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 simultaneously 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

K-CAP'07, October 28–31, 2007, Whistler, British Columbia, Canada.
ACM 978-1-59593-643-1/07/0010

2. USE CASE SCENARIO
We have collected requirements from our industrial partners, the car manufacturer FIAT and the aerospace manufacturer Rolls-Royce, using a user centred design process [5]. 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. The required data is to be found scattered throughout the Internet (e.g. in blogs and forums), and covered by international and national automotive 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.

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 can issue queries over the extracted knowledge, e.g. "find competitor car models with ergonomic air ducts". The desired output for this query would be to present Yaris as a potentially interesting model through a set of relevant images and text snippets. For that, knowledge extraction systems must gather evidence from across the media. In our example, identification of the car model depicted in the images can only be done using the text, which explicitly mentions "Yaris", while 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, and pockets.

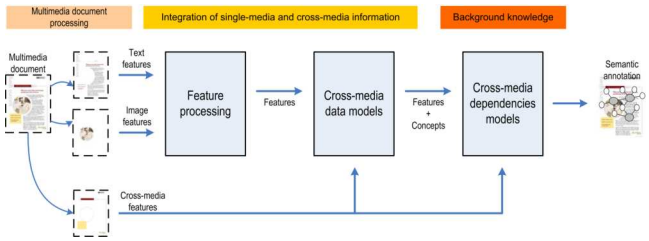

Figure 1: Cross-media knowledge extraction framework in the X-Media project.

3. CROSS-MEDIA EXTRACTION FRAMEWORK

To perform extraction from across the media, we have designed the machine learning-based framework depicted in Figure 1. The framework receives as input a set of multimedia documents and produces semantic annotations over the documents.

The document layout and extracted cross-references (e.g. captions) can hint at how each text segment relates to each image/raw data [1, 2, 4]. Arasu and Garcia-Molina [1], Crescenzi et al. [2] and Rosenfeld et al. [4] approaches are based on templates that characterize each part of the document. These templates are either extracted manually or semi-automatically. Rosenfeld et al. implemented a learning algorithm to extract information (author, title, date, etc). They ignored text content and only use features such as fonts, physical positioning and other graphical characteristics to provide additional context to the information. X-Media follows an approach similar to the one proposed by Rosenfeld et al., we extract a set of cross-media features for the types of documents we need to process. These cross-media features include: layout structure, distance between segments, cross-references, same type of font, font colour, and background colour/pattern.

Sparse feature data such as text and dense feature data such as images have very different characteristics. Learning algorithms handle different types of data simultaneously more effectively when data is pre-processed to produce a single common representation for all the data. We follow Magalhães and Rüger [3] and process text and image independently with probabilistic latent semantic indexing to produce an unified canonical representation of both text feature space and image/raw data feature space.

Once the feature data has been processed, some modelling algorithm can be used to create the knowledge models for 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. Several approaches have addressed similar problems, e.g. [3, 6]. The maximum entropy framework adopted, described in [3], fully addresses these issues by using a Gaussian prior (or a Laplacian prior) and a well known quasi-Newton optimization procedure.

4. 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.

5. 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).

6. REFERENCES

[1] A. Arasu and A. H. Garcia-Molina. Extracting structured data from web pages. In *ACM SIGMOD International Conference on Management of Data*, San Diego, California, USA, 2003.

[2] V. Crescenzi, G. Mecca, and P. Merialdo. Roadrunner: Towards automatic data extraction from large web sites. In *27th International Conference on Very Large Databases (VLDB)*, 2001.

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

[4] B. Rosenfeld, R. Feldman, and J. Aumann. Structural extraction from visual layout of documents. In *ACM Conference on Information and Knowledge Management (CIKM)*, 2002.

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

[6] Y. Wu, E. Y. Chang, K. C.-C. Chang, and J. R. Smith. Optimal multimodal fusion for multimedia data analysis. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 572–579, New York, NY, USA, 2004. ACM Press.