Education and Information Technologies 9:3, 271–289, 2004.

Electronic Roads in Historical Documents: A Student Oriented Approach

MANOLIS WALLACE * and KOSTAS KARPOUZIS
Image, Video and Multimedia Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Greece
E-mail: {wallace,kkarpou}@image.ntua.gr

MARY STEFANOU
Faculty of History and Archaeology, School of Philosophy, National Kapodistrian University of Athens, Greece
E-mail: marystef@arch.uoa.gr

ILIAS MAGLOGIANNIS
Department of Information and Communication Systems Engineering, University of the Aegean, Greece
E-mail: imaglo@aegean.gr

STEFANOS KOLLIAS
Image, Video and Multimedia Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Greece
E-mail: stefanos@cs.ntua.gr

Abstract

The new educational approaches, as far as teaching of history in secondary education is concerned, are characterized by a shift away from sterile memorization and towards a critical approach of historical facts and phenomena; the aim is to contribute to both the development of students' historical concept and conscience and the promotion of critical thought. These teaching goals are pursued by promoting initiative of the students through self study assignments in the form of projects; students can be greatly supported in such tasks by computer-based applications, which can offer access to vast amounts of historical texts and data to be used next to the main scholar textbook and be analyzed by students. Still, existing applications seem to be quite inadequate for this purpose, as they require that the student be already informed on a matter, before the initiation of a quest for data. In this paper, we describe an intelligent information system that is designed to facilitate browsing of educational material and historical sources, thus allowing students to efficiently retrieve information on topics that are not yet known to them and expand in this way their historical knowledge. This can help in fulfilling the aforementioned teaching goals. Our system relies on the notion of the Electronic Road.

1. Introduction

A few years ago, history was just another theoretical course in which students in secondary education, namely children between the age of 11 and 17, were expected mainly to memorize facts and dates. With the latest changes in educational systems, in Greece and elsewhere, the objectives and methods of teaching history have been substantially altered.

* Corresponding author: 45, Achilleos Str. 176 74 Kallithea, Athens, Greece.

It is now aimed that through the study of history students can acquire more than historical knowledge; they can also develop their historical concept and conscience and promote, especially through intercourse with primary historical sources, their critical thought, compare various historical facts, quite often from different time periods and about different people, and find similarities and differences among them; this way they may start understanding the causalities that lead to the repetition of specific events over the years.

In order for this to be feasible, students are required to develop a series of new skills, that may be summarized in the following: students need to be able to (Ministry of Education and Religious Affairs, 2002a, 2002b):

1. Compare historical facts or phenomena and find similarities and differences among them.
2. Approach, comprehend and utilize different kinds of historical sources.
3. Associate and correlate facts, periods, ideals and civilizations of different periods and areas, and compare important historical figures based on their actions and total offer.
4. Comprehend the correlation of historical elements synchronically and diachronically and the correlation of the historical facts with contemporary facts.

Together with these, it is required that students realize that the credibility of the historical narration relies on the proper use of the sources; if the original historical sources are not studied and analyzed correctly, then the historical information that is acquired may be incorrect. For all of these reasons, it is required that students become familiar with original historical sources and are accustomed to utilizing them directly in order to reach a conclusion on a historical subject.

In conducting the course, the teacher needs to exploit all available means, in order to stimulate the students to achieve all the above. Next to traditional methods for teaching, much value is given now to the preparation of short projects, either personal or prepared by groups of students, via the research of various historical sources. In this process, sterile memorization of historical facts and dates is being set aside and a critical approach is being required, by asking students to commend on primary sources or visual material and to relate them to the historical facts. Of course, this implies that students need to have continuous access to such material, which can be facilitated to a great extent if historical sources are available in electronic form.

It is already established that electronic – based, independent learning environments can take their place in a traditional educational environment (Masterman and Sharpies, 2002; Vertsetis, 1998). Moreover, as the new generation becomes more affiliated with computers, their perception of the structure and content of electronic material is enhanced (Schuyten *et al.*, 1999). Information technology can help teachers locate the appropriate material to use, next to the main instructive textbook; this refers to primary sources, photographs and other visual material. Furthermore, it can help students prepare their assignments and projects, having a closer intercourse with the primary sources and having access to a broader range of learning material (Blow, 1991; Yang, 2003). On the other hand, various obstacles have to be overcome before information systems can be utilized directly and efficiently by students, with respect to the teaching and learning of history (Perkins *et al.*, 1992).

In this paper we present an information system that is designed to address such issues. We start by presenting existing systems that may be used by students that are trying to locate historical sources, as well as other computer systems that are related to the instruction of history, and conclude that their main weakness is that they are usually not appropriate for use by unsupervised students. The main reasons are that

1. They require that students are already knowledgeable on a topic before they may utilize the system efficiently.
2. They do not facilitate the study and comparison of facts that, although similar in nature, may differ in time, location and participating historical figures.
3. They require that the teacher is guiding pupils all through their research and study.

We continue by explaining how the electronic road metaphor, a notion borrowed from modern document browsing systems (Wallace *et al.*, 2003b; Constantinou *et al.*, 1999), can be integrated into a history related information system and used to alleviate some of the aforementioned problems. The resulting system may be used by students in a self – study manner, in the process of preparing a project/presentation on a history – related topic, as well as by teachers in the process of gathering supplementary course material.

The structure of the paper is as follows: in Section 2 we review related applications and technologies that can be used by students in the framework of history courses. In Section 3, we present the electronic road metaphor and explain how it can facilitate the aforementioned teaching goals, when integrated in an information system. Finally, in Section 4 we present the architecture of the proposed system, and in Section 5 we provide comparative experimental results for the proposed system. In Section 6, we list our concluding remarks.

