Ontology Aided Information Retrieval in Digital Historical Archives

Katifori Akrivi
Department of Informatics and Telcommunications,
University of Athens
vivi@mm.di.uoa.gr

Golemati Maria
Department of Informatics and Telcommunications,
University of Athens
margo@di.uoa.gr

Lepouras Georgios
Department of Computer Science and Technology,
University of Peloponnese
G.Lepouras@uop.gr

Keywords: historical archives, visualization, ontology, information retrieval

ABSTRACT

In the age of digital information more and more digital libraries and historical archives are using information systems in order to facilitate the document retrieval and provide better visualization of the search results and document presentation. Much research has been done in the field of digital libraries, but in the case of historical archives, which have particular needs, this is not the case. To this end, we investigate the use of new tools, which are based on the ontology of the historical archive in order to provide a new and effective method for document retrieval in a dynamic environment which will take into account the collaboration needs of the users.

INTRODUCTION

New versions of advanced telematics applications appear everyday in our life and affect the way one performs tasks such contacting governmental services or visiting the doctor. To this end, information systems have become a common place in most libraries. However, searching, retrieving and analyzing historical documents and archive material in general is not an easy task [1]. The researchers have to face a large amount of documents that have never been studied and, in some cases, have not even been read, in search of "interesting" historical facts, which, if they are discovered and published, may confirm or refute existing theories. Currently, efforts are being made in archives all over the world to digitize their material, integrate electronic search and display facilities to the existing paper-based archives and, in some cases proceed even further to make them available on-line. Especially, the use of Internet technologies in the environment of a historic archive can allow access to the material to a large number of users, whose only option until now was to physically visit the archive itself.

According to a recent study [2] on how historians locate primary resource material in the digital age, in order to find documents relevant to their research interests historians use both keyword searches and browsing of the archive categorizations and indexes. However, index browsing in the context of a digital library has not been bibliographically covered to the same extent with keyword searching. Furthermore, in a historical archive, the search for documents relevant to particular research problems is usually carried out, both by the researches and by the archive personnel, in a try and error manner. Users have to rely mainly on the conceptual model of the archive that they have developed through their experience from past searches and less on a tool that, to some extent, will guide and automate their searches.

Another important fact that hinders the search in a digitized historical archive is that most documents are not available in the form of electronic text, but only as images and, as a result, full-text keyword search is not available. Consequently, it is of the essence to develop alternative methods of providing support to the users during their searches in the environment of the historical archives.

The present paper provides a brief overview of the techniques currently used for browsing and visualizing results in historical archives and proposes a new approach aiming to help users in performing their search by providing a conceptual model of the entities found in the archive, and bridging this to the existing categorization.

The next, second section describes methods for browsing historical archives and digital libraries in general, while the third section describes the proposed approach as a generic framework for searching and viewing documents specifically in historical archives. The fourth section summarizes key points that have to be taken into consideration during implementation and outlines future research aims.

INFORMATION RETRIEVAL METHODS FOR DIGITAL LIBRARIES

Historical archives are considered to be a special case of a library, where the documents contained are the product of the activities of the same entity (organization or person). To this end, when an historical archive is being digitized the methods provided to support the functionality needed for the digital archive are usually similar to the ones used for digital libraries.

Up to this day, the research in the field of user interfaces for digital library systems has developed a great variety of proposals about document search and visualization of its results. The most widely used solution is that of the web search engines, where the

input of one or more keywords results in a document list that may contain even hundreds of results, divided in pages. The browsing of hierarchies is accomplished through multi-level lists. The user may view the contents of a category by selecting it. However, these methods do not offer effective ways of an overall presentation of the results. Consequently, efforts have been made, by using 2D and 3D visualizations, to represent more efficiently the search results and facilitate the hierarchy browsing.

Many 2D presentations, as IVEE [6], use graphs that visualize the documents as points in the 2D space, in order to provide an overview of the results of a keyword search. The axes are defined according to certain pre-selected document characteristics and they provide a query formulation area with various tools to aid the keyword search.

On the other hand, there are several 2D systems that represent hierarchies in order to facilitate the browsing of large document spaces. All these techniques employ a similar point and click navigation method, suitable for browsing the visual representation of the document space.

