Creating an Historical Archive Ontology: Guidelines and Evaluation

Torou Elena
Dept. of Informatics and Telecommunications, University of Athens, Athens, Greece
etorou@di.uoa.gr

Katifori Akrivi
Dept. of Informatics and Telecommunications, University of Athens, Greece
vivi@di.uoa.gr

Vassilakis Costas
Dept. of Computer Science and Technology, University of Peloponnese, Tripolis, Greece
costas@uop.gr

Lepouras Georgios
Dept. of Computer Science and Technology, University of Peloponnese, Tripolis, Greece
gl@uop.gr

Halatsis Constantin
Dept. of Informatics and Telecommunications, University of Athens, Greece
halatsis@di.uoa.gr

Abstract

Ontologies have been proven invaluable tools both for the semantic web and for personal information management. In the context of a historical archive an ontology may provide meaningful and efficient support for search tasks as well as be used as a tool for storage and presentation of historical data. The creation however of such an ontology is complex, since the digitized archive documents are not in text format and the concepts that must be captured may vary among different time periods. This work presents a user-centric methodological approach for extracting the ontology of an historical archive focusing on the evaluation issues related to this process. The approach is exemplified through cases from its application in the University of Athens Historical Archive.

1. Introduction

In the past few years, libraries are increasingly incorporating digital technologies to make their material available to more users through the Internet and facilitate search and retrieval within the environment of the library itself. Historical archives all over the world are also starting to make an effort to digitize their material, integrate electronic search and display facilities to the existing paper-based archives and in some cases- proceed even further, making them available on-line.

The digitization process in the context of historical archives is inherently more demanding than the equivalent in common digital libraries mainly due to the large volume of the original material and its poor preservation state, as well as to the convoluted and archaic handwriting often found in documents of historical archives. At the best case, keywords or other metadata (creation date, author etc) will be available.

Commonly, documents in a historical archive are fitted into a categorization scheme, which has proven to provide little help for information retrieval (IR) purposes, as it is typically compiled by archivists to suit archiving purposes. As a result, even browsing becomes very difficult without help from experienced archive personnel, which mainly relies on their conceptual model of the archive, rather than on some explicit representation of knowledge about the archive content and tools facilitating search tasks.

Taking the above into account, it is essential to develop methods for supporting users during their searches in the historical archive environment. A useful instrument in this context is the *archive ontology*, which captures the concepts within the archive, relationships between them, properties that describe them as well as individual data for specific instances.

Ontology creation within this context has to address various issues, including differences in concepts among time periods, changing roles of entities and the large number of information sources. It is essential not only to reflect in the ontology the evolution of the organization, but also to create an ontology that will be meaningful for users and useful for IR. To this end, an ontology development methodology should include an evaluation stage, which will provide feedback to the ontology formulation phases, to allow for better tailoring of the ontology to the actual user needs.

This work presents a methodological approach to ontology development, focusing on the evaluation issues related to this process. The approach is exemplified though cases from its application in the University of Athens Historical Archive. The rest of the paper is organized as follows: section 2 provides background information and overviews work related to ontology design and evaluation. Subsequent sections describe the proposed methodology for the creation and evaluation of a historical archive ontology, along with a case

study of creating an ontology for the University of Athens Historical Archive.

2. Background and Related Work

An ontology is a formal explicit description of a domain, consisting of *classes* which are the concepts found in the domain [1]. Each class may have one or more *parent classes*, has *properties* or *slots* describing features of the modeled class, and *restrictions* on the slots. Class *instances*, correspond to individual objects in the domain of discourse; each instance has a concrete value for each slot of the class it belongs to.

In relation to digital libraries, historical archives possess certain special characteristics: (a) libraries contain generally independent texts or series of texts whereas historical archives contain material pertaining to a single organization (b) the majority of the historical archive documents are not available in full text form but only as images [9] and (c) temporal relationships, entity evolution and timelines are of particular importance in historical archives. These historical archive characteristics necessitate the provision of additional tools and services for IR, as compared to a digital library.

Work published insofar for facilitating the ontology engineering process has employed both manual and semi-automatic methods. Semi-automatic methods focus on the acquisition of ontologies from domain texts. In [2], for example, a framework incorporating several information extraction and learning approaches is proposed with this objective. Interesting surveys of existing methodologies can be found in [3] and [6]. Throughout the ontology creation process, the designers may take into account a set of ontology design criteria, such as clarity, coherence and extensibility [7].

