Visualizing OPAC Subject headings

Ioannis Papadakis, Michalis Stefanidakis, Aikaterini Tzali

Ioannis Papadakis graduated from the Department of Informatics, University of Piraeus in 1997. He obtained his PhD at the same institution, and delivered the dissertation "Digital libraries: Architectures, Security and Information Retrieval" in 2003. Since 2004, he is a Lecturer at the Department of Archival and Library Sciences, at the Ionian University.

Michalis Stefanidakis holds a Diploma of Computer Engineering and Informatics and a Ph.D. in Design and Performance Evaluation of Distributed Memory Parallel Computer Architectures (2000) from the University of Patras. Currently he is occupied as a lecturer at the Ionian University, Department of Informatics, with research interests including High-Performance Distributed Computer Systems, Real-time Embedded System Architectures and Pervasive Computing.

Aikaterini Tzali graduated from the Department of Archival and Library Sciences, Ionian University in 2002. She obtained the MSc in "Information science - Information services in digital environments" at the same institution in 2005. Since 2003, she works as a Librarian at the academic library of Department of Informatics, Ionian University.

Abstract

Purpose – The proposed paper presents kamij, a prototype web-based application, which aims at providing a robust, user-friendly and at the same time efficient information retrieval process in an online library catalogue that is based on semantic information encapsulated within Subject headings. Instead of relying on a static, predefined hierarchical navigation structure, which forces users to follow certain paths, the proposed application relies on a dynamic, interactive graph-based structure. Subject headings and their relationships constitute the nodes and arcs respectively of the graph. Such a structure may be encoded in a variety of semantic-aware technologies implementing subject-based classification such as thesauri, taxonomies, ontologies etc. The ultimate
The goal of Kamij is to guide users in locating useful information as accurately and efficiently as possible.

**Design/methodology/approach** – The building block of the proposed approach is the employment of ontologies. Specifically, users select subject headings by interacting with the prototype application which is based on an AJAX-powered web Graphical User Interface (GUI). The GUI communicates with an underlying ontology which is comprised of the subject headings and their interrelations. Such headings are used within the OPAC to semantically describe the contained assets (books, journals, etc). By selecting a specific subject heading, the proposed application addresses a typical search query to the OPAC containing the selected heading. A working functional application embedded within a real-world OPAC will be provided illustrating the effectiveness of the proposed approach.

**Findings** – According to the proposed approach, a GUI interface exposes the hierarchy of the subject headings employed within an OPAC, as well as all stated relations between such headings (e.g. “seeAlso”), as links that the user can follow, effectively traversing the ontology and formulating at the same time the actual query to the underlying OPAC. This act of interactive navigation through the library’s assets aids searchers in accurately formulating their queries, by offering broader or narrower concepts for selection or indicating alternative or related concepts they might be initially unaware of. The augmented exposition of inter-relations between concepts provides multiple paths for information retrieval and enables searchers to fulfill faster, more efficiently and in an intuitive manner their information needs.

**Practical implications** – The paper includes implications for the development of modern, semantic web applications focused on the library domain. The novel approach of visualizing subject headings could be further extended to visualize a number of other conceptualizations of the library domain (e.g. author-based classifications).

**Originality/value** – This paper fulfils an identified need to take advantage of the “hidden knowledge” existing within the library domain but for a number of reasons is never exposed to the library users.
1 Introduction

Nowadays, libraries across the world employ Online Public Access Catalog (OPAC) systems in order to facilitate quick access to their content [1]. Such content refers to a registry of information entities considered as library material such as books, computer files, magazines, etc. Many times, information entities within a library are well-described in various ways through the employment of semantic-enabled technologies such as thesauri, controlled vocabularies etc. As far as information retrieval is concerned, however, most of the times OPACs do not take advantage of the expressiveness of information deriving from such technologies and remain antiquated browsing interfaces relying on static structures with crude search tools useful only for locating specified materials [2]. Thus, the Graphical User Interface (GUI) seems to prevent users from exploiting the functionality deriving from the employment of such technologies. The proposed paper aims at providing a robust, user-friendly and at the same time efficient navigation procedure in an online library catalogue that is based on semantic information. Instead of relying on a static, predefined hierarchical navigation structure, which forces users to follow certain paths, the proposed navigation procedure relies on a dynamic, interactive graph-based structure. Such a structure may be encoded in a variety of semantic-aware technologies implementing subject-based classification [3] such as thesauri, taxonomies, ontologies etc. The ultimate goal of the navigation procedure is to guide users in locating useful information as accurately and efficiently as possible. In order to demonstrate the functionality of the proposed navigation procedure, kamij has been developed, which refers to an ontology-based, information retrieval application integrated within a typical OPAC.