For example, GRIDL [5] does not offer a sophisticated query formulation area, but may represent hierarchies in the form of a zoomable table with a technique called Hieraxis, allowing the user to navigate in the hierarchical data.

Other visualizations, like the hyperbolic browser [7], use graphs in order to group the results in hierarchies. An alternative way to represent a hierarchy is by using nested colored areas, which could be rectangular surfaces, as in the Tree Map [8], semicircular disks as in the Information Slices [9] technique or circular as in Grokker [10].

A number of methods, such as Compus [1], use a type of bar graph to represent visually document attributes. Other representations, like Scrollbars [12] and Tilebars [11], use the bars combined with a document list as an aid for the user in order to determine the relevance of the document. Such visualizations are used to display the results of keyword searches.

More realistic metaphors are used by Infosky [13], which represents the documents as stars and their groups as star clusters in the night sky and Websom [14], which uses a geographic map metaphor.

3D visualizations could be grouped in two large categories: the realistic and the abstract ones. The realistic visualizations use real world metaphors, such as a wall, a city or a library for the representation of the document space. The abstract visualizations use representations such as 3D graphs to represent the documents, their hierarchy and grouping.

The realistic 3D visualizations employ a great variety of metaphors. A common one is the landscape metaphor, where documents are placed on a plane as color and size coded 3D objects. This is the case of File System Navigator (FSN) [15] and Harmony Information Landscape [15].

The Data Mountain [16] visualization method for document management again uses a plane for the document placement, but in this case the method is not fully 3D, as the interaction technique is 2D.

Another group of techniques uses perspective and distortion to combine focus and context in the presentation of the document space. Such techniques have a central panel for the information on focus and side panels, two as in the Perspective Wall [17] or more for other systems, for the information context.

The Task Gallery [18] uses a corridor metaphor, the walls, ceiling and floor of which are used for the presentation of the documents.

Another real-world metaphor used as a 3D visualization method is the book metaphor. The WebBook [21] visualization simulates a book containing a collection of documents. The Web Forager [21] provides an environment for the presentation of the WebBooks, which contains a library and a desk

3D graphics have also been used extensively for abstract visualizations of the document space. A common representation method is a 3D graph with the documents as nodes and their relations or similarity as a defining factor for the edges, such as Starwalker [22]. A special category of 3D graphs is the 3D trees Cone Trees [19], with their nodes arranged in cones in order to take advantage of the third dimension. The Cat-A-Cone [30] technique combines a Cone Tree with the Web Forager Environment. The Hyperbolic Tree [24] visualization places the Cone Tree in the 3D hyperbolic space in order to achieve greater information density.

Another group of visualization methods, to which belongs among others PRISE [23], uses 3D coordinate systems to place documents in the 3D space according to preselected attributes. Other methods, such as Information Cube [25] and Information Pyramids [15] use nested 3D geometrical objects in order to represent document hierarchies. Information Sphere [25] applies the Perspective Wall focus and context technique on the surface of a sphere.

PROPOSED APPROACH

All the visualization methods presented above help the user get an overview and browse the space of the existing information, assuming that certain characteristics and attributes such as the author, keywords, and even the text of the documents are known. However, this is not usually the case with historical archives. In most cases historical archives are grouped in broad categories that only provide hints of the document's contents. To this end, meta-data have to be introduced that will aid users in their search and guide them in finding the documents in question.

Analysis of the methods employed by domain experts at the historical archive of the University of Athens, revealed that when faced with a user's requirement for information that cannot be directly answered through the existing categorization, domain experts tend to use their own knowledge of the entities found in the domain and their interrelations. This abstract representation of the domain resembles the structure of an ontology.

Ontologies have been used for knowledge representation in various fields because they offer a higher level of abstraction and relationships than a conventional vocabulary or thesaurus [29]. They provide a versatile and flexible representation that may aid the sharing of the structure of information among people and software agents.

There exist two main user groups for a digital historical archive system. The first group is that of the domain experts and the second is that of the casual user and researcher. The user interface has to offer its functionality in a manner that caters for the needs of both groups. Analysis has shown that while domain experts may start their search directly at the categorization of the historical archive with good results, casual users can seldom achieve the same performance. Casual users need first to get an overview of the domain space, then zoom in and filter out.

