

# CAPTURING THE HISTORICAL RESEARCH METHODOLOGY: AN EXPERIMENTAL APPROACH

Torou Elena<sup>1</sup>, Katifori Akrivi<sup>1</sup>, Vassilakis Costas<sup>2</sup>, Lepouras Georgios<sup>2</sup>, Halatsis Constantin<sup>1</sup>

<sup>1</sup> Department of Informatics and Telecommunications, University of Athens, Athens, Greece

<sup>2</sup> Department of Computer Science and Technology, University of Peloponnese, Tripoli, Greece  
[etorou@di.uoa.gr](mailto:etorou@di.uoa.gr), [vivi@di.uoa.gr](mailto:vivi@di.uoa.gr), [costas@uop.gr](mailto:costas@uop.gr), [gl@uop.gr](mailto:gl@uop.gr), [halatsis@di.uoa.gr](mailto:halatsis@di.uoa.gr)

## Abstract

This paper presents the results of a study on how historians conduct research in a historical archive, and the methodologies they use while searching. Historic research involves finding, using, interpreting and correlating information within primary and secondary sources, in order to understand past events. The collection of historical data is accomplished through methodical and comprehensive research in primary and secondary sources. An important factor in our study was to understand what kind of data and/or information historians are looking for in a library/historical archive, either printed or digitized, and which research methodologies or research models they use while they investigate a historical archive. Since this issue has not been addressed insofar, and therefore there are no methods for elucidating research methodologies or research models that historians employ / use, we formulated a questionnaire comprising of seven information retrieval tasks commonly addressed in the context of historic research. History researchers were asked to describe in detail how they would proceed in searching for the information they need for completing these tasks. Through this procedure we aimed to investigate the different ways a historian can use to tackle a specific question, examine whether there exists a common research methodology, and the historic researchers' expectations and preferences. The insight gained from this investigation can be used for educational purposes, since it could be useful in the creation / development of a methodology for conducting research on historical information. Furthermore, the findings can be exploited in the context of organizing documents within historical source repositories, so as to facilitate the retrieval of documents by historians; finally the presented findings can serve as a preliminary requirement analysis phase for building tools that will enable historians to access more rapidly and fully the information they need.

**Keywords** - Research methodology, historical research, education

## 1 INTRODUCTION

Libraries and historical archives (HAs) are regarded as the main repositories for preserving and maintaining historical documents. Their documents may constitute either primary or secondary sources, and be maintained in the form of books (pages bound together), manuscripts, single pages, photos, paintings, video etc. A source is characterized as *primary* if it has been created during the period of interest, whereas *secondary* sources are those created later on and are based on the analysis of primary sources [2].

Historians conducting research systematically examine past events to give an account; historic research may involve interpretation to recapture the nuances, personalities, and ideas that influenced these events, and the expected research outcome is to communicate an understanding of past events [1]. Their main objective is to recreate the past, through existing records and their interconnections. In this process, historians employ their scientific knowledge, experience and intuition to decide which information they will need to find and study during each next step, and subsequently attempt to locate sources that contain this information.

In this work, we attempt to investigate the historians' search methods in the context of printed and digitized libraries and historical archives. An important factor in our study was to understand what kind of data or information historians are looking for in an historical archive (or *library* in general), either printed or digitized, and which research methodologies or research models they use, while they investigate an historical archive. Since this issue has not been addressed insofar [3], [4], no methods for elucidating research methodologies or research models that historians employ have been reported in the literature. Therefore, we formulated a questionnaire comprising of seven information retrieval tasks commonly addressed in the context of historic research, and historical researchers were asked

to describe in detail how they would proceed in retrieving information for completing these tasks. The approach of a semi-structured interview was adopted for this stage, to maximize the completeness of responses, while allowing interviewees to freely communicate their replies. The experiment took place in the Historical Archive of the University of Athens (HAUA). Through this procedure we aimed to investigate the different ways a historian may use to tackle a specific question, examine whether there is a common research methodology, and the historic researchers' expectations and preferences. Printed and electronic corpuses were examined separately, so as to determine whether different methods and habits are employed in each case; this has found to be important, since historical researchers proved to proceed differently in printed and electronic document collections, and specific aids should be provided for electronic document collections.

