A Visual Tagging Technique for Annotating Large-Volume Multimedia Databases

A tool for adding semantic value to improve information filtering

(Post Workshop revised version, November 1997)

Konstantinos Chandrinos, John Immerkær, Martin Dörr, Panos Trahanias
Institute of Computer Science
FORTH
71 305 Heraklion, Crete
Greece
{kostel | jimm | martin | trahania}@ics.forth.gr

Abstract

The Computer Vision & Robotics Lab along with the Information Systems & Software Technology Lab at ICS-FORTH have recently developed a technique for marking-up different media to provide rapid entering of semantic information on large-volume multimedia databases. We expect this per medium semantic information to improve filtering of query results and facilitate retrieval of compact, context-sensitive information from large-volume multimedia databases. To that end, we present open design issues. Our technique offers the ability to interactively mark-up parts of the representation on the medium of choice (e.g. areas on a document image) and associate local information with text, images, sound-clips, video or subparts of them without altering the original. The output is XML files that can be parsed by viewers to navigate through the semantic information. The XML files can also be used by intelligent search agents for different media in order to refine the accuracy of retrieval in a query. The basic principles of implementation allow for database-flexibility through the use of XML. Developing all the software in Java allows for net-centric administration and use of a number of heterogeneous databases through standard Internet browsers. The Java implementation also guarantees platform-independence. This work has been developed under the ARHON project and is part of an ongoing effort to answer the need for novel ways of data entering and registering in large multimedia databases. The methods presented are intended to carry over the analogy of text mark-up by a database administrator or approved user, to other media such as still images, video and sound.

Introduction

In recent years ICS-FORTH has designed, implemented and installed a number of large-volume multimedia databases. On launching the ARHON project (A Multimedia System for Archival, Annotation and Retrieval of Historical Documents, Jan. 1997 - June 1998) we were faced with a good 1,500,000 archive manuscripts from the late 1600s to early 1900s in varying degrees of deterioration. The ARHON project [1] is to provide the framework under which these documents can be turned into a digital library accessed through an intranet at first, and then over Internet. As a test bed for a model we work on a sub-archive of about 100,000 documents. These documents have to be scanned to an electronic form, processed for image correction and then transcribed or even translated as some of them are in a foreign language compared to that of potential users.

The problem analysis

A first approach would be to combine OCR technology with text-based mark-up (e.g. SGML). However, the inaccuracy of state-of-the-art OCR software on all but the most modern and uniformly printed documents (let alone manuscripts) still demands thorough proof-reading and error correction. Also, the intensive manual labor and financial cost required for SGML text mark-up were deemed as barriers to our purpose. Even if one could get a perfect OCR result on 17th century manuscripts so as to apply Full Text Indexing techniques, the most important view of the resulting digital library, the semantic information would be missing. As a result additional mark-up would be necessary and in our case would have to be carried out by specialists who first would have to become familiar with SGML specific tools. As a last argument against textual mark-up we considered that it is a very invasive procedure with respect to the raw data, since textual mark-up changes the original. Even worse, in our case we have to cater for multiple annotations for the same document coming from several researchers. The above situation gives rise to a number of questions:

- *Accommodating different levels of authority*. Scholar researchers come at different levels of authority as well as different fields. Consulting the login data, assigning different weights to scholars as a default, according to the standing estimations of the archivists, seems an acceptable choice. However, authorities seem to "decay" in time as new scholars with new interpretations show up. This indicates that a proper weighting system had better be a function of the utilisation of annotations by other users. The users of the digital archive should be allowed to change such system rating of scholars either per session or through personalised profiles.

- *Ranking the query results according to relevance*. This could be done using the number of annotations for each result. Priority should be given to images or other more complex media (e.g.

video, sound) since these present higher fidelity to the originals as opposed to transcribed ASCII text.

Our approach

To overcome the shortcomings of OCR and SGML manual markup we have designed, developed and implemented a novel approach for visual tagging. In our approach we encourage the user to mark-up parts of the representation on his/her medium of choice (e.g. areas on a document image) and associate local information (e.g. filename, spatial or temporal co-ordinates etc.) with text, images, sound-clips, video or subparts of them without altering the original. To illustrate our approach we present here ***ImageTagger***, an implementation of our ideas that focuses on image documents (cf. Fig 1) and associates selectable faceted keywords to them. Nevertheless, the philosophy of our design and implementation is valid for any other media.

This mark-up technique produces a file in XML (eXtensible Markup Language, [2]) format that can be parsed to update a varying range of databases through the appropriate Java Database Connectivity driver (JDBC, [3]). Currently, we are implementing a JDBC for the Semantic Index System (SIS, [4]), a system developed at FORTH for describing and documenting large evolving varieties of highly interrelated data, concepts and complex relations. Since the format of the produced files is open, implementations for any other database system, whether RDBMS or object-oriented is considered trivial. In Figure 1 we also demonstrate a labelling technique implemented in our viewer. The viewer parses the output file to create individual labels for marked regions.

Using JavaServer 1.0 as an http server and servlet technology [5], we are now testing simultaneous media based queries. With the innate multithreading capability of servlets and Java this can be done without significant programming overhead.