2. Existing Approaches and Applications

When preparing a project, apart from books and libraries, students typically use general – topic search engines, such as Google, in their quest for information. Such engines most often return numerous documents, most of which are unrelated to the user's quest. More importantly, returned documents are unreliable, as far as their content is concerned, as they have not been checked and approved by teachers and historical experts. Further more returned documents may be a combination of different kinds of historical documents, of various and often conspicuous origins, and of primary historical sources. A pupil does not possess the ability to distinguish among them the correct and relevant ones, and most importantly to evaluate their importance. As a result, it is often observed that pupils include in their projects irrelevant data or even worse that they have been misled into drawing wrongful conclusions. As we have already mentioned, it is important for students to realize that the credibility of the historical narration relies on the historical sources; thus, they need to be discouraged from utilizing unreliable information systems to locate and retrieve historical sources.

Documents that are related to historical events can also be accessed via web search engines that are dedicated to this subject. An example of such a search engine is Argos, which is developed and maintained by the University of Evansville. Students can also

access resources for ancient Greek and Byzantine history online from the Perseus Project web page. Perseus is a digital library, which started out as a digital library of Classical Civilization and has been expanding its holdings to include Latin texts and lexicographical resources. It provides web access to a large number of images and translated Greek and Latin texts, which can be queried with the use of keywords. It also provides access to tools such as the electronic version of Liddell–Scott Dictionary, web based versions of Smyth's Greek Grammar and Allen and Greenough's New Latin Grammar, a web exhibit about Hercules and the Olympic Games, a tour of Olympia and access to Thomas Martin's overview of archaic and classical Greek history.

Furthermore, students and teachers can access primary sources for ancient and Byzantine history with the help of the TLG and PHI projects. TLG has digitized most Greek literary texts that have survived from Homer up to the fall of Byzantium. Its goal is to create a comprehensive digital library of Greek literature from antiquity to the present era (Pantelia, 1995). The PHI project is a similar digital library containing a selection of Greek and Latin inscriptions and papyri and nearly all works of Latin authors.

Both TLG and PHI data can be accessed through special engine tools such as Musaios, TLG Workplace, Antiquarium and others. Musaios offers full support for the TLG Word Index, including the ability to view citations for all indexed words. Antiquarium generates reports for search results that users may print out. It allows sorting of authors by date and makes use of the TLG word index and bibliographical information. In TLG Workplace, works can be automatically selected by date, classification, provenience and gender.

More information on all of the above, as well as a vast collection of web and other resources for history (Chiron, Lector, V&F, Diotima, The Duke Papyrus Archive, The Ecole Initiative, History/Social Studies Web Site for K-12 Teachers, The Oriental Institute Home Page, The Ovid Project, The Roman Forum, The Seven Wonders of the Ancient World) is available at *http://www.tlg.uci.edu/index/resources.html*.

3. The Electronic Road Metaphor

All the tools and applications presented in the previous section are, as far as the technological aspect is concerned, classic information retrieval systems. The conceptual document structure that characterizes them is the one presented in Figure 1. An index exists, listing all the keywords that one may use to perform a search; each of these keywords is linked to the set of corresponding documents. User queries are evaluated, with respect to the index, and the resulting document set is presented to the user.

When it comes to the case of students that wish to locate historical documents, systems of this type are proven to be inadequate for a series of reasons, the most important of which are summarized bellow:

Issue 1: In order to form a proper query, a student needs to already possess some knowledge. As in all other search – based information systems, if students are not aware of the existence of a specific topic, they will never form a corresponding query, and thus, they will never retrieve it. Pupils do not usually have the ability to raise proper questions be-

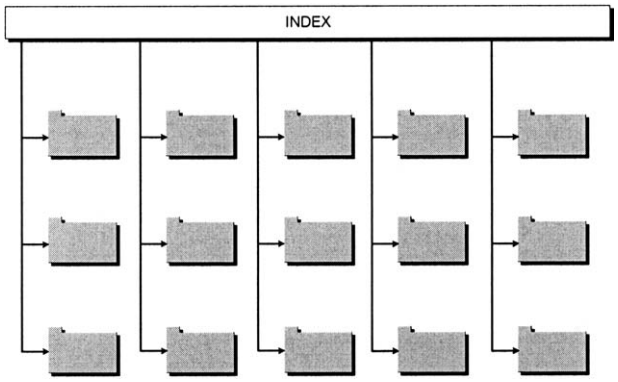

Figure 1. The indexed model.

cause of their ignorance of historical method and their inefficient historical knowledge, so they cannot get proper material that can lead them to meaningful conclusions.

Issue 2: Especially when the topic of the search is related to a chronologically distant period, as is for example ancient Greek history, even names of locations may differ from their modern counter parts. Thus, even if students are aware of the topic, they may also need to know all the diverse ways with which they could query for it.

Issue 3: Due to information overload (too many documents are selected for each query), it is quite difficult for the student to distinguish the ones that are truly related to the question at hand. "Historical events and situations must be reconstructed from a vast corpus of evidence, which may be incomplete, inconsistent and, for learners, difficult to interpret. An expert historian meets this challenge by bringing into play his or her "second record" (Hexter, 1971). That is, the expert historian utilizes the sum total of his or her existing knowledge, both historical and everyday, character and personal experience. In contrast, a secondary school pupil whose "second record" is still relatively impoverished will often resort to naive theories, or "alternative frameworks" (Hallden, 1986), according to which historical personages can be deemed intellectually inferior or be judged according to modern standards (Sansom, 1987). Finally, returned documents are a combination of primary historical sources and of different kinds of historical documents, of various and often conspicuous origins.