## 2 A STUDY OF HISTORIAN RESEARCH METHODS

As described in [7], in order to identify the historical researchers' needs and requirements, we combined two different approaches: (a) the study of queries made by historians to the Historical Archive of the University of Athens and (b) the use of semi-structured interviews with historians. The results of the first method were very interesting; the analysis revealed that historical researchers very frequently surveyed the issues of «Person biographies» and «historical evolution of institution / organisation». The study of these queries was the first step for understanding what kind of data or information historians are looking for in a library/historical archive, either printed or digitized material. After having analyzed *what* information historical researchers typically searched for, we proceeded in investigating *which methodology* historical researchers follow to complete each task and *how* they proceed in retrieving documents from the historical archive. To this end, we compiled a set of information retrieval tasks that corresponded to the most frequently occurring queries; these tasks were generalized enough to facilitate our task of recording information on how different search types were conducted. Interviewees were asked to describe in detail how they would work for each of the tasks; interviewees were asked to describe separately how they would work (a) in printed corpuses and (b) electronic corpuses, so as to document any differences in their methodology and habits they follow and determine which tools should be made available in the context of electronic historical archives. The seven information retrieval tasks are as follows:

- Q1. Describe how would you search for information regarding «Kostis Palamas as University of Athens, General Secretary»
- Q2. Describe how would search for information on the historical evolution of the Chemistry Department of University of Athens
- Q3. Describe how would search for information regarding the historical evolution of an organization or city.
- Q4. Describe the way that you would seek for the PhD thesis of professor X who lived between 1850 and 1920
- Q5. Describe how would you search for the Curriculum Vitae of a teacher (e.g. P. Papageorgiou), who taught Greek Studies in the Department of Philosophy during the academic year 1909-1910.
- Q6. Describe how you would look for information in case of synonymy. How would you verify that two pieces of information actually refer to the same entity or synonymous but discrete entities?
- Q7. Describe how you would search for information regarding «female graduate students of University of Athens coming from Smyrna».

### 2.1 User group

A user group of 20 historical researchers was interviewed for gaining insight on how they search for historical information. The user group consisted of 10 men and 10 women. 4 of them were employees of the HA, and 16 were historical researchers who had visited the HA of the Athens University more than 3 times. The participants were chosen to be familiar with information retrieval from digital sources (including searching the internet using search engines and using custom Historical Archive or Digital Library systems, either available through the internet or through local installations at the archives' premises). This requirement was set to enable the gathering of a more complete view on information retrieval methods both on digital and printed sources.

## 2.2 Study results

Most of the interviewees found the idea to participate in the survey, recording on a piece of paper their line of thought, the data or information they are looking for in a historical archive, and the queries they formulate to achieve the retrieval of this data/information, very exciting. They told us that it could be very helpful if a methodology for conducting research on historical information was developed, since this could provide valuable guidance for conducting research, especially for the less experienced historical researchers. They also commented that the availability of digital tools to guide researchers through the steps of the methodology and assist them in performing each step would be of great assistance.

More specifically, in order to collect the information on the different ways a historian may address a specific historical question, as well his/her expectations and preferences, we collected and analyzed information regarding the number and type of terms that researchers employ for retrieving the information they require, while searching either in a printed or in a digitized historical material.

Table I depicts the average number of terms used for each IR task, while Table II illustrates the number of terms that are being used by each interviewee for all tasks. Interestingly, it appears that researchers use *fewer* terms when searching in digital corpuses, although searches in this type of corpuses are easier to be formulated and less time-consuming.