Many works are available on ontology evaluation. [13] presents a survey of existing approaches, categorizing them in (a) *golden standard*, based on comparing the ontology to a reference "golden standard" ontology (b) *application–based*, base on using the ontology in an application (c) *data-driven*, involving comparisons with a source of data relevant to the ontology domain and (d) *assessment by humans*, who evaluate the ontology based on a set of predefined criteria.

In the context of historical archives, however, semi-automatic ontology extraction is only of limited use, since documents are generally available only in image format, and OCR cannot be efficiently applied because documents are mostly handwritten and often employ calligraphic or convoluted handwriting. Moreover, none of the existing ontology development methodologies addresses the matter of time and evolving ontologies. Finally, the issue of ontology evaluation is considered to be totally separate from its development.

Our work aims at complementing existing approaches providing an integral framework for (i) developing ontologies accommodating evolution and temporal relationships and (ii) integrating evaluation procedures into the ontology development process, for improving the quality of the final outcome. For the evaluation procedure, in particular, the presented work mainly employs the application-based and assessment by humans approaches, since "golden standard" ontologies are currently not available and data sources required for the comparisons of the third approach are scarce in the context of historical archives.

3. Creating an Ontology for the Historical Archive

In order to create the ontology depicting the various states of the University of Athens, we used a top–down, present to past approach. The basic idea was to create a basic set of upper level classes and then expand it by firstly enriching it with entities related to the current state of the organization and then recording the evolution of these entities from their current state back to the initial state of the organization. The ontology of the current state of the organization is a good starting point, as it represents the end of the organization evolution, which is, in most cases, more complex than the ontology of the organization's initial state. Then, working backwards in time, it is easier to identify the history of individual entities and their transformations with the passage of time.

Each step of this process is performed relying on various sources and results in the identification of several ontology versions. The steps followed are (a) identify most important *upper level classes* (b) enrich the ontology with classes and relationships relevant to the *current state of the organization* (c) record the evolution of the ontology by investigating the history of individual classes and (d) evaluate the ontology.

This process, depicted in Figure 1 is an iterative one, as most probably the evaluation stage will raise issues that need to be addressed. Each such issue will be used as input to the pertinent stage, which will produce a rectified output; this output, in turn, may trigger remedial activities in subsequent stages, e.g. if the "current state of ontology" is modified then the "create and link past versions" phase needs to be revisited.

Note that in the context of the proposed approach, "ontology versions" are used and considered only by ontology developers for better organizing the material in the ontology creation process. The final outcome of the process is a *single ontology*, including all classes

relevant to the organization's history. Each class in this ontology is timestamped with its validity period and linked to its previous and next versions, as appropriate.

Producing a single ontology was opted for because the evolution of the entities that are relevant to the university has been very complex, especially in the first years of its establishment. Creating a new version for every structural change that had occurred would lead to too many versions, rendering the final model excessively complex. Furthermore, current ontology management tools (e.g. Protégé [11] and Kaon [5]) do not support browsing and searching across different versions, thus use of multiple versions would hinder the users' IR tasks. The following sections describe the ontology development phases in detail.

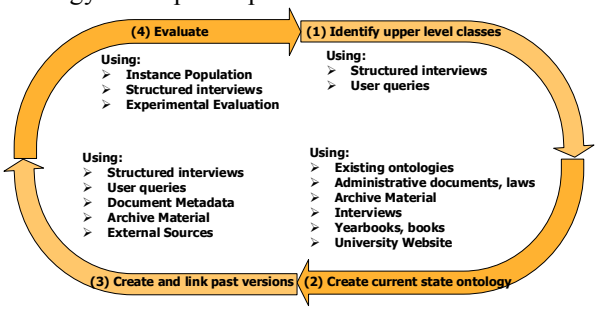

Figure 1. Ontology creation life cycle

4. Upper Level Classes Identification

A first step in creating the ontology of the historical archive is to identify a core set of high-level classes that are relevant to the material contained in the archive as well as user needs. It is important to create an ontology that not only expresses wholly and comprehensively the domain but it is also useful and effective as an aid for IR. Once the relevant classes are identified, they are structured in hierarchies and linked through relationships in a way that best serves user requirements. These requirements are captured by investigating user interests and their ways of researching archive material. To this end, two sources of information were used: structured interviews and analysis of queries already made to the archive.

