Semantic Navigation in an online Library Catalogue

Ioannis Papadakis¹, Michalis Stefanidakis², and Aikaterini Tzali²

¹ Ionian University, Department of Archives and Library Sciences, Plateia Eleytherias, 49100 Corfu – Greece
² Ionian University, Department of Computer Science, Plateia Eleytherias, 49100 Corfu – Greece

{papadakis,mistral,ktzali}@ionio.gr

Abstract. In this paper, an innovative ontology-based, information seeking process is proposed, capable of aiding searchers in their quest to fulfill their information needs. Such a process supports multilinguality and concept disambiguation in a way that prevents overcrowding the screen while exposing the expressiveness of the underlying ontology. In order to assess the effectiveness of the proposed approach in a real-world scenario, a prototype web-based application is developed by applying the approach to a working Online Public Access Catalog – OPAC located at the Ionian University. The underlying ontology is encoded in OWL format. Interaction is facilitated through an intuitive, web-based GUI implemented in AJAX technology.

Keywords: semantic web, owl, OPAC, ontologies, information retrieval

1 Introduction

Information exploration on the web is always an open research agenda due to the diversity of the collections it contains. Such collections refer to various domains. Depending on the domain a collection refers to (e.g. domain: computer programming, collection: java), quite often a specific (formal or informal) terminology is employed by the members of the domain. However, such terminology is usually a ‘foreign language’ to people outside this domain that nevertheless wish to retrieve information from such collections.

This paper introduces a novel ontology-based information seeking process for medium to large collections on the web. Serving as an application of the semantic web [1], the proposed process is capable of bridging the gap between users’ information needs (elsewhere defined as information goals or information retrieval context [2]) and the terminology that is usually employed to describe and index information assets within such collections.

More specifically, instead of relying on a typical static, predefined hierarchical navigation structure, which forces users to follow static, predefined paths, the proposed information seeking process relies on a dynamic, interactive graph-based structure. Such a structure visualizes the semantic structure of an underlying ontology
The ultimate goal of the proposed work is to guide users in locating useful information as accurately and efficiently as possible.

The proposed process is capable of being integrated to common search interfaces as a query formulation tool. Users are able to express their information needs in search queries that are closer to the indexed terms of the underlying collection.

A prototype system has been developed in order to serve as a proof of concept for the proposed information seeking process. It combines interactive, user-centered searching and browsing in a rather intuitive manner. The prototype is integrated to the Online Public Access Catalog – OPAC of the Library of Ionian University.

The rest of this paper is structured as follows: The next section summarizes similar projects aiming at providing information seeking alternatives on the web based on ontologies. Section three defines the proposed ontology-based, information seeking process. This section contains three subsections referring to a prototype web-enabled application that is based on the principles of the proposed information seeking process. More specifically, the prototype’s architecture is presented, along with the GUI’s functionality and the underlying ontology’s structure. Finally, this paper draws conclusions and points directions for future work.

2 Ontology-based information seeking on the web

Ontology-powered information exploration is one of the most promising approaches to enhance existing search GUIs with features capable of enabling users to better express their information needs or to improve exploratory search styles [2]. The expressiveness of information within ontologies should find its way towards the GUI while at the same time the GUI should retain basic usability features [5] that constitute the whole information retrieval system appealing to its users. It should be also kept in mind, that GUIs based on ontologies should not only expose the concepts they contain. It is equally important to take advantage of the relations between concepts (i.e. properties-slots) in order to provide a more efficient interactive search tactic.

Several efforts have been made to employ ontologies within a broader information seeking system (for more information, see [6]). Among them, OntoQuery [7] employs a search interface where searchers are prompted to type terms that are fed into a Natural Language Processing – NLP system. Despite the promising results in certain occasions, such an approach can only be applied to a very strict number of domains were natural language is unambiguous.