Issue 4: Pupils are not in the position of knowing the available primary historical sources or bibliography in every field and historical subject. It is possible that this ignorance may lead to the omission of important data, if there is nobody to guide them through their research and study. Existent systems not only do they do not offer such a guidance, but also often lead to the confusion of pupils due to the vast amount of documents that they are offering; among these documents the truly important ones may be missing. In this case pupils have no way of realizing this omission and they do not even have a way of searching for the right material. When using such systems it has repeatedly been observed that the guidance of the teacher is of primer importance in order not to have the above negative results. But in this way, the initiative of students is being held

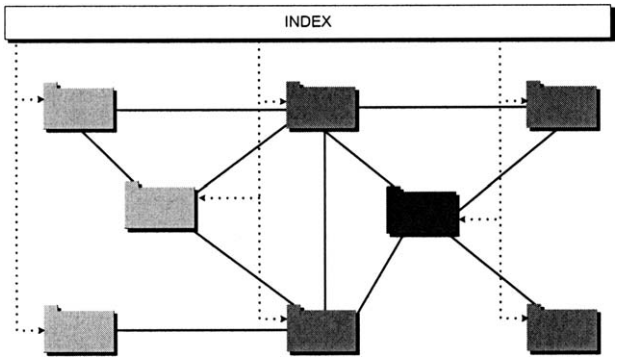

Figure 2. The relational model.

back, and students do not have the opportunity to develop their sense of responsibility, through personal work.

Issue 5: Similar documents cannot be located automatically, as relations among documents are not defined. Thus, students are not facilitated in locating and correlating similar events of different eras; as we have already mentioned, this is of major importance, given the new goals of teaching history.

Overall, the index-based perception of the document space degenerates project research to a simple search for documents, rather than the stimulating journey in historical information that it is initially meant to be.

In our approach, documents, together with their relations, form a conceptual graph, as seen in Figure 2; a user may navigate in the graph forming a personal path in the historical document space. Lately, the utilization of graphs of concepts to be taught together with the main relationships between them is becoming a common approach for intelligent tutoring systems; herein we extend on this concept and port it to the case of educational content management systems, so that it may contribute to the unsupervised yet efficient browsing of the content by pupils.

An index still exists; using it one may initiate a search. Still, the main service offered is that of adaptive navigation through documents, as the system can automatically determine and propose the documents that are related to the one the user is currently viewing.

In this approach, documents that are encountered during a student's interaction with the system are not considered to be unrelated from each other. On the contrary, they are considered to be the distinct steps of the student's path among available documents. This path is often referred to as an *electronic road* (Wallace *et al.*, 2003b; Constantinou *et al.*, 1999). This notion is used by the system in order to best determine the set of relevant documents to propose to students; as the path grows, the documents it comprises are compared in order to detect the specific topic of the browsing session.

Especially for the case of history, the notion of electronic roads can be extremely helpful. Historical documents may be correlated to other ones for numerous reasons, such as same authors, same location, same time period, similar events, same historical figures and people and so on. Consequently, for each document in an information system, numerous other

related documents can be found. This implies that considering only the document a user is currently viewing in order to determine the set of documents to propose fails to tackle the issue of information overload. Applying the notion of electronic roads, all recently viewed documents may be considered in order to determine a more specific *context* for the session, using which to filter the set of candidate documents. In this process, documents that appear last in an electronic road have a greater probability of being closely related to the user's wish and, as the principles of relevance feedback dictate, are taken into account to a greater extent when extracting the electronic road context (Yong *et al.*, 1998).

The main difference between our approach and others in the literature is that, together with indexing information, relations between documents are also stored/handled by the system. This, together with other implementation and technological choices we have made and which further explain in the following section, address the aforementioned disadvantages of the indexed model as follows:

Issue 1: Students do not necessarily need to have before hand plenty of information on the topic that interests them, as a proper query is not an absolute requisite. One may start from a query that is merely close to the topic, and then locate other, more "on the topic" documents via document relations. Thus, even if a student cannot raise the proper question in order to have best possible system response with regards to the project at hand and uses a indefinite or more general query, it is almost certain that the system will provide the documents relevant to the subject of the research through in interactive process that relies on the utilization of document relations.

Issue 2: The indexing of the historical sources and documents is not performed in a fully automated manner via word matching and stemming, but mainly by historical experts. Thus, modern counterparts of locations, names and terms are also included in the index, to facilitate younger and less knowledgeable users.

Issue 3: The proposed system assures that the pupils are not being faced with documents of conspicuous origin, documents that are irrelevant to the subject or documents of different kinds, i.e. primary sources and bibliography. For example, when a pupil has already selected to study a number of historical sources, the system will present similar documents that are primary sources as well. Overall, the system minimizes the probability that pupils will include irrelevant data in their projects.

Issue 4: The system prevents the omission of important documents, either bibliography or historical sources, as the historical experts have defined the set of documents that need to be presented for each topic. This way, it is not necessary for the teacher to supervise pupils throughout their whole research; pupils can be safely offered the opportunity to develop initiative and work without any supervision.

Issue 5: Similar documents are automatically presented to the pupil by the system. Documents that differ in most aspects – namely era and leading personages of stated event, author and era of document's composition, etc. – but have at least one common element, will be presented to the student, in the proper context. In this way, students are able to conduct comparative essays about historical facts, phenomena and historical personages, even though they may not initially know which facts, phenomena or personages to

use for this comparison. What remains for the student to do is comprehend and evaluate what it is that relates the documents.