Subsequently, we proceeded in categorizing the terms that users had utilized during their investigation, into «terms categories». «Terms categories» actually correspond to the different *meanings* that terms appearing in queries may have. To exemplify this, when the researcher copes with the first IR task *Describe how would you search for information regarding «Kostis Palamas as University of Athens, General Secretary»*, the term “Kostis Palamas” is categorized under “Name of target subject” (and denoted as an *F* element); similarly, in query Q2 “*Describe how would search for information on the historical evolution of the Chemistry Department of University of Athens*”, when the term “Department” appeared on its own, it was classified under the “Part of the target subject’s name” category (and denoted as an *N* element). The goal of this categorization was to deduce the semantic relationships between the entities that are being surveyed and the terms that are used for creating these queries. For digital environments where semantic links between information elements are available (such as ontology-based representations), the tools available to the researcher may exploit this information to present the user with options for formulating and/or enriching his/her queries.

Table III lists some of the most frequently used term categories that were used in the experiment results, while Table IV elaborates one important term category, the concept of «time». According to the users’ answers, time is a specific category and of special importance to the historical research. Primary materials, which include the remaining records of archives, mail, books and time, constitute the basis for original historical research [7]. As depicted in Table IV, time can be used in different ways in a historical research. It appears as either a specific point of time, or to determine a time period, a whole year or more. Even though in the historical research time is of a great importance, the survey showed that researchers do not use it as frequently when searching in digital sources as they do when searching in printed ones. When researchers were asked to comment on why this happens, most of them responded that they had tried in the past to use a time period as a search term in a historical database or in a search engine, but the majority of the information returned was chaotic and irrelevant to their query; therefore, in a digital research they prefer to use specific points of time and not periods of time.

Finally, Table V depicts the term categories most frequently used in each IR task.

Table I. Number of terms used by the by IR task and research method

IR task	Research in printed corpuses	Research in digital corpuses
Q1	2	1,56
Q2	2,62	2,23
Q3	1,42	1,5
Q4	2,46	2
Q5	2,42	2,09
Q6	2,37	2,05
Q7	2,49	2

Table II. Number of researches by subject and research method

Subject	Research in printed corpuses	Research in digital corpuses
101	2,44	1,94
102	2,28	1,67
103	2,08	1,86
104	2,11	1,50
105	2,31	2,13
106	2,27	1,95
107	2,23	2,05
108	2,43	2,15
109	2,65	2,19
110	2,37	2,10
111	1,96	1,71
112	1,96	1,78
113	2,76	2,18
114	2,00	1,85
115	2,21	1,87

### 2.3 Experiment evaluation

According to the researchers' responses during the interview, their majority prefers to search in printed collections, rather than in digital ones. This mainly stems from their experience with searching in digital collections, where the results were poor, since (a) a lot of documents irrelevant to their queries were retrieved (low *precision* [12]) and (b) many important relevant documents were missed (low *recall* [12]). This is reflected in the results illustrated in

Table I and Table II, which clearly show that in digital search researchers used fewer keywords than in the traditional/printed one. According to the users, the metadata that is used by digital libraries or digital historical archives [5], [6], do not cover the needs and requirements of the historical research; especially referring to questions related to the entity evolution [8],[10] like the evolution of an institution or a person. Additionally, in digital search, they used fewer *combinations* of keywords. Another interesting aspect is the fact that even if an *advanced search* was available in the digital tools, researchers confined themselves to use simple search only, neglecting the advanced one. For instance, in the case of Google Scholar [13] where it is possible for the user to pose a query with more detail, e.g. requesting articles by a specific author or published in a specific period, researchers only used the simple search, where terms are given in a "flat" fashion (e.g. if an author name is given, documents containing the designated name in the author list, paper body, footnotes or references are returned). Researchers stated that this is owing to the fact that in many cases they have missed important documents due to metadata incorrectness (e.g. the author's name has been used instead of his/her surname; the year of publication of a conference's proceedings has been used instead of the year that the conference was held) or incompleteness (e.g. the year of publication has not been recorded at all). As for the synonyms and concepts related to the query terms, in digital search, they used less synonyms or related concepts (in some cases none at all) and limited themselves to the keywords presented in the topic.