2 The origin of Semantics

Semantics have been in the neighborhood for quite some time now, long before the emergence of library-related tools such as thesauri or
ontologies. The Knowledge Management (KM) community has produced considerable work aiming at instructing machines on how to comprehend the semantics of information that they possess, in order to develop software capable of managing knowledge instead of plain information.

Despite the initial promising results, however, things did not evolve the way most members of the KM community anticipated. Mapping human behavior and wisdom proved to be particularly complex and subjective among KM researchers. Consequently, mostly monolithic systems and theories have been developed, all featuring certain advantages but also drawbacks when compared against each other. The point is that the absence of collaboration between the members of this community and the inherent heterogeneity of the resulting products, prevent the wider dissemination and establishment of very significant findings.

Such diversity is inherited to the library domain where semantic-aware technologies (e.g. thesauri, controlled vocabularies, etc) do not seem to be based on well-accepted standards. More specifically, the existence of standards for defining the general principles upon which such technologies should be based (ISO 2788:1986, ISO 5964:1985), does not seem to be enough to ensure semantic interoperability between applications and between applications and their users. There is limited interoperability between tools and digital resources employing different thesauri [4]. The lack of standardization in a more technical level is perhaps the most important reason why such technologies despite their vast capabilities in expressing information, so far do not seem to be popular among the members of the library community.

The advent of the semantic web [5] could be considered as a movement towards the right direction as far as the library domain is concerned. The diversity, bulkiness and chaotic nature of the web forced the web community to rely on standards for the foundation of this environment. In this context, the semantic web provides the necessary technological infrastructure for the development of semantic-aware library services. The next section describes an interactive information retrieval process applied to an online library catalogue based on semantic information.

The proposed approach is presented through kamij, a web-based application following the most recent trends of the semantic web such as AJAX technology and the web ontology language – OWL for encoding semantics.

3 Kamij: implementing an ontology-based, information retrieval process for the library domain

Kamij is a web-based, information retrieval application targeted to the library domain. Specifically, kamij interacts with the online library catalogue located at the library of the department of computer science, at the University of Ionio in Greece. Kamij aims at providing accurate navigation and thus efficient information retrieval over the underlying online library catalogue. This is accomplished by exposing existing semantic information in an intuitive and at the same time user-friendly fashion.

The building block of the proposed application is the employment of ontologies. Specifically, as illustrated in fig. 1, users interact with the online library catalogue through the prototype application which is based on an AJAX-powered web GUI. The online library catalogue is managed by a software module named “semantic catalogue manager,” capable of integrating semantic information obtained through ontologies with existing information provided by the various bibliographic databases that constitute the deriving online library catalogue.

![Diagram](image.png)

Figure 1: Architecture of the prototype application.

Based on the aforementioned architecture, kamij is able to provide interactive, guided navigation over the online library catalogue based on semantic criteria. It is also important to notice that users are in
control of the navigation procedure at all times. The application’s main components and their corresponding functionality are explained in detail in the following sections.

3.1 Ontology

The underlying ontology is capable of modelling the various concepts existing within the library domain together with their relationships. More specifically, kamij interacts with an ontology modelling the subject headings that are employed to provide the subject to the resources participating within the online library catalogue. In this context, four relations have been identified. The first two refer to the standard hierarchical relations existing within every hierarchy, namely “contains” and “is part of”. “Contains” points to concepts with narrower meaning than the selected concept (e.g. Mathematics contains Algebra) and “is part of” points to concepts with broader meaning than the selected concept (Algebra is part of Mathematics). There are also two relations, namely “seeAlso” and “inContextOf”. The former refers to concepts that have similar yet not the same meaning with the selected concept and the latter refers to concepts that determine the precise meaning of the selected concept.

The ontology may also contain multilingual, non-standard and user-preferred terms, which are used in conjunction with the “official” subject headings within the library catalogue.