4. System Architecture

The architecture of the system is presented in Figure 3. The digital library is the document store. Together with the actual documents in electronic form, it also contains the index; the index is an extension of classic text retrieval indices, as keywords in it are organized in categories. The role of the content expert is to provide the indexing of the documents in a manual or semi-automatic manner; automatic indexing has been ruled out because, due to the instructive nature of the system, the validity of the data it contains is of crucial importance. The digital library also contains the edges of the graph that forms the document space, i.e. the relations between the documents. These are used to reduce system response times.

The two intelligent modules are in charge of analyzing data and producing actionable information. The document similarity estimation module analyzes the document space in order to detect similarities among documents; these similarities are logged in the digital library. The electronic road analysis module analyzes the session information, i.e. the feedback the user has provided by forming an electronic road, in order to determine the context of the current browsing session. In order for these tasks to be performed, some knowledge is needed; this is available in the form of a suitable ontology.

Finally, the document recommendation module selects from the digital library the documents to propose to the user. As the documents that are related to the one the user is currently viewing are typically numerous, the recommendation module utilizes the context that is provided by the electronic roads analysis module in order to rank/filter candidate documents.

This type of filtering typically requires that one is already quite knowledgeable on a topic. In the case of students looking for information, it is usually the teacher that has the load of supporting them by helping in the filtering of irrelevant documents. The presented application of the electronic road metaphor substitutes to a great extent this assistance, thus allowing students to act more on a self – study manner, i.e. without the need for direct supervision by a teacher.

In the following subsections we elaborate on each one of the system components.

4.1. Digital Library

The Digital Library comprises the electronic documents and annotation data, together with document to document relations. All the documents that the system currently contains refer to ancient history. They are mainly original works (for example, Thucydides) together with a modern translation, or, in some cases, works of contemporary historians on the same subject.

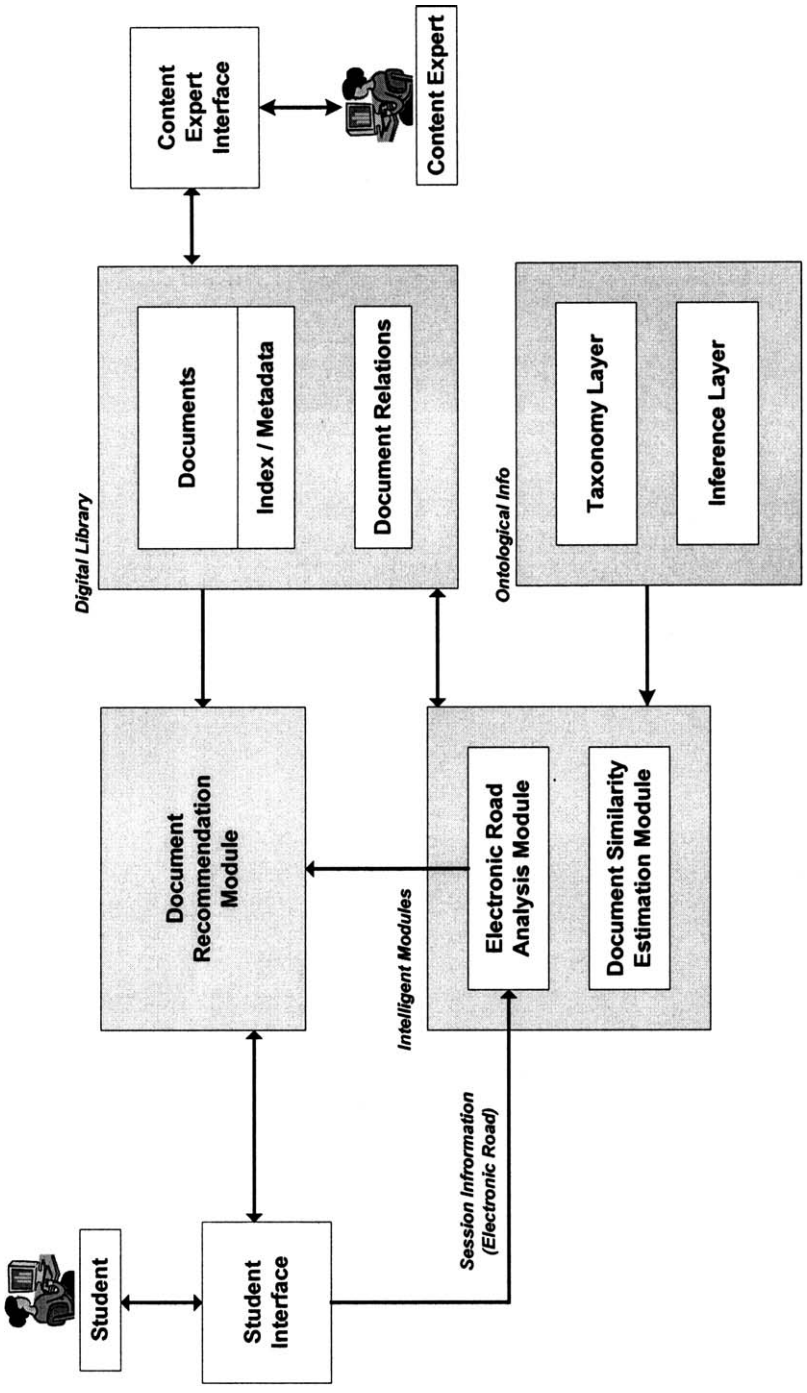

Figure 3. The system's architecture.

In the index, concepts extracted automatically through string matching, as well as manually provided creation and content information, are structured in groups related to the following categories:

- historical term (revolution, treaty, civil war, etc.),
- participating figures and people,
- era,
- location,
- conditions/factors (under which an event occurred or phenomenon appeared),
- author,
- time of creation of the document,
- location of creation of the document,
- causes and consequences.