**Table III** clearly shows the difference in researchers' behavior when searching in printed and digital corpuses. In all categories but one (term category "Part of the target subject's name"), researchers use significantly fewer terms in digital corpuses as compared to the terms they use in printed corpuses.

Table III. Frequency of terms categories' presence per searching method (note: in the traditional method more than 14,89% terms are being used than the digital method)

Term category	Element code	Frequency of terms	
		Research in printed corpuses	Research in digital corpuses
Name of target subject (Kostis Palamas)	F	216	180
Name of a property of the target subject (e.g. Secretary General)	G	84	69
The name of the corpus against which the query will be run (e.g. Historical Archive of the University of Athens)	H	37	19
The name of an entity indirectly related to the target subject; used for enriching the query (e.g. "Works" when searching for "Kostis Palamas")	I	43	20
Any given word that is related to the Subject Name, but not in the query (e.g. term "Professor" for subject "Kostis Palamas")	K	6	1
« Part of the target subject's name » (Department from Department of Chemistry)	N	10	52

Table IV. Frequency of time usage in the digital and traditional search

Term category	Element code	Frequency of terms	
		Research in printed corpuses	Research in digital corpuses
Specific time point (23 October 1845)	A	48	50
Lifespan of the Subject (i.e. period during which the subject existed in the real world)	B	15	6
Time period during which the Subject had a specific property (e.g. Kostis Palamas was a rector of the University)	C	13	7
Every single time point at a specific granularity (e.g. which was the position of Kostis Palamas in the University of Athens <i>for each year</i> )	D	37	3

These results suggest that the digital search has not yet managed to win the trust of the historical researchers. Researchers seem to feel more confident with using the traditional sources in printed format, since they believe to have better access to the historical data required for their research. The lack of trust towards digital historical material can be attributed to the following reasons:

1. High volumes of irrelevant documents are retrieved. Researchers commented that while in the printed versions they considered it "natural" to proceed with an irrelevant number of documents and reject them, in the case of a digitized HA they expected the system to be able to filter-out most of the documents not related to their queries. Coarse-grained metadata (i.e. specification only of the organizational unit that has created a document, rather than a specific author) or lack of metadata is the most usual causes.

Table V. Frequency of terms categories' presence per searching method and IR task

IR task	Term category	Element code	Frequency of terms	
			Research in printed corpuses	Research in digital corpuses
Q1	Name of target subject (e.g. Kostis Palamas)	F	45	30
Q1	Name of a property of the target subject (e.g. Secretary General)	G	19	13
Q2	Part of the target subject's name (e.g. "Department" when searching for "Department of Chemistry")	N	52	81
Q2	Secondary part of the name of target subject (e.g. "University of Athens" when searching for "Department of Chemistry of the University of Athens")	Q	29	30
Q3	Name of target subject (e.g. Kostis Palamas)	F	43	38
Q3	Time frame / Period that is presumed to contain the information searched for	A	12	10
Q4	The name of an entity related to the target subject (n.b. the relation semantics are not <i>isa</i> – e.g. "University of Athens" when searching for "Kostis Palamas")	E	25	35
Q4	The name of an entity indirectly related to the target subject; used for enriching the query (e.g. "Works" when searching for "Kostis Palamas")	I	25	8
Q5	Name of target subject (e.g. Kostis Palamas)	F	61	51
Q5	The name of an entity related to the target subject (n.b. the relation semantics are not <i>isa</i> – e.g. "University of Athens" when searching for "Kostis Palamas")	E	29	17
Q6	Name of target subject (e.g. Kostis Palamas)	F	52	37
Q6	Όνομα οντότητας σχετιζόμενης με ρόλο (όχι <i>isa</i> ) (π.χ. Πανεπιστήμιο Αθηνών (ΠΑ) στον Κ.Π.)	E	37	20
Q7	The name of an entity related to the target subject (n.b. the relation semantics are not <i>isa</i> – e.g. "University of Athens" when searching for "Kostis Palamas")	G	56	47
Q7	The name of the corpus against which the query will be run (e.g. Historical Archive of the University of Athens)	H	25	11