4.1 Structured Interviews

Besides elucidating the core set of high-level classes, structured interviews additionally served the purpose of identifying user requirements for a digital historical archive, and helping to understand the way users search the material available. 12 interviews have been carried out, the interviewees being 5 professors, 4 employees of the Administration and 3 belonging to the Historical Archive of the University of Athens. Our aim was to record different perspectives about the University, leading possibly to different ontology concepts. The interviewees were firstly introduced to the concept of the ontology and then were asked to describe their notion of the University in relation also with their work and discuss terms which they find relevant. They were available for follow-up interviews to review sets of classes and their hierarchy compiled.

The final results of this research are not available yet; however the preliminary results provide some very useful insight regarding the general classes that researchers are interested in. A part of the results is presented in the first column of Table I.

4.2 User queries

The queries that users have made to the historical archive requesting documents were used to identify the classes most frequently considered in IR tasks. Knowing these classes is essential to the ontology creation process, since they can be used to extract the upper level classes, as well as slots and relationships associated with them. For example, for the query "what was the name of the professor that served as Dean in the University in 1912" the key high-level classes identified are "Professor" (and its super-class "Person") and "University" (and its super-class "Academic Institution"). Both classes "Professor" and "Academic Institution" have a property "Name", while a relationship among these classes labeled "Dean" is also revealed. The analysis of approx. 100 user queries resulted in several high level classes that were of interest to users. These classes were selected to be the core of the ontology. The second column of Table I contains some of the terms extracted form queries.

Table I presents an example of the process of combining structured interview and query term analysis results. The first column presents terms derived from structured interviews, while corresponding query terms are presented in the second column. These terms have been organized into classes and sub-classes in the third and fourth columns. Some of the terms, like *name*, have been added as slots. Interview-derived terms such as "name" and "occupation" were related to their corresponding query terms and were the basis for creating the super-class "Person" and its sub-classes depicted next to it in Table I. When all direct subclasses of a class had a common slot, this slot was moved upwards to the parent class (e.g. the *person name* was moved upwards from the subclasses to the *Person* class).

Table 1. Results from the structured interviews and user queries analysis.

Structure Interview Terms	Query Term	Class (Slots)	Sub-Classes(Slots)
Name	Person Name	**Person** (Name)	**Professor(…)** *Student(…) Secretary(…) Dean(…)*
Occupation	Professor		
	Person Occupation		
	Student		
	Secretary		
	Dean		
Date Year	Year	**Time**	*Time Instant* ***(Date, Time)*** *Time Period(* ***Start, End)***
Place Name	Place Name	**Place** (Name)	*Building(…)*
	Building		
	Faculty		
University, School, Faculty, Institution	School	**Academic Institution(…)**	*Faculty(…) School(…) University(…)*
	Institution		
	University		
Administration	Senatus	**Administrative Body(…)**	*Senatus(…)*
	Administration		
Outliers	Museum	**University Outlier(…)**	*Museum(…)*

5. Current Organization State Ontology

In order to create the organization's current state ontology, numerous sources can be employed. The following are considered as the most important ones:

1. *Existing university ontologies*. We searched for detailed university ontologies that could serve as sources for classes and could provide supplementary views for their structuring and their interrelations. The only university ontology we located is the one in [4], which served as a first basis mostly for the "Person" and "Publication" sub-hierarchies, after being translated and adapted to the Greek university environment.
2. *Information about the structure of the University of Athens*, as it is presented in its website [8].
3. *Yearbooks*, containing information about university personnel were used for extracting classes relevant to the professors and employees of the university.
4. *Books* relevant to the university.
5. *Laws, directives and regulations*, providing useful insight as to the university structure and practices.
6. *Interviews with university professors and administrative personnel* (cf. section 4.1).
7. *Administrative documents and existing categorizations*.

In recent years, the largest part of the University documents, are kept electronically. These documents are not yet part of the historical archive, but will be incorporated into it after a certain time period (currently, 30 years after document creation). For these documents, semi-automatic ontology extraction could be applied to identify classes. We used Kaon's [5] *text-to-onto* ontology extraction tool, which however does not include a Greek dictionary, so its full potential could not be exploited for our tasks. Automatic exclusion of stop words ("and", "or", particles, etc), identification of different forms of the same word (e.g. "president" and "presidents") and identification of multi-word terms (e.g. "Assistant Professor")- were thus disabled. The results produced from this stage were further refined manually, to compensate for the unavailable functionality. The categorizations currently used for the filing of documents were also used as input for this phase.