System evaluation

What we consider of significant importance is that this design allows databases of different technology, such as a thesaurus and an RDBMS, to be updated through parallel parsing of the XML file. Additionally, implementation in Java allows the same code to be executed on different platforms. Currently Java ver. 1.1 is supported by Windows 95/NT and Solaris. To ease installation and maintenance and cater for repositories that are not implemented in these two platforms we have built into our program the ability to execute as a Java applet inside a standard Internet browser (cf. Fig. 1 presenting a snapshot of the ***ImageTagger*** running inside Netscape Communicator 4.03). With this last kind of use there is no client software installation and the user always gets the latest version.

Future directions

Next in our plans comes the implementation of taggers for other media, namely sound and video and the integration of all such media taggers using existing Web technology (HTML and Java) combined with the latest evolution in mark-up languages (Synchronised Multimedia Integration Language, SMIL [6]). In order to test query refinement with these tagging techniques we intend to apply search agents for each medium that will execute the query on their part and filter the results. Points we shall investigate in this direction include:

- *Automation of the search and filtering process through the use of intelligent agents*. Autonomous media-specific agents can be used to execute the actual search and collaboratively filter the results. Hence the agents will be required to reason about the use of the resources and negotiate among them in an agreed protocol. Provision should be taken for the machinery of agent-creation to allow further production of agents according to the expansion of the digital library. There is definitely a cost in the negotiation part of the agents function which even in the best design architecture is inversely proportional to the available bandwidth.

- *Creation of a user profile*. This profile will be published along with the query to the appropriate agents so as to guide them through the search process. We share the view that casual users should be given less retrieved information and a "More like this" option, compared to expert users that can pinpoint what they are looking for and can identify it easier in fine detail. Additionally, similarities in past queries from the same or a different user may provide stimulation for query refinement. Searching through the profile base poses an extra overhead in implementation and real time efficiency. In this direction we intend to implement some directory service solution, based perhaps on LDAP [7].

References

[1] K.V. Chandrinos, J. Immerkær and P.E. Trahanias, ARHON: A Multimedia Database Design for Image Documents, submitted to EUSIPCO 98, Special Session on Multimedia Signal Processing

[2] http://www.w3.org/xml/

[3] http://www.javasoft.com/products/jdbc/

[4] http://www.ics.forth.gr/proj/isst/Systems/SIS

[5] http://www.javasoft.com/products/java-server/

[6] http://www.w3.org/smil/

[7] Joao Ferreira, Jose Luis Borbinha, Jose Delgado: Using LDAP in a Filtering Service for a Digital Library, 5th DELOS Workshop, Budapest, Nov. 1997

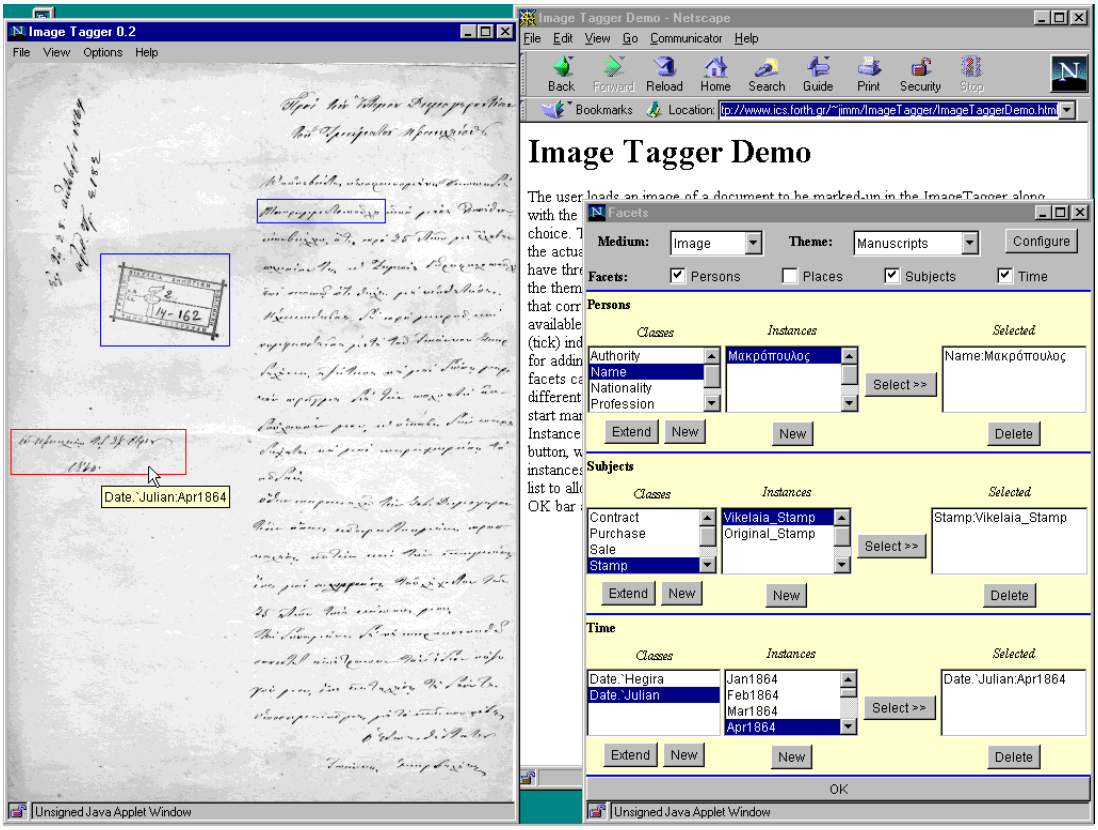

Figure 1. ImageTagger demo running as an applet in Netscape Communicator