2. Not all relevant material is retrieved. Obscure classification schemes that are meaningful only to the personnel of the archive, missing or erroneous metadata, lack of consistency (e.g. use of abbreviations in some instances and expanded forms in others) and lack of linkage between entities are the most typical causes for failing to retrieve relevant material. It has to be noted here that when searching in printed archives, errors in the documents (e.g. misspelled names) or inconsistencies are tackled, since the researcher can understand the

meaning of the document; in the case of digital archives though, search is performed by software and matching is performed at strict level only. The fact that a researcher first reads a document in the printed archive and then decides if it is relevant or not, alleviates the need for considering all search criteria beforehand, as it must be done in the case of searching in digital archives. The lack of metadata or errors in them could be tackled by broadening the scope of searches, but this action results in retrieving even more irrelevant documents. Some researchers finally commented that they used fewer keywords, synonyms or keyword combinations, because they expected the system to be “smart enough” to retrieve documents that would match synonyms, keyword combinations, or terms semantically associated to the ones given.

3. The fact that query formulation needs the intervention of the HA personnel, who indicates which search terms should be used in which fields is an additional impediment; since it introduces delays (HA personnel may not be always available) and it is an additional source of errors (the personnel’s tacit knowledge may be incomplete or imperfect). We have to note here that the intervention of the HA personnel may in cases be helpful, since the personnel’s experience and tacit knowledge will be exploited to the benefit of the researcher; on the contrary, when the researcher employs digital means, his/her own experience and intuition are the only guides for formulating queries.
4. Finally, there may be issues with the users’ skill levels in computer usage. Besides users’ own responses to questions related to their computer competence during the survey, this can be concluded from the fact that users confine themselves to using simple search instead of advanced search, although the latter can return more accurate results (e.g. entering “Educational” in the simple search may match appearances of the term in *both* the “Organizational unit” and the “Thematic category” metadata slots, whilst if advanced search were employed the user would clearly designate his/her intention).

We also recorded several observations related to the search methods employed by historical researchers. They were using a specific research model based on the given terms of their questionnaire. They tried to identify specific words like names and find if there was any other word related to that name. They tried to connect terms, separate terms, create new terms or synonyms from existing terms, so as to gain as much information as possible related to their historical investigation. [9] presents a project on historical scene investigation, which encourage students in the process of «doing history». In that project, students are provided with a set of questions to guide their analysis and their step-by-step analysis on historical clues.

What historians use, is a step-by-step investigation based on the historical material. Their main objective is to recreate the past, through existing records and their interconnections. The collection of historical data is accomplished through methodical and comprehensive research in primary and secondary sources.

### **3 RESEARCH METHODOLOGY**

Regarding the historical researchers’ methodology for formulating queries in order to retrieve documents relevant to their research, we observed the following practices, which are more or less followed by all researchers:

1. They identify and isolate keywords in the topic of their research. These keywords are very often entities like persons, places or organizations and in many cases the search is restricted by a time point (date, year, etc) or period.
2. They focus on one keyword at a time and look for material in the primary and secondary sources available.
3. They separate compound terms like «Department of Chemistry» into to individual terms («Department», « Chemistry»), in order to isolate two terms and investigate each one of them separately so as to introduce new related concepts.
4. They attempt to perform searches combining more than one of the identified keywords, for example name – date, or place – name – date.

5. They use synonyms and derivatives of the keywords. For example, for the topic “history of the department of Chemistry”, apart from the word “Chemistry” they also use the word “chemical”.
6. The enrichment of the initial terms with new ones is performed incrementally, introducing to the search firstly those that seem more relevant and then the less relevant ones; e.g., for the “Department of Chemistry”, they would introduce “study programme”, “professor” or “book”.
7. They organize terms in a *hierarchical taxonomy* by using a mental model on depicting the related terms closer than the others.
8. They use connections between related terms, by connected terms to the initial ones with relations like “belongs to” or “works at” (For the “Department of Chemistry”, “Faculty” or “University” could be possible related terms).