Another interesting approach, OntoIR [8], provides a selection-based interface according to which searchers have the ability to engage themselves into an interactive process that ultimately leads to query refinement, based on the concepts of the underlying ontology. OntoIR manages to avoid exposing the complexity of the underlying ontology to the screen. Moreover, it takes advantage of the richness of information within the ontology to provide a rather efficient information exploration process to the searcher. Similarly to other information seeking efforts, direct access to
the underlying repository is required, since the contained information assets should be annotated with the ontology.

In the following section, an ontology-based information seeking process is presented, aiming at aiding searchers expressing their information needs within the context of the formal terminology of the referring domain, as also suggested by Saracevic et al. in [9]. The proposed work combines searching and browsing within the context of interactivity between the system and the searcher.

3 Ontology-based information seeking process

The proposed approach aims at enabling searchers expressing the semantics of their information needs in an unambiguous fashion. This is achieved by providing certain search tactics that ultimately lead to the satisfaction of the searchers’ information needs. According to Bates [10], tactics are defined as “one or a affordance handful of moves made to further a search” (e.g. query formulation, query refinement).

According to the proposed model, searchers take control over the search process through an interactive GUI. Interaction is generally appealing to searchers on the condition that the GUI provides straightforward interactions, as stated in [8].

The proposed GUI is comprised of boxes that correspond to concepts and labelled lines corresponding to the relations linking such concepts. Searchers may hop from one concept to a related one by selecting one of the various relations each concept is being involved. Such relations are outlined inside each box. Upon selection of a relation, a pop-up menu is presented containing the concept(s) that relate to the initial concept with the selected relation. Finally, by selecting an item within the pop-up menu, a new box is created together with a line that connects the two boxes. The whole iterative process is illustrated in Fig. 1.

As the searcher hops from concept to concept, hers/his information needs become formal terminology terms constituting the search query that will be ultimately addressed to the underlying search engine.
The aforementioned concepts and relations correspond to the classes and properties (or slots) of an ontology that manages to capture the conceptualization of the referring domain. It should be mentioned that the output of the proposed information-seeking process is essentially one or more concepts that may serve as query terms to the traditional search engine interface. Since the ‘glue’ between the information assets residing at the repository and the provided GUI are the concepts-terms, it is very important to design the ontology in a way that accurately corresponds to the terms being indexed by the search engine.

The proposed approach would have maximum effectiveness in the case of employing the ontology to annotate the underlying information assets. Although such a scenario would lead to much more relevant results and consequently greater satisfaction to the searcher, it is the authors’ belief that serving as a complement to pre-existing search engines, the ontology-based information seeking process can be successfully applied to a wider range of real-world applications.

The proposed approach is capable of supporting many (if not all) of the search tactics proposed by Bates [10] provided that the corresponding relations are accordingly modelled in the ontology.

It is important to note that the GUI exposes the expressiveness of the ontology without cluttering the searcher’s screen. This is accomplished by enabling the searcher to select which part of the ontology will be revealed to her/him. The searcher can expand and/or collapse the ontology through an interactive process at hers/his own will.

In order to narrow the interactions with the searcher and thus the overall time performance, it is very important to provide efficient entrypoints [8] to the system. Entrypoints are the concepts that the searcher confronts when the information seeking process is initiated. The closer the entrypoints are to the resulting concepts, the better for the searcher.

As stated in [7], the GUI of any ontology-based information seeking model should be web-browser based. In this context, the proposed approach should take advantage of semantic web standards for the encoding of the ontologies (e.g. Web Ontology Language – OWL) and new trends in web GUI design based on the most promising Ajax technology.

In the following section, a prototype system is presented based on the line of thoughts described throughout this section.

### 3.1 Prototype system Architecture

In order to test the proposed approach at the field, the OPAC of Ionian University was selected. The selected OPAC provides rich descriptions of the underlying information assets based on the widely employed Library of Congress Subject Headings – LCSHs (available at: http://authorities.loc.gov/). Thus, each information asset corresponds to one to three subject headings corresponding to the formal LCSH.