The grouping of index entries is essential for the implementation of the electronic road metaphor, as the groups can be used to define a context for a browsing session. For example, a student might be browsing for historical sources concerning treaties in one case and for texts written by a specific author on another.

Automatic analysis of documents (the translated versions) is not impossible (Wallace DTC), but is not free of uncertainty and errors either; as natural language processing (NLP) remains an open research problem, document analysis is reduced to simple, context insensitive matching of terms in the text to those in the ontology. Therefore, indexing of documents is not performed in an automatic manner; initial system recommendations for the index entries of the document have to be verified/corrected/extended by a human expert.

Document relations are structured as multidimensional vectors, describing the degree to which two documents are related to each other, according to each one of the aforementioned categories. Thus, two documents may be related to each other as referring to the same location, while at the same time they may differ as referring to different time periods. Given that the validity of both the index and the ontological information have been assured by human experts, the estimation of the document relation using them as inputs is a much more efficient process and is performed in an automated manner by the document similarity estimation module.

4.2. Ontological information

An ontology O is a formal, shared conceptualization of a particular domain of interest (Gruber, 1993). It can be though of as a set of concepts together with the combination of two semantic layers; one provides the conceptual hierarchy and the other one provides for inference:

$$O = \{S, O_1, O_2\},$$

where $S = \{c_i\}$ is the set of concepts c_i described by the ontology, with $i = 1..|S|$. The conceptual hierarchy layer, describes the relations among the semantic entities. It can be described as:

$$O_1 = \{R_j\},$$
$$R_j : S \times S \to \{0, 1\}, \quad j = 1..|O_1|,$$

where O_1 is the count of distinct relations in the ontology. Although any type of relation may be contained in an ontology, R_j are typically taxonomic (i.e. partial ordering) or compatibility (i.e. symmetric) relations. O_2, the inference layer, describes inference rules for which taxonomic information is not sufficient. For example, consider the statements:

1. A is the father of B.
2. A is the father of C.

The fact that A is of gender male can be described in O_1 as a relation between the concepts of male and father. The fact that B and C are siblings, on the other hand, cannot be described in this way; the corresponding rule has to be inserted in O_2.

It is well understood that relations among real life entities are in most cases a matter of degree, and are, therefore, modeled best using fuzzy binary relations. Moreover, we have established in previous works that, if an ontology is to be used in an information retrieval system, it is best if the relations it contains are made fuzzy, as to reflect the degree of relevance among concepts (Wallace *et al.*, 2003a; Akrivas *et al.*, 2002).

Ontological taxonomies, on the other hand, are crisp in principle. Thus, they fail to fully describe real life concepts, and are limited to α-cuts of the desired fuzzy relations. Therefore, we extend the aforementioned definition of the first ontological layer, as to support for fuzzy binary relations as well:

$$O_1 = \{R_j\},$$
$$R_j : S \times S \to [0, 1], \quad j = 1..|O_1|.$$

Fuzziness of relations R_j is of major importance. Both for compatibility and ordering relations, $R_j(A, B) \to 1$ implies that A and B are almost identical in meaning. $R_j(A, B) \to 0$, on the other hand implies that A and B are barely related, if R_j is a compatibility relation, or that B is a much narrower concept that A, if R_j is an ordering relation.

The ontology that has been developed for this system aids, first of all, in handling Issue 2 of Section 3; concepts of the taxonomy are mentioned with all known textual forms. Furthermore, the ontology is used in the correlation of history related documents to each other. Of course, the main topic of the ontology is history; the concept hierarchy layer contains information such as the dates and places of the most important historical events, life span, origin and office of the most important historical figures, etc.

Still, other topics had to be included as well. The most important among them is the related geographical information, such as that Athens and Sparta are in Greece while Alexandria is in Egypt or that Istanbul can also be found as Constantinople or that the ancient name of France is Galatia. Some common knowledge information is also included, such as the fact that June and July are parts of the summer. Finally, the inference layer contains history related rules such as the fact that Olympic games imply truce or peace.

Of course, the generation of such complex ontologies is not a simple task. Quite the contrary, it is a task that requires that a person has some basic knowledge management skills, and also requires a trial and error approach. In order to generate the knowledge of the system we have developed a set of tools that allow historical experts to specify knowledge components via simplified user interfaces. Still, the need for the cooperation of knowledge engineers in this process has not been overcome.

Readers interested in more information on the properties of the fuzzy relations used to represent knowledge may refer to (Akrivas *et al.*, 2004). More information on the handling of these relations can be found in (Wallace and Kollias, 2004), while information on the techniques leading to the utilization of the relations can be found in (Wallace *et al.*, 2003a; Akrivas *et al.*, 2002).

4.3. Intelligent modules

The document similarity estimation module operates off – line and aims to automatically determine the relations among documents. It operates on the index, i.e. the characterization of the documents, with respect to the knowledge that exists in the ontology.

In order to facilitate the utilization of the ontological information in the estimation of inter – document relations, all available taxonomic information is combined into a fuzzy, quasi ordering binary relation T (Wallace *et al.*, 2003a):

$$T = \left(\bigcup R_j^{p_j}\right)^{|S|},$$

where $p_j \in \{-1, 0, 1\}$, $R_j^n = R_j \circ R_j^{n-1}$ and R_j^{-1} is the inverse relation of R_j. Using T we may find all concepts that are related to a document

$$d_k = \sum_{c_i \in S} c_i / d_k(c_i), \quad k = 1..|D|,$$

where $D = \{d_k\}$ is the set of documents in the system:

$$M(d_k) = d_k \circ T.$$

The degree of similarity between two documents is computed as the height of the intersection of the sets of concepts that are related to them:

$$\mathrm{Sim}(d_1, d_2) = h\big(M(d_1, d_2)\big)$$

and is stored in the digital library.