The phases presented above are the steps of a mental technique that historians use, while searching in a digital or in a printed historical material. These findings can be used for education purposes, since they can be incorporated in a methodology for conducting research on historical information. Furthermore, these findings can serve as the basis of *user requirements*, when building tools to support historical research, since such tools should help researchers perform these phases more efficiently in terms of completeness (all steps are performed and all possible options are available to the researcher to try) as well as in terms of time (document retrieval should be performed more rapidly).

#### 4 ASSISTING HISTORICAL RESEARCH

Based on the methodology results, a number of requirements for the functionality of the tools, that will be available to researchers, can be identified. These requirements are presented in the following paragraphs.

1. The digital repository contents should be tagged with *complete* and *structured* metadata. *Metadata completeness* refers to the need that information typically used in researchers’ queries should be available as metadata. In particular, the topic of the item, its author, date of creation, period to which the content refers, and involved entities (entities referred to in the documents) should be listed within the item’s metadata. *Metadata structuring* refers to the need that this information should be stored as separate fields in the item’s metadata, not as mere keywords. Structuring allows the researcher to gain more control over the search procedure and get more relevant results. For example, if structured metadata are available, the researcher may request documents authored by “Palamas”, whereas if no structure is imposed the query will return all documents that refer to “Palamas”, most of which will be irrelevant to the researcher’s search.
2. The topics available, the document categorization scheme, the entities of interest in the domain of discourse and the timeline covered by the repository should be expressly represented using an appropriate scheme (taxonomy or semantic network), and be made available to researchers, alleviating thus the need to rely on the personnel’s tacit knowledge for conducting a successful search. The scheme should be populated with both generic and specific terms, suitably organized in hierarchies, to allow the researchers to tune the scope of their searches accordingly to the information they have available and the goal of their queries. (Naturally, a search for a generic term should fetch all documents that are tagged with a term that specializes this term). Items should be tagged with the more specific term possible.
3. The manner that users place queries against the archive should be kept as simple as possible, to allow users that are not highly competent with IT systems to work with the system. Advanced search features should be included, but access to *the same functionality* should be also provided to more naive users. For instance, instead of requiring the user to explicitly specify the metadata field against which a keyword must be matched, the user could simply enter keywords as in simple search, and then the system could offer a menu through which the user could disambiguate his/her intentions (e.g. “Does *Educational* refer to the organizational unit *Educational Directorate* or the thematic category *Educational Affairs* (or *both*)?”).
4. The choice of the terms that are used for describing the domain of discourse should be careful, and only terms that are clear and unambiguous should be used.

5. Since the mental model of each researcher for the domain of discourse may differ from the model adopted by the digital repository, the tools must provide means for researchers to align their mental model to the digital repository's model. Drilling down the concept hierarchy, searching, synonyms and thesaurus, as well as descriptive texts for the adopted concepts could greatly assist the researchers in choosing the right concepts.
6. The tool should provide means to limit the scope of searches to points or periods of time, both regarding the time that a document was authored and the time to which the document refers to (the latter is particularly useful for secondary sources). Proposing of terms whose spelling matches most closely the given terms by the researcher could be also used to handle misspelling cases; this is particularly important in the context of historical archives, since names are often found to be written with different spellings in different time periods.
7. Since the retrieval of person biographies and evolution of institution appears to be a frequent query, the tools should assist researchers in locating documents that refer to different periods of the same entity. [7] lists some heuristics that can be employed by such a tool.
8. Given that researchers limit the number of query term combinations when searching in digital repositories, as compared to when they search in printed repositories, the tools provided to researchers could compensate for this reluctance. A possible method could be to ask the researcher for all query terms to be employed and subsequently formulate automatically all possible term combinations.
9. Since researchers have been found to use less synonyms or related concepts when searching in digital archives, the tools could automatically extract synonyms from thesauri and use them in the search; suggesting related terms, extracted from semantic network connections could also assist researchers in the phase of enhancing their queries.
10. The results listed in Table V indicate which term categories are most frequently used for each query type. These results may be exploited to guide the researcher on which terms should be used for formulating and/or enriching his/her query. For instance, if the historical researcher is searching for "Kostis Palamas", the system may detect that this entity is linked to other concepts and/or entities such as "University of Athens", "Works", "Poetry" etc, and present these terms to the user, to be considered for query enrichment.