As a consequence, it would be very useful to study in depth the conceptual model of the historical archives user, both researcher and personnel, on which their searches are based. The definition of this model, which constitutes in fact the ontology of the historical archive, could result to the development of alternative browsing methods for the environment of the digitized historical archives. The ontology of the historical archive, visualized effectively could be a valuable tool not only for browsing but for exchanging information with other users as well.

In the proposed approach the ontology tool would be an integral part of a historical archive digital visualization system. Various theories have been developed suggesting basic features and facilities a visualization system should have in order to be usable and effective. In [34] Shneiderman presents four high level tasks that an information visualization application should support:

Overview: Gain an overview of the entire collection. Overview strategies include zoomed out views of each data type to see the entire collection plus an adjoining detail view (context plus focus display).

Zoom: Zoom in on items of interest. Users typically have an interest in some portion of a collection, and they need tools to enable them to control the zoom focus and the zoom factor.

Filter: filter out uninteresting items. By allowing users to control the contents of the display, users can quickly focus on their interests by eliminating unwanted items.

Details-on-demand: Select an item or group and get details when needed. Once a collection has been trimmed to a few dozen items it should be easy to browse the details about the group or individual items.

Therefore, adopting a holistic view of the historic archive information retrieval system, the user should be able to acquire an overview of the information space, then zoom-in on specific areas of interest, narrow this to a specific part of the document space and view details of the documents. Starting with the domain ontology the user should be able to select certain nodes of the ontology, then move through the existing categorization to view specific documents. To further, illustrate this approach necessary definitions should be given.

Ontology Definition

According to [26], an ontology is an explicit specification of a conceptualization. The term "conceptualization" is defined as an abstract, simplified view of the world that needs to be represented for some purpose. It contains the objects, concepts and other entities that are presumed to exist in some area of interest and the relations that hold them. The term "ontology" is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems what "exists" is exactly that which can be represented.

Therefore, as defined in [28], an ontology is a formal explicit description of concepts in a domain of discourse (called classes or entities), properties of each concept describing various features and attributes of the concept (called slots, roles or properties), and restrictions on slots (called facets or role descriptions). An ontology together with a set of individual instances of classes constitutes a knowledge base.

A more mathematical definition can be the following [27]:

An ontology is a triple O = (C, R, isa) defined as follows:

1. $C = \{c_1, c_2, \ldots, c_n\}$ is a set of concepts, where each concept c_i refers to a set of real world objects (concept instances),
2. $R = \{r_1, r_2, \ldots, r_m\}$ is a set of binary typed roles between concepts.
3. *isa* is a set of inheritance relationships defined between concepts. Inheritance relationships carry subset semantics and define a partial order over concepts.

Ontologies can be represented as directed graphs where nodes correspond to concepts and links to *roles* and *isa* relationships.

Ontology – Categorization binding

The ontology to be used in the tool is comprised of general concepts, their inheritance relations (*isa*), allowing for multiple inheritance, their role relations and the instances (*io*). The categorization of the Historical Archive is a hierarchy having the form of a tree.

Once an ontology describing the domain has been defined, it has to be linked to the existing categorization of the domain. This is not a trivial task since the ontology is not necessarily a genuine superset of the existing categorization, resulting in nodes that do not correspond directly to an existing category. Furthermore, nodes in the ontology may correspond to more than one categories and one category to more than one node creating a complex visualization problem.

A weight may be assigned to each link in order to represent the degree of relevance of the specific entity to the corresponding category. This link may be computed taking into account a number of factors, such as the subjective certainty of the user that a category contains material relevant to the entity, the number of relevant documents found during an automated search,

the possible concurrence of category and entity name, etc.

The ontology of the historical archive will be presented to the user, which, in this case, will be the personnel of the archive, in an integrated environment for its management. The user in a semi-automated way will create the initial connections between the ontology entities and the corresponding categories. After this initial step, the users, who in this case could be either the personnel or the researchers, will be able to use the ontology in order to navigate in the document space while browsing, or locate specific documents.

To this end, an algorithm was developed that defines and clarifies the necessary steps for the linking of the institution's ontology to the corresponding historical archive and is presented below.