The electronic road analysis module accepts the last visited documents as input. These are considered as the last steps of a path and are compared to each other, in order to determine the topic of the current session; this is achieved through a soft intersection operation. The term soft is used due to the weighting of documents, based on their position in the electronic path. Specifically, were there no weighting of documents, the context $M(ER)$ of an electronic road ER could be calculated as the intersection of the topics that are related to each participating document:

$$M_{\text{crisp}}(ER) = \bigcap_{d_j \in ER} M(d_j).$$

Now that we want the last documents in the electronic road to be considered to a greater extent in the calculation of the context, the above formula is adjusted as follows.

When the electronic road is initiated with a document d_1, the context is computed easily as

$$M(ER) = M(d_1).$$

In a consequent step, when document d_2 is added to the electronic road, the context of the extended electronic road ER' is calculated as:

$$M(ER') = M(d_2) \cap \big(0.8M(ER) + 0.2S\big).$$

This way, the influence of a document on the context of the electronic road diminishes as the distance (in count of steps) between that document and the current one increases. On the other hand, all documents of the electronic road contribute to the overall context, each one to its own degree.

The function of both modules presented in this subsection is based on fuzzy relations algebra operations. More on the topic may be found in Klir and Yuan (1995).

4.4. Document recommendation module

Using the result of the electronic road analysis module as a guideline, the document recommendation module ranks documents that appear to be related to the one the student is currently viewing, in order to form a meaningful recommendation. The role of the module is to detect the documents that are most relevant to the session of the current user interaction, thus helping alleviate the problem of information overload.

The set of documents that are related to the one the student is currently viewing is readily available in the digital library as a weighted list. More formally, if d_0 is the current document, then the set of documents that are related to it is

$$N(d_0) = \sum_{d_j \in D} d_j / \operatorname{Sim}(d_j, d_0).$$

This set may be considered as the document's neighborhood in the document space. A document $d_j \in N(d_0)$ is related to the context of the electronic road to the degree it is related to concepts that belong to that context:

$$\alpha\big(d_j, M(ER)\big) = \frac{|M(ER) \cap d_j|}{|M(ER)|},$$

where $|f|$ is the scalar cardinality of fuzzy set

$$f = \sum_{c_i \in S} c_i / f(c_i);$$

it can be computed as

$$|f| = \sum_{c_i \in S} f(c_i).$$

When $\alpha(d_j, M(ER))$ is high, then the weight $\mathrm{Sim}(d_j, d_0)$ of d_j in the set should be promoted. When, on the other hand, $\alpha(d_j, M(ER))$ is low, then it should be diminished. Keeping in mind that $\mathrm{Sim}(d_j, d_0)$ is a fuzzy degree, and therefore lies in the [0, 1] range, it is easy to see that the above are satisfied by

$$N'(d_0) = \sum_{d_j \in N'(d_0) \cap \overline{ER}} d_j / \mathrm{Sim}'\big(d_j, d_0, M(ER)\big),$$

where $\overline{ER}$ is the set of documents not already included in the electronic road and

$$\mathrm{Sim}'\big(d_j, d_0, M(ER)\big) = \big[\mathrm{Sim}(d_j, d_0)\big]^{2-(3/2)\alpha(d_j, M(ER))}.$$

The document recommendation module adjusts the weights in this fuzzy set, according to the context of the already formed electronic road. This way, the documents that are most suitable for the continuation of a semantically meaningful navigation in the historical document set are promoted and presented first to the student. Given the typically large number of documents that may be related to the current one to a degree other than zero, this ranking may be considered as equivalent to filtering, as a user will rarely access a document that is positioned towards the end of a long list.

5. Experimental Results

In this section we provide two moments of the system; the selected examples demonstrate the function of the electronic road in the document recommendation process. In the first example, a student is asked to prepare an essay with the following subject: "Battles between Greeks and Persians in the historical work of Herodotus". The student forms an electronic road by selecting the documents shown in Table 1. One can easily see that the

Table 1. The first electronic road

1	*Herodotus 7, paragraph 223–225* Greeks against Persians in the battle at Thermopylae in 480 BC. Greeks are being defeated.
2	*Herodotus 6, paragraph 111–113* Athenians against Persians in the battle at Marathon in 490 BC. Persians are being defeated.
3	*Herodotus 5, paragraph 102* Battle in Ephesus between Ionians and Persians Ionians are being defeated.
4	*Herodotus 6, paragraph 18* Persians lay siege to the Ionian city Miletus.
5	*Herodotus 8, paragraph 84–96* Sea battle near Salamis between Greeks and Persians in 480 BC. Persians are being defeated.

Table 2. The most relevant documents for the first electronic road

100%	*Herodotus 9, paragraph 61–70* Greeks against Persians in the battle at Plataea in 479 BC. Persians are being defeated.
100%	*Herodotus 9, paragraph 98–106* The battle at Mycale between Greeks and Persians in 479 BC. Greeks win.
100%	*Herodotus 6, paragraph 14–17* Sea battle near the small island of Lade between Ionians and Persians. Ionians are being defeated.
100%	*Herodotus 9, paragraph 115–121* Athenians lay siege to Sestus where there are still Persian garnisons, in 479 BC.
65%	*Aeschylus, Persai 249–531* Description of the sea battle in Salamis between Greeks and Persians in 480 BC.
65%	*Herodotus 1, paragraph 29–33* The Athenian legislator Solon visits the court of the king of Lydia Kroisos.
55%	*Herodotus 3, paragraph 40* Polykratis, the tyrant of Samos, is being advised by Amasis, the Pharao of Egypt, about his extremely good fortune.
40%	*Thucydides 1, paragraph 100* Athenians and their allies win over Persians in a sea battle and a terrestrial battle near the river Eyrymedon in Pamfylia.
40%	*Arrianus Cynegeticus 24.5* Greeks win over Persians in sea battle near Salamis.