3.2 Bibliographic Database

Each library consists of one or more specialized databases containing information about a specific subject. Each database is essentially an electronic catalogue to journal or magazine articles, containing citations, abstracts and often either the full text of the articles indexed, or links to the full text or simple bibliographic information to such resources.

The proposed application provides keyword terms as entry points to the information retrieval service of the underlying bibliographic database. The selection of the terms is based on semantic criteria. Specifically, users traverse the semantic structure visualized by the proposed GUI
This navigation process results in a set of semantic-flavoured keyword terms that correspond to the users’ information needs. Such terms are then transparently fed into the information retrieval service through the “semantic catalogue manager” that will be described in the following section.

3.3 The semantic catalogue manager

The semantic catalogue manager provides the core functionality of the prototype application. Such a module acts as a mediator between the underlying helper ontologies and the AJAX-based web GUI. Specifically, it is the manager’s responsibility to populate the GUI with information deriving from the helper ontologies in an appropriate fashion. Moreover, the semantic catalogue manager receives user requests from the GUI and transforms such requests into semantically rich keyword terms. Such terms will be directed to the appropriate database’s search facility. Eventually, the user will be presented with search results as provided by the chosen search facilities.

From the user’s point of view, the semantic catalogue manager aids users in transparently selecting the most appropriate keyword terms for their quest of knowledge within each library’s content. Such a task is accomplished through the employment of semantically rich structures (i.e. ontologies) together with traditional online library catalogue’s search facilities.

3.4 AJAX-powered web GUI

The web interface empowering the semantic navigation within the online library catalogue employs a number of visual widgets:

a) a ClassBox (an example is shown in fig. 2), representing a thematic subject heading (corresponding to a class within the helper ontology). Such a widget contains the relationships in which the subject heading is involved (e.g. narrower meaning, broader meaning, etc). The user, via the familiar hyperlink interface, may select a relationship and observe the subject headings (corresponding to classes within the helper ontology), that are linked with the initial subject heading through the selected relation. This way, user navigation can be performed all the way down to the bottom of the hierarchy. Subject headings may also list the alternative (non-preferred and multilingual) equivalences of a
term, facilitating this way user–oriented navigation. It should also be noted that wherever available, translations of the subject headings appear beneath the authorized headings in a different color.

Figure 2: A ClassBox example.

b) a ContextMenu (an example is shown in fig. 3), which lists the subject headings that are related through the selected relation to the initial subject heading represented as a ClassBox. The ContextMenu is activated by clicking on a relation within a ClassBox. Such relations exist within the ontology as “properties” linking two different concepts.

Figure 3: A ContextMenu example.

The ability to navigate according to the relations within the ContextMenu instead of a mere hierarchical path, is the true strength of our proposed interface.

Figure 4: A LinkLine example.

c) An interconnecting LinkLine (see fig. 4) together with an appropriate description of the relation, employed to denote the relation between two
adjacent ClassBoxes. The LinkLine conveys such information via its label and positioning, depending on the selection of the user.
3.5 Implementation details

Kamij is a web-based application hosted at the library of the Ionian University, department of Informatics, available at http://195.251.111.53/server/entry/index.html. The underlying ontology refers to a part of the well-known Library of Congress authorized subject headings collection, available at: http://authorities.loc.gov/. It is encoded in OWL format in conjunction with the rdf and rdfs namespace. The “semantic catalogue manager” component is implemented in Python 2.4 and the GUI is based on javascript employing various XMLHttpRequests. Finally, Greek translations of the employed subject headings are available.

4 Conclusions

This paper introduces an innovative navigation procedure applicable to online library catalogues – OPACs. According to the proposed approach, users are able to semantically navigate the digital assets within a library collection based on a dynamic, interactive graph-based structure. The graph-based structure is essentially an ontology that models the subject headings and the corresponding relations of the collection’s digital assets.

Users are able to semantically explore the underlying collection without having to know the exact subject headings that are being employed to describe the individual assets. Additionally, the bilingual interface assists users that are unaware of the English terminology that is usually employed to express the original subject headings (i.e. LSCHs). Moreover, the proposed application can be easily integrated to any existing OPAC system. The only requirement is that the subject headings’ ontology should be synchronized with the actual subject headings employed within the OPAC. Finally, the employment of most recent advances in web development renders the proposed application as very appealing to average library users.

References