The algorithm developed contains a number of steps that in most cases can be automated so as to enable the system to carry out the linking without the need for user intervention. These steps are outlined below:

1. For each entity the system checks the existing categorization for category labels that also appear in the ontology and sets a link between them. Each link is considered to be *under final approval* by the historical archive domain expert. An initial relevance weight is set during this stage, as the similarity of the names implies a strong degree of relevance. The system can employ a thesaurus with synonyms in order to identify links more effectively.
2. The system performs the reverse process: for each category label it searches in the ontology for corresponding entities and sets links between them. In this case also an initial weight is set and a thesaurus with synonyms may be used.
3. If full-text is available for a portion of the category documents, this can used to help improve the relation of the links (and the corresponding weight) by first analyzing the documents. If documents with a reference to the specific entity are found inside a category, then this category may be linked with the entity and the weight of the link is set according to the number of documents found relevant.
4. The final stage of this procedure is the validation and approval of the links by the domain expert user and the identification of links that were not detected automatically. The user first checks the links that the system detected in order to approve or reject them and adjust their weights.
5. The next step is to add links that were not identified but the domain expert user considers useful. To this end, the user employs his/her personal experience in satisfying requests for specific documents by searching the archives through the categorization or trying to locate documents relevant to a more general request by browsing the categorization. This conceptual model of the archive is of the essence during this linking process as the added value resulting from the expert's intuition may aid in the creation of a useful retrieval tool for the end users.

The result of this procedure is a representation of the user's conceptual model of the historical archives structure and method to locate in the categorization documents relevant to a particular entity.

Document Retrieval

When the ontology is constructed and a sufficient number of links to the category have been created by the archive personnel, the tool can be made available to the end users, in this case the historians and researchers in order to conduct searches in the archive material. The first step of the search process would be for the user to identify one or more entities relevant to his current search in the archive ontology. This process could be done manually or with the aid of an agent designed for this purpose. For example the user can type in a question of interest and the agent will identify from the question keywords relating to certain ontology nodes. Then, following the links between the ontology and the categorization, the user will be guided to the corresponding categories, the contents of which he/she may browse to locate the documents he/she is looking for.

Update of the Ontology – Categorization Links

During the retrieval process the system will monitor the user movements in order to identify new successful links between entities and categories.

After each search the user can be prompted to state if the search was successful and what category produced results for a specific entity. With this information the system will create a new link between the entity and the category or categories that produced relevant documents. These will later be examined and approved or rejected by the archive personnel. On the other hand, if a category to which a link exists did not produce any relevant results, it should also be examined by the personnel to verify that it actually contains documents relevant to the specific entity.

The new links created after a successful search and subjected to validation by the archive personnel could be used by future users to access more quickly and effortlessly the same material. As a result, a dynamic visualization tool is created, which grows with each usage and memorizes successful searches in order to be later used by other members of the researcher community.

CASE STUDY: UNIVERSITY OF ATHENS HISTORICAL ARCHIVE

Ontology and Categorization Binding

In Figure 1, we present as an example a part of the National and Capodistrian University of Athens Historical Archives ontology. This visualization was

created using the Protégé [32] ontology tool and OntoViz visualization technique. The inheritance relations are represented with an arrow labeled "*isa*" and with direction from the child to the parent entity. The instances of the various entities are represented with a similar link labeled "*io*". The role relations between classes have been added as dashed lines

In Figure 2, a part of the categorization is presented with the ontology with some of their interconnecting links.

These links were created as follows:

1. The system first searches for categories and entities with the same name in order to create the corresponding links. The entities Faculty, Philology, Physics and Informatics are linked to categories at this stage. The weight assigned is **strong** in this case.
2. By analyzing some documents available in full text in the Faculty category concludes that they are relevant to the Rector entity and creates the corresponding link.
3. The domain expert user checks and verifies these links and adds two more, the ones between Bambiniotis and the Philology and Student Education categories respectively. Based on his/her previous knowledge that Bambiniotis was a student in the University and now belongs to the department of Philology considers that there is a strong possibility that these two categories contain relevant documents.