## 5 CONCLUSIONS AND FUTURE WORK

This work presents a user study aiming to record the historians' information retrieval methods in the context of a Historical Archive. The study was conducted by examining the research methodologies they use while they investigate a historical archive. Through gaining insight to the practices employed by researchers, requirements for information organization and tool support so as to facilitate historical research within digitized repositories of primary and secondary sources can be formulated. Based on an initial set of these requirements, regarding the terms they use, the frequency that each term appears in each IR task and the importance of time in historical research, a prototype tool architecture has been drafted [7] and an initial ontology schema has been designed [11]. The ontology schema has been populated by automatically processing the metadata present in the filenames of the digitized documents; however these metadata are coarse-grained and partial, necessitating thus their refinement and completion. Future work will include the completion of the prototype tool implementation, and the testing of this tool in the context of the Historical Archive. Extending the presented surveys to include subjects working in other archives and/or different historical subjects (e.g. national history) will also be considered.

## 6 REFERENCES

- [1] Investigative Techniques Glossary, <http://www.pbs.org/opb/historydetectives/techniques/glossary.html>
- [2] Primary and Secondary sources, <http://ipr.ues.gseis.ucla.edu/info/definition.html>
- [3] Tibbo, H. R., Primarily History: Historians and the Search for Primary Source Materials, in Proceedings of the 2nd ACM/IEE-CS joint conference on Digital Libraries, 1-10, 2002

- [4] Mark Vajcner, The importance of context for digitized archival, Journal of the Association for history and computing, Volume XI, Number 1, April 2008
- [5] Ian H. Witten, David Bainbridge « How to build a digital library», Morgan Kaufman publishers, 2003
- [6] Victoria Irons Walch, Marion Matters, Standards for Archival Description: A Handbook, The Society of American Archivists, 1994 <http://www.archivists.org/catalog/stds99/toc.html>
- [7] A. Katifori, E. Torou, C. Vassilakis, C. Halatsis, Supporting Research in Historical Archives: Historical Information Visualization and Modeling Requirements, Proceedings of IV 08
- [8] Torou, E., Katifori, A., Vassilakis, C., Lepouras G., Halatsis, C., Creating an Historical Archive Ontology: Guidelines and Evaluation, Proceedings of the First IEEE International Conference on Digital Information Management (ICDIM 2006) December 06-08, 2006, Bangalore, India
- [9] Kathleen Owings Swan and Mark Hofer, The historical scene investigation (HSI) Project: Facilitating historical thinking with web-based, digital primary source documents, Journal of the Association for history and computing, Volume XI, Number 1, April 2008
- [10] Katifori A., Torou E., Vassilakis C., Lepouras G., Halatsis C., Daradimos E., Historical Archive Ontologies – Requirements, Modelling and Visualization, Proceedings of the RCIS 2007 Conference
- [11] Torou, E., Katifori, A., Vassilakis, C., 2007a, University of Athens Historical Archive Ontology Version 1, <http://oceanis.mm.di.uoa.gr/pened/?category=pub#ontos>
- [12] Wikipedia, 2008. Precision and recall, [http://en.wikipedia.org/wiki/Precision\\_and\\_recall](http://en.wikipedia.org/wiki/Precision_and_recall)
- [13] Google inc, 2009. Google Scholar advanced search. [http://scholar.google.gr/advanced\\_scholar\\_search](http://scholar.google.gr/advanced_scholar_search)