Table 3. The second electronic road

1	*Aeschylus, Persai 249–531* Description of the sea battle in Salamis between Greeks and Persians in 480 BC.
2	*Diodorus 11 17–19* The sea battle near Salamis between Greeks and Persians in 480 BC. Persians are being defeated.
3	*Arrianus Cynegeticus 24.5* Greeks win over Persians in sea battle near Salamis.
4	*Herodotus 8, paragraph 84–96* Sea battle near Salamis between Greeks and Persians in 480 B.C. Persians are being defeated.

electronic road comprises documents written by Herodotus in the same, more or less, time period. Moreover, their content is related to either battles or sea battles.

The system selects as most relevant the documents that appear in Table 2. The system's response contains documents that are related to the last document in the electronic road in various ways. Still, the ones that are selected first (with greater degrees) are the ones that are most related to the aforementioned context.

In the second example, the student is asked to prepare an essay with the subject: "The battle of Salamis as presented in primary historical sources". Similarly to before, Table 3 presents the electronic road of this example and Table 4 the corresponding system recom-

Table 4. The most relevant documents for the second electronic road

100%	*Ctesias fr.13.141–152* Sea battle near Salamis between Greeks and Persians in 480 BC. Persians are being defeated.
100%	*Plutarchus, Life of Themistocles 13–15* Sea battle near Salamis between Greeks and Persians in 480 BC. Persians are being defeated.
100%	*Plutarchus, Life of Aristides 9* Sea battle near Salamis and terrestrial battle in the small island Psytaleia between Greeks and Persians in 480 BC. Persians are being defeated.
65%	*Xenophon, Hellenica, 2.2.9* Spartan general Lysandre ravages Salamis in 405/4 BC.
50%	*Strabo Geographica 9.1.9* Description of island Salamis.
50%	*Thucydides 1, paragraph 100* Athenians and their allies win over Persians in a sea battle and a terrestrial battle near the river Eyrymedon in Pamfylia.

Figure 4. The response of Google.

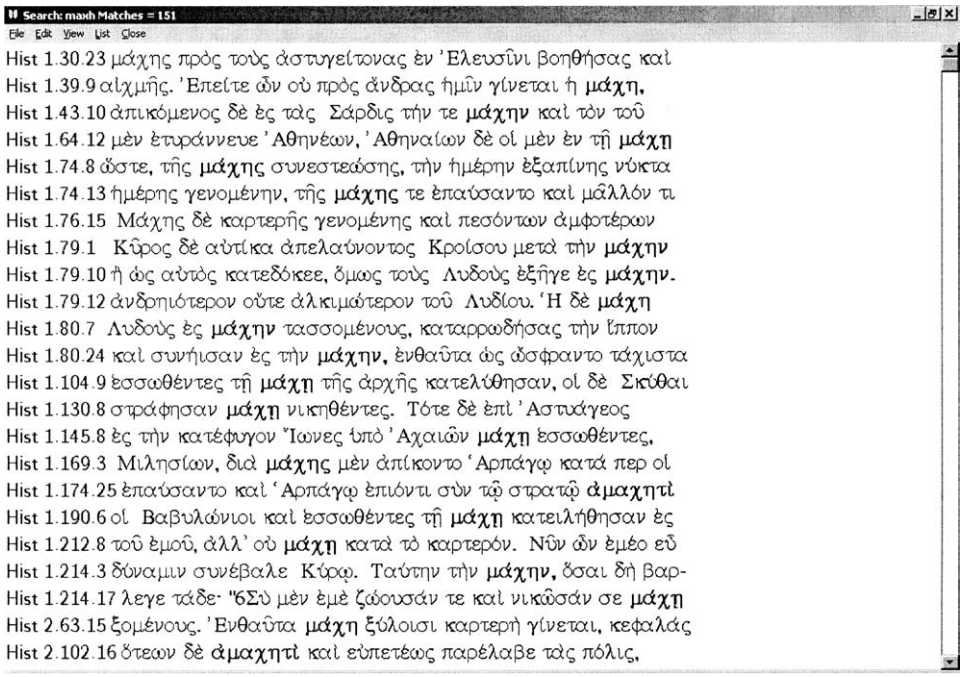

Figure 5. The response of TLG workplace.

mendation. In this case, all documents in the electronic road are related to the sea battle at Salamis; authors and creation time periods vary. The documents that are related to the sea battle at Salamis are correctly selected first, with others that are less relevant to the overall context following with smaller degrees.

It is worth noting that the current document is the same in both cases. Still, the overlap between the two result sets is minimal, due to the differences in what is common among all the documents of the session; the system automatically determined which of the documents that are related to the current one are also related to the context of the overall browsing session. This is the effect of the utilization of the electronic road metaphor.

Alternatively, the pupil is asked to use Google in the preparation of the first of the two essays. The query the pupil forms is "Battle Herodotus Greeks Persians". In Figure 4 we present the beginning of the first page of the results. We before anything else we observe that about 12600 documents are selected, which makes it impossible for the pupil, or anyone else in that matter, to process all of them. Moreover, most of the documents in the first pages of results are hardly related to the topic in question, and many are not related at all. Finally, the validity of the results, with the exception of *www.herodotuswebsite.co.uk*, is not verified by history experts. Similar results are obtained when trying with queries "Salamis" or "Salamis sea battle" which refer to the second assay requested from the student.