Document Retrieval

As an example of the document retrieval procedure one can consider the following: Using again the example of the National and Capodistrian University of Athens Historical Archives, we suppose that a researher, is looking for documents relevant to a dean, named Bambiniotis. There are two possibilities:

1. Bambiniotis exists as an entity in the ontology, as an instance of the Rector entity (Figure 2). In this case, the user is guided through direct links to the categories Philology and Student Education. These two links have different weights, as it is expressed from the thickness of the link. This gives a hint to the user as to what category should investigate first. In the respective categories the user may search for relevant documents.
2. Bambiniotis does not exist as an entity. As a result the researcher has to search in the ontology for other, relevant entities, for example Professor or Rector and through these he can be guided to categories that contain documents possibly relevant to his/her goal. In our case there is a link form the Rector entity to the Faculties category. If the search leads to satisfactory results, then the successful connections between ontology and the categorization can be stored automatically in order to be used in future searches. A new instance named "Bambiniotis" could be added to the ontology, in this case, as a sub-entity to the Rector or Professor entities.

Ontology Visualization

There are not many electronic historical archive systems that are operational and provide access to digitized documents through the web. Two of them are NARA National Archives & Records Administration [3] and Public Record Office, The National Archives [4]. The common characteristic of their user interface is that they have the functionality of a Web browser. The user can search for documents using keywords or a simple hierarchy and the results of the search are displayed in a list that contains the title and icon of the document. They do not include any type of 2D or 3D visualization of the search results that could offer the user an overall view of the results of a query and they use documents and for the presented documents much information is available about their contents and type.

Given the ontology of the contents of an historical archive and its interconnection with the existing historical archive categorization, our goal is the development of a visualization tool that will be used for browsing the archive structure through the ontology, presenting both interconnected structures with an adequate and effective visualization.

In the field of ontology visualization, not much work has been published, especially in 3D. In [30], a 2D hyperbolic tree is used for the ontology visualization. Jambalaya [31], a visualization plug-in for the Protégé ontology tool [32] uses the ShriMP (Simple Hierarchical Multi-Perspective) 2D visualization technique. ShriMP uses a nested graph view and the concept of nested interchangeable views, combined with geometric, fisheye and semantic zooming. The OntoViz [32] is another Protégé visualization plug-in using a very simple 2D graph visualization method. ToughGraph [33] is another visualization used in Protégé, which presents a graph with nodes that may be expanded or retracted using animation during their repositioning.

CONCLUSIONS-FUTURE WORK

In this paper a methodology for information retrieval in historical archives has been presented. The methodology caters for cases where the existing categorization is very abstract and not enough information about the content of each individual document is present, as it is usual the case with historical archives where only an image of the document is available and not the full text. To this end, the methodology proposes the creation of an ontology and its linking with the existing categorization. Research so far has focused on the problem of creating successful links. Future research will focus on the problem of visualization of the ontology. 3D technologies offer a prominent solution since they enable the presentation of large amounts of information in a restricted space as it is the one of the monitor screen.

REFERENCES

[1] Compus, Visualization and Analysis of Structured Documents For Understanding Social Life in the 16th Century, Jean-Daniel Fekete, Nicole Dufournaud

[2] Primarily History: Historians and the Search for Primary Source Materials
Helen R. Tibbo, ACM 2002

[3] NARA, U.S. National Archives and records administration, http://www.archives.org

[4] Public Record Office, The National Archives, http://www.pro.gov.uk/

[5] Visualizing Digital Library Search Results with Categorical and Hierarchical Axes, Ben Shneiderman, David Feldman, Anne Rose

[6] IVEE: An Information Visualization & Exploration Environment, Christopher Ahlberg, Erik Wistrand, Proceedings of IEEE Viz 1995

[7] A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies, John Lomping, Ramana Rao, Peter Pirolli, Proceedings of CHI. 1995

[8] Tree visualization with Tree-maps. A 2-d space-filling approach, Ben Shneiderman

[9] Information Slices: Visualizing and Exploring Large Hierarchies using Cascading, Semi-Circular Discs, Keith Andrews, Helmut Heidegger

[10] A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable Interfaces, Walky Rivadeneira, Benjamin B. Bederson

[11] *TileBars: Visualization of Term Distribution Information in Full Text Information Access,* Marti A. Hearst, 1995

[12] A Scrollbar-based Visualization for Document Navigation, Donald Byrd, Proceedings of Digital Libraries, 1999

[13] The Visualization of Large Hierarchical Document Spaces with Infosky, Keith Andrews, CODATA Prague Workshop, Information Visualization, Presentation and Design, 30-31 March 2004