6. Tracking the Evolution of Entities

After creating the ontology of the current state of the organization we proceeded in tracking the evolution of the individual classes and instances backwards in time. Class and instance evolution includes changes in naming, addition/deletion of slots and relationships, merging or splitting of entities, and differentiations in class hierarchies. Relationships between concepts were additionally exploited to identify classes that were valid at past time periods but not today (and thus have not been modeled in the initial ontology). For performing these tasks, various information sources were investigated, as described in the following paragraphs.

a) Interviews with the archive personnel and historians. Domain experts provide their knowledge on the archive material together with their understanding of user needs. Through interviews they gave guidance as to where to find the necessary information and the concepts we should focus on.

b) Reviewing the queries made to the archive. These queries have proven useful as well, as they contain concepts that do not exist today in the context of the university but have previously been essential, for example "Chair" as in "Chair of Physics".

c) Reviewing the archive material. The most important source and the basis for the historical archive ontology is the archive material itself. In our case, the size of the university of Athens historical archive has been estimated to 4 million documents, in either handwritten or in typed form. Only 10% of them were digitized in image format, and for 70% of the digitized material certain metadata were available-more specifically, their Administrative Body Provenance, their type (e.g. "Proceedings") the date and thematic category. These metadata were exploited, as described in the following paragraph. Some parts of the material, namely the typed documents that are in a relevantly good state, were used to perform OCR and then term

extraction. The rest were reviewed selectively, following the archivists' suggestions, in order to identify classes that were not found up to that point.

d) Using possible metadata information on the documents. Metadata information in the archive material is generally not available; however, in some cases there exist annotations on the documents made for archiving reasons from past archivists. These annotations have proven very useful for extraction of ontology classes and/or slots. There is an ongoing effort, parallel to ours, to systematically record and digitize this metadata, but only preliminary results have been made available; these results were also exploited. In general, metadata were used for class and relation extraction as they provide insight as to the important terms of that period and their relations and groupings.

e) Yearbooks and Dean Speeches. Yearbooks were being compiled once a year, since 1863 and contain information about university structure, education programs, administration, professors, etc. *Dean Speeches* were also compiled each year since 1837, and contain all speeches given by the deans. During the first years, Dean Speeches contained detailed fiscal information.

f) *External sources to the archive, wherever applicable.* In some cases, historical studies related to the material of the archive and the history of the organization may be useful for extracting information about the ontology classes and class hierarchy. In the case of the University of Athens there was only one such work available, a doctorate study relevant to the reforms in the University at the beginning of the 20th century.

Table 2. Example of the evolution of instances and classes from the university establishment until today

Instance Name	"Othonian University", renamed "National University" in 1862	"National University" containing the Medical Faculty and the Faculty of Natural Sciences	"National and Capodistrian University of Athens"
		"Capodistrian University", containing the Philosophy Faculty and the Faculty of Law	
Period	22/05/1837 to 1911	1911 to 1922	1922 to today

Class Name	Chair (Name, Professor, Faculty)	Department(Name, Teaching Staff, Faculty)
Period	22/05/1837 to 1982	1982 to today

Using all these sources, we compiled tables representing the evolution of individual classes and instances in time. New classes were also listed, along with their properties and relations to existing classes. Dates and time periods accompanied each piece of information. An excerpt of this recording is illustrated in Table II. These tables were used to create an ontology using Protégé [1] as an editing tool.

It should be noted here that the representation of the temporal aspect of classes and instances required modelling from our part, as it is not adequately supported by existing ontology management tools.

The complete ontology evolution was modeled in a single ontology, for the reasons presented in section 3. The ontology was first expanded with two auxiliary classes for time representation, namely Time Instants and Time Periods. Instances of these classes were used to flag the validity periods of classes and instances alike. Figure 2 illustrates a portion of the ontology, as designed in the Protégé tool.