Finally, the pupil is provided access to the TLG Workplace, a content management system designed for history postgraduate students and researchers, and is asked to prepare the

same essay; TLG CD-ROM contains all the ancient Greek literary texts that have survived from Homer up to the fall of Byzantium. As the system does not allow users to use multiple words in their queries, the pupil is forced to use the very general word "μάχη", i.e. battle. The pupil also restricts the search to works by Herodotus. The result set, as seen in Figure 5, contains 151 documents, all of which come from the original work of Herodotus (no secondary bibliography is provided). Most of the selected documents are not related to the project of the pupil, as they refer to battles between opponents other than the Greeks and Persians. Important related texts have also been omitted because they do not contain the word "μάχη". None of the returned documents is available in a translated form, so the pupil needs to be fluent in ancient Greek in order to exploit the results. As a final remark, the TLG Workplace needed around twenty minutes to search the complete document base before providing the result. Thus, only users that are able to acquire a satisfactory result with the first query they form are able to use it in real life; it is not adequate for use by a pupil that will most probably tackle the search problem via a trial and error approach.

6. Conclusions

In this paper we explained that the information services that students of history need, and their teachers wish for them to have, are quite different from most existing general purpose, or history specific, information systems; students need to interact with a system that (i) allows them to locate information on topics they have not yet mastered, (ii) drives them to critically approach new information, (iii) aids them in specifying the topic of their research in an interactive manner.

Continuing, we have presented a system that attempts to offer such services. This system relies on a relational perception of the document space, as well as on an application of the electronic road metaphor, towards the determination of the context of the student's research. Of course, although the system has been designed in order to facilitate pupils in navigating through historical documents, teachers may find it useful as well in the process of locating and selecting supplementary material for a class.

The efficiency of the application of the electronic road metaphor, in order to aid students that are not yet knowledgeable in the topic in browsing for information, has been indicated via two representative examples, as well as via comparisons to a general purpose search engine and a history specific content management system.

References

Akrivas, G., Wallace, M., Andreou, G., Stamou, G., and Kollias, S. (2002) Context – sensitive semantic query expansion. In *Proceedings of the IEEE International Conference on Artificial Intelligence Systems (ICAIS)*, Divnomorskoe, Russia, September.

Akrivas, G., Stamou, G., and Kollias, S. (2004) Semantic association of multimedia document descriptions through fuzzy relational algebra and fuzzy reasoning. *IEEE Transactions on Systems, Man, and Cybernetics*, A, **34**(2).

Blow, F. (1991) Information technology: Catalyst for change in history teaching in schools. *Computers & Education*, **16**(1), 57–64.

Constantinou, C., Kakas, A., Katsikides, S., Papadopoulos, G., Pattichis, C., Pitsillides, A., and Schizas, C. (1999) Constructing electronic roads in the information society: A case study in Cyprus. In *Proceedings of Neties '99*, Krems, Austria, March.

Gruber, T. R. (1993) A translation approach to portable ontology specifications, *Knowledge Acquisition*, **6**(2), 317–335.

Hallden, O. (1986) Learning history. *Oxford Review of Education*, **12**, 53–66.

Hexter, J. H. (1971) *The History Primer*. Basic Books, New York, NY.

Klir, G. J. and Yuan, B. (1995) *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Masterman, L. and Sharples, M. (2002) A theory-informed framework for designing software to support reasoning about causation in history. *Computers & Education*, **38**(1–3), 165–185.

Ministry of Education and Religious Affairs (2002a) Directions for Teaching of Philological Courses in Lyceum. Athens.

Ministry of Education and Religious Affairs (2002b) Directions for Teaching of Philological Courses in Gymnasium. Athens.

Pantelia, M. C. (1995) Electronic resources for classicists: The second generation. *New England Classical Journal*, **22**, 117–127.

Perkins, P., Spaeth, D. A., and Trainor, R. H. (1992) Computers and the teaching of history and archeology in higher education. *Computers & Education*, **19**(1–2), 153–162.

Sansom, C. (1987) Concepts, skills and content: A developmental approach to the history syllabus. In *The History Curriculum for Teachers*, C. Portal (ed.). The Falmer Press, London, pp. 116–141.

Schuyten, G., Dekeyser, H., and Goeminne, K. (1999) Towards an electronic independent learning environment for statistics in higher education. *Education and Information Technologies*, **4**, 409–424.

Vertsetis, A. B. (1998) *Teaching of History*. Athens (in Greek).

Yang, S. C. (2003) Computer-mediated history learning: Spanning three centuries project. *Computers in Human Behavior*, **19**(3), 299–318.

Yong, R., Huang, T. S., Ortega, M., and Mehrotra, S. (1998) Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, **8**, 644–655.

Wallace, M. and Kollias, S. (2004) Computationally efficient incremental transitive closure of sparse fuzzy binary relations. Submitted to *The IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.

Wallace, M., Akrivas, G., and Stamou, G. (2003a) Automatic thematic categorization of documents using a fuzzy taxonomy and fuzzy hierarchical clustering. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, St. Louis, MO, USA, May.

Wallace, M., Karpouzis, K., Stamou, G., Moschovitis, G., Kollias, S., and Schizas, C. (2003b) An efficient personalised content browsing system based on the electronic road metaphor. *IEEE Multimedia*.

Wallace, M., Stefanou, M., Karpouzis, K., and Kollias, S. (2003c) Towards supporting the teaching of history using an intelligent information system that relies on the electronic road metaphor. In *IEEE International Conference on Advanced Learning Technologies (ICALT)*, Athens, Greece.