[14] WEBSOM – Self-Organizing Maps of Document Collections, Timo Honkela, Samuel Kaski, Krista Lagus, Teuvo Kohonen, Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6

[15] Information Pyramids, Compactly Visualizing Large Hierarchies, Josef Wolte

[16] Data Mountain: Using Spatial Memory for Document Management, Robertson G., Czerwinski M., Larson K., Robbins D., Thiel D., van Dantzich M, Proceedings of ACM UIST '98 Symposium on User Interface Software&Technology, San Fransisco, CA, November 1998

[17] The Perspective Wall: Detail and Context Smoothly Integrated, Jock D. Macinlay, George G. Robertson and Stuart K. Card

[18] The Task Gallery: A 3D Window Manager, George Robertson, Maarten van Dantzich, Daniel Robbins, Mary Czerwinski, Ken Hinckley, Kirsten Risden, David Thiel, Vadim Gorokhovsky, CHI Letters, volume 2, issue 1, 494-501.

[19] Information Visualization Using 3d Interactive Animation, George G. Robertson, Communications of the ACM, Vol 36, No 4, pp 57-71Stuart K. Card, John D. Mackinlay, April 1993

[20] Cat-a-Cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy, Marti A. Hearst, Chandu Karadi, Proceedings of 20th Annual International ACM/SIGIR Conference, Philadelphia, PA, July 1997

[21] The WebBook and the Web Forager: An Information Workspace for the World-Wide Web, Stuart K. Card, George G. Robertson, and William York

[22] Representing the Semantics of Virtual Spaces, Chaomei Chen, Linda Thomas, Janet Cole, Chiladda Chennawasin, IEEE MultiMedia, 1999

[23] Interactive 3D Visualization for Document Retrieval, John Cugini, Christine Piatko, Sharon Laskowski, Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation , ACM Conference on Information and Knowledge Management (CIKM '96), November 1996

[24] H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space, Tamara Munzner

[25] Three Dimensional Information Visualization, Peter Young, November 1, 1996, Computer Science Technical Report:12/96

[26] A Translation Approach to Portable Ontology Specifications, Thomas R. Gruber, Knowledge Acquisition, Special issue: Current issues in knowledge modelling, Vol 5, Issue 2 (June 1993, Pages: 199 - 220

[27] Integrating Ontologies and Thesauri to Build RDF Schemas, B. Amann, I. Fundulaki, Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries, pp 234 – 253, 1999

[28] Ontology Development 101: A Guide to Creating Your First Ontology, Natalya F. noy, Deborah L. McGuiness, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001

[29] Converting a controlled vocabulary into an ontology: the case of GEM, Jian Qin, Paling, Stephen, Information Research, 6(2) Available at: http://InformationR.net/ir/6-2/paper94.html

[30] Visualization of Ontologies through Hypertrees, Kleber X. S. de Souza, Andriana D. dos Santos, Silvio R. M. Evangeista

[31] Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protégé, Margaret-Anne Storey, Mark Musen, John Silva, Casey Best, Neil Ernst, Ray Fergerson, Natasha Noy, http://www.thechiselgroup.org/jambalaya

[32] Protégé project, Stanford University, http://protege.stanford.edu.

[33] TouchGraph,http://www.touchgraph.com/

[34]Shneiderman, B., Creating creativity: user interfaces for supporting innovation, ACM Transactions on Computer-Human Interaction Volume 7, Issue 1 (March 2000) pp. 114 - 138