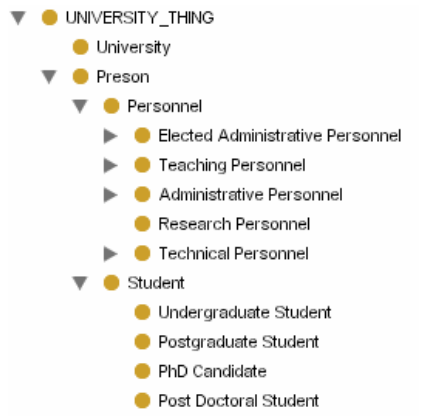

Figure 2. .Part of the ontology in Protégé

Information concerning class and instance evolution was recorded in the form of comments, e.g. for the Chemistry Department a user could read the comment "It was established as the evolution of the Organic and Inorganic Chemistry Chairs". An additional difficulty for providing such comments existed for persons, since cases of synonymy or incomplete data were found, in which it was impossible to deduce if documents actually referred to the same person. It was finally decided that for cases with "adequate confidence" comments should be entered, while for other cases comments would be omitted so as not to mislead the users.

This initial ontology model was evaluated with the help of the archive personnel and users. The evaluation procedure is described in the following section.

7. Ontology Evaluation

The evaluation of the historical archive ontology is a very important step in its life cycle. Since the ontology is intended to be used for storage of information related to the archive material and consequently as a tool for IR tasks, it should be adapted to both the material and user needs. User needs should be reflected in the ontology classes and their structure, their properties and interrelations. This means that of all the possible sets of classes, property slots and inheritance and role

relations between them, an appropriate sub-set should be chosen so that the ontology is created with the appropriate level of detail. Too few classes may not be enough to represent all the necessary domain concepts, whereas too many may hinder the user during IR.

As a result, ontology evaluation should be accomplished through the cooperation with the historical archive personnel and users. We employed three methods of evaluation during the ontology design process in order to capture as closely as possible user needs. These are: (a) evaluation during the population of the ontology with instances, (b) interviews and (c) evaluation through an experiment. These methods are presented in the following sections.

7.1 Evaluation during the ontology population with instances

The first set of modelling errors was found during the preliminary ontology population with instances. At the end of each ontology creation cycle, the archive personnel were asked to assist us in inserting some representative instances into the ontology. The personnel proposed a set of instances and sources for material about these instances and, after studying the material, the main characteristics of each entity were identified and the appropriate instances were created in the ontology to host these characteristics. This procedure revealed omissions and errors in classes and their interrelations: some entity characteristics did not have an appropriate slot for being stored or even some classes were found to be missing; some slots were characterised as "misplaced", i.e. they were considered to fit more naturally within other classes; finally, few class hierarchies were altered, since the originally formulated hierarchy was considered dubious when seen with instances in it.

7.2 Ontology evaluation through interviews

The next step was to perform a set of structured interviews in order to evaluate the ontology. The ontology was presented to domain experts, who were asked to make comments about errors, omissions or parts of the ontology that are elaborated on more than necessary. Note that in this context, the term "domain experts" includes not only historical archive personnel but also professors and university administration personnel, as described in Section 4.1. While the historical archive personnel have knowledge about the history of the university organization, the rest may have an extensive knowledge on particular university domains, such as its administrative structure and procedures or its educational work. The comments collected from this stage can be classified in two subcategories: The first category includes comments collected from the archive personnel, which were mainly related to the changes made in the previous step, e.g. if a class was added in step (7.1), new relationships or classes could be proposed, since –when viewing the new classes- the archive personnel recalled concepts that were not considered in the initial stages of ontology creation.

The second category includes comments made by professors and university administration personnel, which mainly referred to the whole ontology; this was expected, since this was the first time subjects viewed the ontology. A number of classes, slots and relationships were suggested by the subjects, since each one of them has much deeper knowledge on his/her domain than the archive personnel. These subjects characterized some of the classes within the ontology as "over-detailed", mainly those not included in their domain of expertise; interestingly enough, the archive personnel characterised as "over-detailed" a number of suggestions made by these subjects. When such contradictory views were encountered, a short session was held between the disputing parties to settle the issue.

7.3 An experiment for ontology evaluation

The final evaluation stage focused on the ontology's intended usage, i.e. as an aid for storing and retrieving historical data. To this end an experiment was designed, for evaluating the effectiveness of the ontology for this purpose. While the experiment was designed for the purpose of evaluating the ontology, it also provided useful insight regarding two other aspects, namely (a) evaluation of 4 ontology visualization methods and (b) strategies and techniques employed by users while researching historical material.