Searchers are prompted to make use of the provided ontology-based prototype (fig. 2) in order to find the subject heading(s) that best describe(s) their information needs. Such heading(s) will seamlessly direct them to the desired information assets.
The interaction with the provided GUI results in the population of the search box of the traditional parametric search interface with terms corresponding to the various subject headings the searcher visits during the ongoing search process. The GUI is powered by an underlying ontology designed to capture the expressiveness of the formal LCSH structure. Each time a new heading is selected by the searcher, a new query is addressed to the underlying search engine. The whole process is concluded when the search results list contains information assets that match the searcher's information needs.

![Diagram](image)

**Fig. 2.** Prototype system architecture

It should be mentioned that information assets within the OPAC are linked to various indices apart from the index that contains subject headings. For the purposes of this research, the prototype GUI system interacts only with the subject headings index. In case of numerous search results, searchers are free to engage any other index (e.g. title, author) in order to narrow their search.

The two main components of the proposed architecture, namely the underlying ontology and the ontology-based GUI are described in the following sections.

### 3.2 The underlying ontology

The Library of Congress Subject Headings catalogue (LCSH) is employed for building the class hierarchy of the underlying ontology. LCSH is used to classify books and other collections with a subject-based classification approach. The LCSH List classifies all knowledge into twenty-one basic classes, appearing in an alphabetical, rather than hierarchical list. Most of these alphabetical classes are further divided into more specific subclasses. Each subclass includes a loosely hierarchical
arrangement of the topics pertinent to the subclass, going from generic to more specific ones.

For the purposes of this work, from a total of 265,000 LCSH’s records (terms), a subset of about 2000 terms have been employed, corresponding to the subject headings used in the Library of Ionian University, Collection of Informatics. Terms used in the Library of Ionian University are stemming from the LCSH List but have been translated into Greek and modified according to restriction rules with respect to language, technical term and spelling applied to holdings.

LCSH not only allows subjects (terms) to be hierarchically arranged, but also allows other statements to be made about such subjects (i.e. broader-than (BT), narrower-than (NT), seeAlso, use/use-for (USE/UF) relationships). The following paragraphs describe how the semantics of these statements are captured in the proposed ontology.

LCSH’s BT and NT relationships are used to build an asserted hierarchical ontology, where terms are represented as classes and relationships as properties (also known as slots). A BT relationship refers to a superclass having a broader or less specific meaning than its descendant classes. The inverse relationship according to the LCSH system is known as NT and is implied by the BT. NT represents all subclasses having a narrower meaning. Classes may have many super- or subclasses according to the LCSH taxonomy. Hierarchical navigation through the class hierarchy constitutes the basic query refinement tool for the searcher.

The relationship ‘seeAlso’ refers to a term class related to another class, without being a synonym or a broader/narrower term. Moreover, according to LCSH, there are headings accompanied by a term in parenthesis, for example ‘Programming (Mathematics)’. The context of parenthesis signifies a term with a broader meaning, many times (but not all) referring to an authority record. Such a context term association is represented as the property ‘in the context of’. Both RT and context association properties are exposed to the searcher by the GUI system, increasing the chances for serendipitous discovery of interesting terms.

According to the LCSH system, a USE/UF relationship implies that two terms are equivalent; one term refers to another (USE) that is to be preferred instead of this term (UF). In the proposed ontology such a relationship is represented as ‘individual’ or ‘instance’ of a collector LCSH term class, which groups all equivalent terms. Moreover, the role of individuals is expanded by adding the Greek versions of subject headings related to a formal LCSH term, as well as their UF terms.

The rdfs:label annotation property is employed to provide meaningful, human-readable labels to the underlying headings. USE/UF relationships facilitate non expert users to locate the concepts they are searching for.

Since the selected Library's repository indexes just the Greek version of the formal LCSH subject headings, instance labels serve as the keywords to be fed into the Ionian University Library's OPAC search engine.