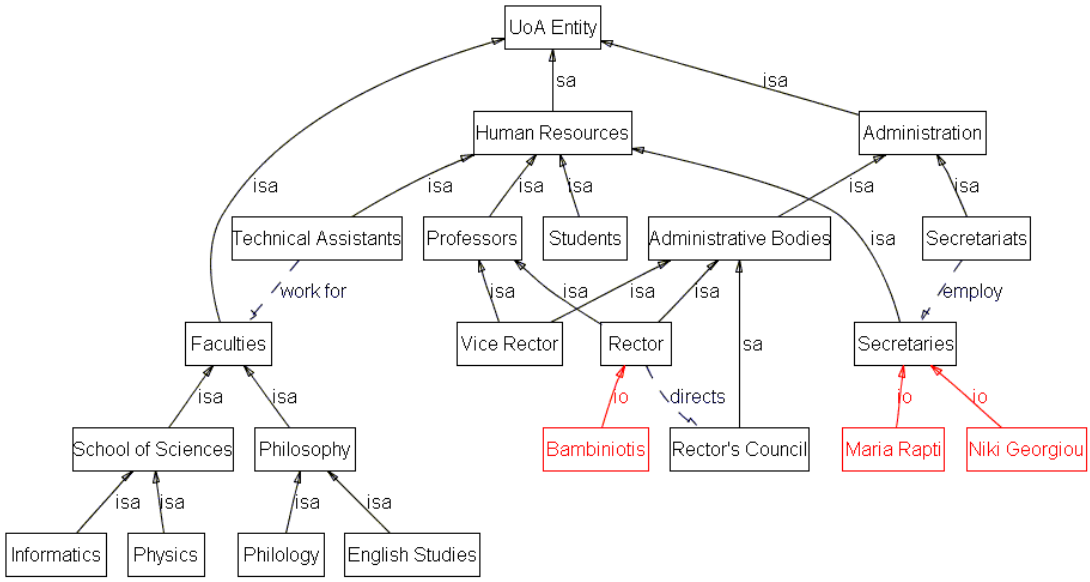

Figure 1. The UoA ontology with the inheritance role relations visible.

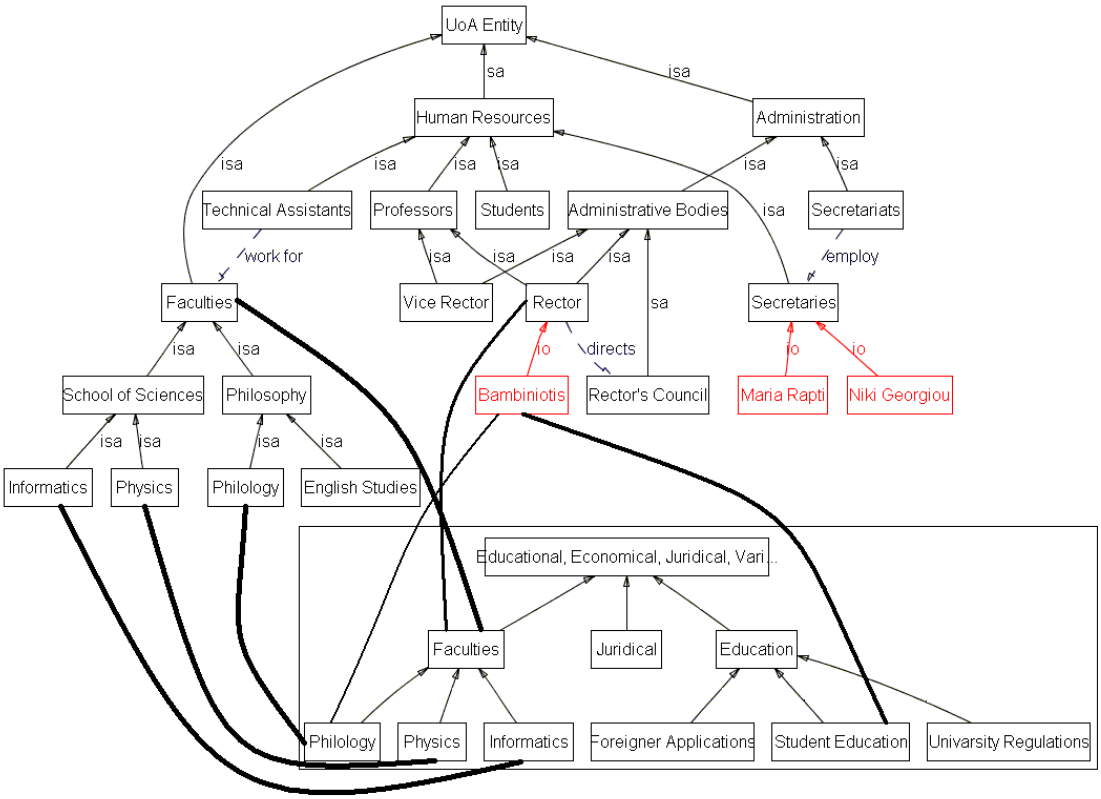

Figure 2. Part of the UoA ontology with part of the Historical Archives Categorization visible at the lower right part of the figure.