Using four different visualization methods was considered useful for removing any effect that a particular visualization might have on the results of ontology evaluation, since users might make unfavorable comments on the ontology because they did not like the visualization tool or vice versa. The experiment setup and results relevant to the evaluation of visualization methods are described in [12]. For the purpose of this work, only the results related to the evaluation of the ontology will be discussed.

Most of the users that participated in the experiment were students of history-related departments and researchers working in the Department of Informatics and Telecommunications of the University of Athens. All these users have some knowledge regarding the classes of the "University" domain but a varying degree of computer expertise. The user group was composed of 5 men and 9 women. 8 of them are students or researchers of computer science departments, while the

remaining 6 are students or researchers that have at least once visited the Athens University Historical Archive or another Archive for research purposes.

During the experiment the users were asked to complete a set of IR tasks ranging form simple ones (e.g. finding the establishment date of a department) to complex ones (such as retrieving a person's biography). These tasks were chosen so as to reflect common query types made to the archive and categorized according to ontology-related criteria. For a more detailed description of the tasks, the interested reader is referred to [12]. The recorded times, notes regarding actions users performed to complete the tasks and comments made by the users were analyzed for assessing the ontology. The conclusions drawn from this analysis are listed in the following paragraphs.

a) The need for entity history and time modeling. All users commented that it would be useful to have a direct connection to previous and future states of an instance, i.e. explicitly model the entity history, in order to be able to access directly this information. The implicit reference to the past state of an entity through the "comment" slot was not generally helpful; some users stated that they were put off from investigating the "comment" slot because it was lengthy (since it accommodated *both* the entity history *and* the "actual comments"). As a result the percentage of incomplete answers to questions involving entity timelines was more than 80%.

Many users expected to find useful information in the "Time" classes, such as events related to certain dates. Since such information was not comprehensively available under the "time" classes (only instance lists could be found), it would be preferable to hide these classes, to avoid user distraction.

In the case of person biographies, where no reference had been made to previous roles of the person in the instance information, users faced the task in mainly two ways: some were satisfied by retrieving only one of the instances related to the person with the given name e.g., if they located the "Undergraduate Student" instance related to the person they were satisfied with the result, ignoring the existing "Postgraduate Student" instance. Other users systematically browsed through the classes and located all the instances, but did not thoroughly check if these instances represent in fact the same person or it was a case of synonymy. Only a small part of the users (the most experienced history researchers) elaborated on this issue and utilized additional information slots (e.g. the date of birth) to verify if the instances indeed refer to the same person.

Taking the above issues into account, the validity period already present in the ontology meta-schema was complemented with the following fields, to further facilitate history- and time-related IR tasks:

Next. This slot is used when an ontology element is no longer valid and should be replaced by another element or elements. For example, if a student named "John Smith" graduates and starts to work as a researcher, the "Next" field in the old instance (of class "Student") is set to point to the new instance (of class "Researcher"). The "Next" slot may also be assigned multiple values to allow for modeling the cases that the class, instance or slot was split in two or more.

Previous. This slot is assigned a link to the ontology element (or elements) that the current element has replaced. Or if the class, instance or slot was a result of merging two or more classes, then this field is assigned with a set of references to the classes, instances or slots the current element has originated from.

However, in the case of persons these fields are not enough. There are cases where a person may have more than one role in the university at the same time (e.g. be a professor, member of the senatus and a director in a laboratory). To accommodate this need, the "**Other Roles**" slot was added so as to store this information.

b) Generic vs. Specialized slots. An issue faced with designing the University ontology was the large number of "has-a" type relationships within certain classes. For instance, the "Department" class *has-a* "Undergraduate Study Programs", "Department Sectors", "Faculty", etc. Two approaches were considered for storing this information within the ontology: the first one was to use a single "contains" slot listing references to all related instances, while the second was to use a separate slot for each different type of semantics. Initially, the first approach was adopted, pursuing the minimization of the overall number of slots. However, this choice was commented negatively by the users: it seemed neither useful nor correct to them to see, for example, in the "Department" instance "Undergraduate Study Programs" and "Department Sectors" under the same "contains" slot. As a result, such slots were split to two or more others depending on the class.

c) Upper level classes naming and grouping. The experimental evaluation of the ontology has proven useful in identifying possible problems with the naming and grouping of the upper level classes, which are of particular importance because they are the starting point for IR tasks, especially in the Windows Explorer-like visualizations like the Protégé [11] Class Browser and Jambalaya visualizations (Jambalaya may be considered analogous to the View/Thumbnails setting).