3.3 The ontology-based GUI

The proposed GUI provides a web-based, user-driven visualization of the LCSH catalogue. Searchers interact with the GUI through a semantic-based process that
enables them going from heading to heading according to predefined semantic criteria (see fig. 1). Such a search tactic is argued to bridge the gap between the searcher's information needs and the Library's formal terminology concerning the LCSHs. As stated earlier in this paper, the GUI attempts to sustain the expressiveness of the underlying ontology without compromising fundamental usability design guidelines.

Thus, the whole interface progresses linearly from left to right in accordance to the applied user interactions. The searcher's focus is most of the times at the right end of the interface, which corresponds to the most recent interactions. Since scrolling to the right is provided automatically, user-initiated scrolling is minimal and the active part of the screen is not overcrowded. At any time, searchers have the ability to scroll to the left and observe their complete navigation path.

Moreover, searchers do not have to type in anything, since the GUI clones each selected heading to the search engine’s input box. This way, possible errors deriving from natural language interpretation are minimized. Searchers are capable of bypassing the GUI and type their own query terms, but such a scenario is beyond the scope of this paper.

The prototype also promotes serendipitous or casual discovery of interesting information connected to the current search process (see [8]). During the search process, searchers underpin their cognitive learning, since they are able to witness which information assets correspond to the formally defined subject headings they traverse.

Like any other search GUI, the proposed one should be assessed in terms of user’s information needs fulfillment speed. The provision of suitable entrypoints located as close as possible to the resulting subject headings, is a key factor affecting such a criterion. Further research is required in terms of entrypoint selection, possibly in the context of personalized information retrieval, as suggested by Garcia and Sicilia in [2].

As a proof of concept, readers are prompted to visit the demo version of the working prototype, which is available at http://195.251.111.53/server/entry/index.xhtml. The underlying ontology is encoded in OWL format, the gluing middleware is implemented in Python, and the provided GUI is based upon the most promising Ajax technology. The prototype is tested with the Mozilla Firefox Web browser (available at: www.firefox-download.com).

4 Conclusions – Future work

This paper presents a novel ontology-assisted information seeking process. Such a process utilizes not only the content but also the organizational schema of a domain ontology. This way, users may express their abstract needs to concrete concepts, subsequently indexed in digital collections. The presented interface can be seamlessly embedded into traditional search form interfaces, as a complementary query formulating tool.

According to the proposed method, a GUI exposes the hierarchy of the concepts included into a domain ontology, as well as all stated relations between different concepts, as links that the user can follow. Thus, searchers are able to effectively
traverse the ontology and at the same time formulate the actual query to the underlying search engine. This act of interactive navigation through the domain collection helps searchers in accurately formulating their queries. This is achieved by offering broader or narrower concepts for selection or indicating alternative or related concepts that searchers might be initially unaware of. The augmented exposition of inter-relations between concepts provides multiple paths for information seeking and enables searchers to fulfill faster, more efficiently and in an intuitive manner their personal information needs.

The process of information seeking by traversing a domain ontology is demonstrated via a prototype web-based application, which helps users traversing the online OPAC Library Catalogue of the Ionian University. Within the underlying ontology, terms are arranged as a complete hierarchy of concepts, together with interconnecting relations (i.e. “seeAlso”, ”inContextOf”) between concepts. Each concept class contains as instances the equivalent and multilingual terms of a particular concept.

The GUI of the prototype application visualizes the aforementioned information schema and allows for interactive traversal of the ontology. In each step of the information flow, corresponding terms are fed automatically into a search field of a typical search engine. Searchers may submit such terms to the library's OPAC, examine the results and redefine the query in subsequent steps by navigating to other parts of the ontology. The prototype GUI allows for a progressive expansion or reduction of the part of the ontology displayed, avoiding cluttering the interface with unneeded information.

In the proposed ontology-based traversing search process, the selection of strategic entrypoints is a critical aspect for it’s speed and effectiveness. For the time being, entrypoints are fixed and manually selected from the underlying collection. While this approach is sufficient for a prototype system, future work includes research for automated entrypoint extraction or dynamic definition via personalized information retrieval methods. Moreover, a comparison with similar approaches is under way together with a qualitative evaluation of the working prototype.

References