data has been processed, some modelling algorithm can be used to create the knowledge models of all concepts. Special care was taken when designing the algorithm to model each concept: it must support high-dimensional data, hundreds of thousands of examples, and low computational complexity. Finally, semantic metadata provide information about concepts co-occurrence and how they co-occur across different modalities, which the algorithm exploits in order to enhance the model of each individual concept.

3. ACKNOWLEDGMENTS

This work was funded by the X-Media project (www.x-media-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-026978.

4. ADDITIONAL AUTHORS

Alberto Lavelli (F. Bruno Kessler)
Sebastian Blohm (U. Karlsruhe),
Aba-sah Dadzie (U. Sheffield),
Thomas Franz (U. Koblenz-Landau),
Ioannis Kompatsiaris (CERTH),
João Magalhães (U. Sheffield),
Spiros Nikolopoulos (CERTH),
Christine Preisach (U. Hildesheim),
Piercarlo Slavazza (Quinary).

5. REFERENCES

[1] A. Laender, B. Ribeiro-Neto, A. Silva, and J. Teixeira. A brief survey of web data extraction tools. In *SIGMOD Record*, volume 31, June 2002.

[2] J. Magalhães and S. Rüger. Information-theoretic semantic multimedia indexing. In *ACM Conference on Image and Video Retrieval (CIVR)*, Amsterdam, Holland, 2007.

[3] M. Rosson and J. Carroll. *Usability engineering: scenario-based development of human-computer interaction.* Morgan-Kaufman, 2002.