It was evident in the experiment that when a user did not understand or notice the upper level class (or classes) relevant to his/her task, the task completion time increased significantly. As a result of the evaluation, these classes were fine-tuned to be more helpful for IR needs. For example, in our case the concepts

"Lesson", "Study Program" and "Program Direction" were grouped under a general class named "Educational", as suggested by the archive personnel who participated in the experiment, since this was considered to be a more intuitive way to present these classes.

Top-level classes cannot always be changed though. A trade off exists between aiding IR and having an ontology useful for representation of historical data. An example in this case is the "University Outlier" class, a term used throughout the Athens University history to denote institutions that were property of the university, like museums and hospitals. However, the majority of the users did not recognize immediately the role of this class. It was thus discussed whether this name should be changed to something more meaningful to the users. The archive personnel opposed to this change, stating that this is the correct name for this class and that it should remain as is; they also pointed out that the ontology should also serve educative purposes for the researchers, presenting the University structure and classes as more closely as possible to reality. They suggested the addition of some type of comment slot or short description to the upper level classes in order to clarify their content.

d) Identification of commonly used classes. Taking into account that the university ontology contained data for a single university, no special attention was given to the creation of a detailed "University" class – all information could be located through the other classes. However, users often turned to this top-level class to complete various IR tasks. They browsed the "University" class intending to retrieve specific "Faculties", "Administrative Bodies", "Museums", which they expected to find linked through "has-a" slots. As a result of the evaluation, the "University" class was modified by adding to it role relations pointing to other classes, including the ones listed above.

The ontology versions are currently available in Greek in RDF format [10]. Soon, they will also be available translated in English as well.

8. Conclusions and Future Work

Creating an ontology for a historical archive is not an easy task, mostly due to the nature of the material, in the majority of cases not available in text format and due to the temporal nature of the ontology. In this paper, we have presented an approach to this task that takes into account all the available sources as well as user needs in order to create an ontology that would be useful for IR purposes. Furthermore, we present a user-centric evaluation method for an historical archive ontology along with some guidelines as to what ontology features should be given more attention. We are currently on the stage of re-evaluating the second version of our university ontology in order to gain further insight into the ontology evaluation process. Future work includes elaborating on the evaluation of ontology visualization methods for the context of the historical archive, as well as developing visualization aids for presenting entity evolution. More information on these visualization aids may be found in [14].

9. References

[1] Noy, N. F., McGuiness, D. L.: Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, March 2001, protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

[2] Maedche, A., Staab, S.: Mining Ontologies from Text. EKAW 2000, 189-202

[3] Cristani, M. & Cuel, R.: A Survey on Ontology Creation Methodologies, International Journal on Semantic Web and Information Systems, Vol. 1, No. 2, 49 – 69, 2005

[4] University Ontology (draft), www.cs.umd.edu/projects/plus/SHOE/onts/univ1.0.html

[5] Kaon, http://kaon.semanticweb.org/

[6] Noy, N. F., Hafner, C.: The State of the Art in Ontology Design, A Survey and Comparative Review, AI Magazine, 18 (3), Fall 1997, 53-74.

[7] C. Fluit, C., Sabou, M., van Harmelen, F.: Ontology-based Information Visualisation, In Visualising the Semantic Web, Springer Verlag, 2002

[8] University of Athens, http://www.uoa.gr

[9] Daniel V. Pitti, Encoded Archival Description, An Introduction and Overview, D-Lib Magazine, Vol 5, No 11, November 1999

[10] Athens University Ontology, oceanis.mm.di.uoa.gr/pened/index.php?category=publications

[11] Protégé, http://protege.stanford.edu/

[12] Katifori, A., Torou, M., Halatsis, C., Vassilakis, C., Lepouras G. A Comparative Study of Four Ontology Visualization Techniques in Protégé: Experiment Setup and Preliminary Results, in Proceedings of IV 2006

[13] Brank, J., Grobelnik, M., Mladenić, D., A Survey of Ontology Evaluation Techniques, In Proceedings SiKDD 2005, Ljubljana, Slovenia, 2005.

[14] Katifori, A., Vassilakis C., Lepouras, G., Daradimos, I., Halatsis, C., Visualizing a Temporally-enhanced Ontology, In Proceedings of the AVI 06, Venezia, Italy, 2006.