



# Enhancement of subject searching in OPAC using ontology

## Mejora de la búsqueda de temas en OPAC utilizando ontología

Rosalien ROUT 1

Received: 07/05/2018 • Approved: 29/06/2018

### Contents

- [1. Introduction](#)
  - [2. The current trend of subject searching in OPAC](#)
  - [3. Subject Searching Problems](#)
  - [4. Ontology-based IR](#)
  - [5. Proposed Searching framework](#)
  - [6. Process](#)
  - [7. Conclusion](#)
- [References](#)

#### ABSTRACT:

In this paper, we present a conceptual searching framework based on ontology to help users in query formation and make subject searching more effective in Online Public Access Catalogue (OPAC). The study proposes an ontology-based GUI through which users can visualize subject headings and interact with the GUI through a semantic-based process while searching. This model represents the library of Congress subject headings (LCSHs) in an ontological form which exposes the hierarchy of the subject headings as well as all stated relations among them. It facilitates context-based searching by replacing the keyword-based search and help users to formulate their query in such a format that, it will be more responsive for information retrieval.

**Keywords:** OPAC, Subject searching, LCSH, ontology, information retrieval, ontology-based semantic model

#### RESUMEN:

En este documento, presentamos un marco de búsqueda conceptual basado en ontología para ayudar a los usuarios en la formación de consultas y hacer que la búsqueda de temas sea más efectiva en el Catálogo de Acceso Público en Línea (OPAC). El estudio propone una GUI basada en ontología a través de la cual los usuarios pueden visualizar los títulos de los sujetos e interactuar con la GUI a través de un proceso semántico durante la búsqueda. Este modelo representa la biblioteca de los encabezados de materias del Congreso (LCSH) en una forma ontológica que expone la jerarquía de los títulos de las asignaturas, así como todas las relaciones establecidas entre ellos. Facilita la búsqueda basada en el contexto al reemplazar la búsqueda basada en palabras clave y ayuda a los usuarios a formular su consulta en un formato que responda mejor a la recuperación de información.

**Palabras clave:** OPAC, búsqueda de sujetos, LCSH, ontología, recuperación de información, modelo semántico basado en ontología

## 1. Introduction

Online Public Access Catalogue (OPAC) plays a vital role as an information retrieval tool in libraries. It has passed through different stages of development starting from the first generation featured OPAC to today's social OPAC (SOPAC) since its inception. But still, it is

being criticised on the ground of difficulty to use (Borgman,1996, Zumer,2007). Past studies have revealed that subject searching is the most problematic area in case of Online Public Access Catalogue (OPAC) searching. During last twenty-five years although various approaches have been proposed to improve the OPAC subject searching these are yet to be implemented (Yu & Young, 2004). The main problem in current OPAC is the searching interface design is inadequate for subject browsing (Mi & Weng 2008). The GUI of OPAC system is incapable of aiding the searcher in choosing the right keywords for successful results. Above all users could not utilize the advantages of the expressiveness of LCSH used in OPAC. Hence, they are unable to frame queries which appropriately reflect their real information needs (Schallier,2005).

The present OPACs work on a keyword-based retrieval system which produces results based on exact matching leading to low recall and precision rate. This limitation of the keyword-based search technique can improve by replacing it with ontology-based semantic searching. The ontology-based searching can interpret the meaning of user's query and the relations among the concepts that document contains. It helps to bridge the gap between users' information needs and the terminology that is usually employed to describe and index information assets within OPAC. In the above context, an ontology-based subject searching framework proposed in this study for library OPAC system. The very objective of the proposed model is to create a user-friendly OPAC system that has high retrieval performance and can able to understand the query intent. Along with this, it will also help in query expansion, searching and ranking of search results based on the developed ontology. This framework will be viable because it understands the query intent through the semantically extended design of the LCSH. It is expected to act as an improvement over the classical keyword-based approaches for the library OPACs.

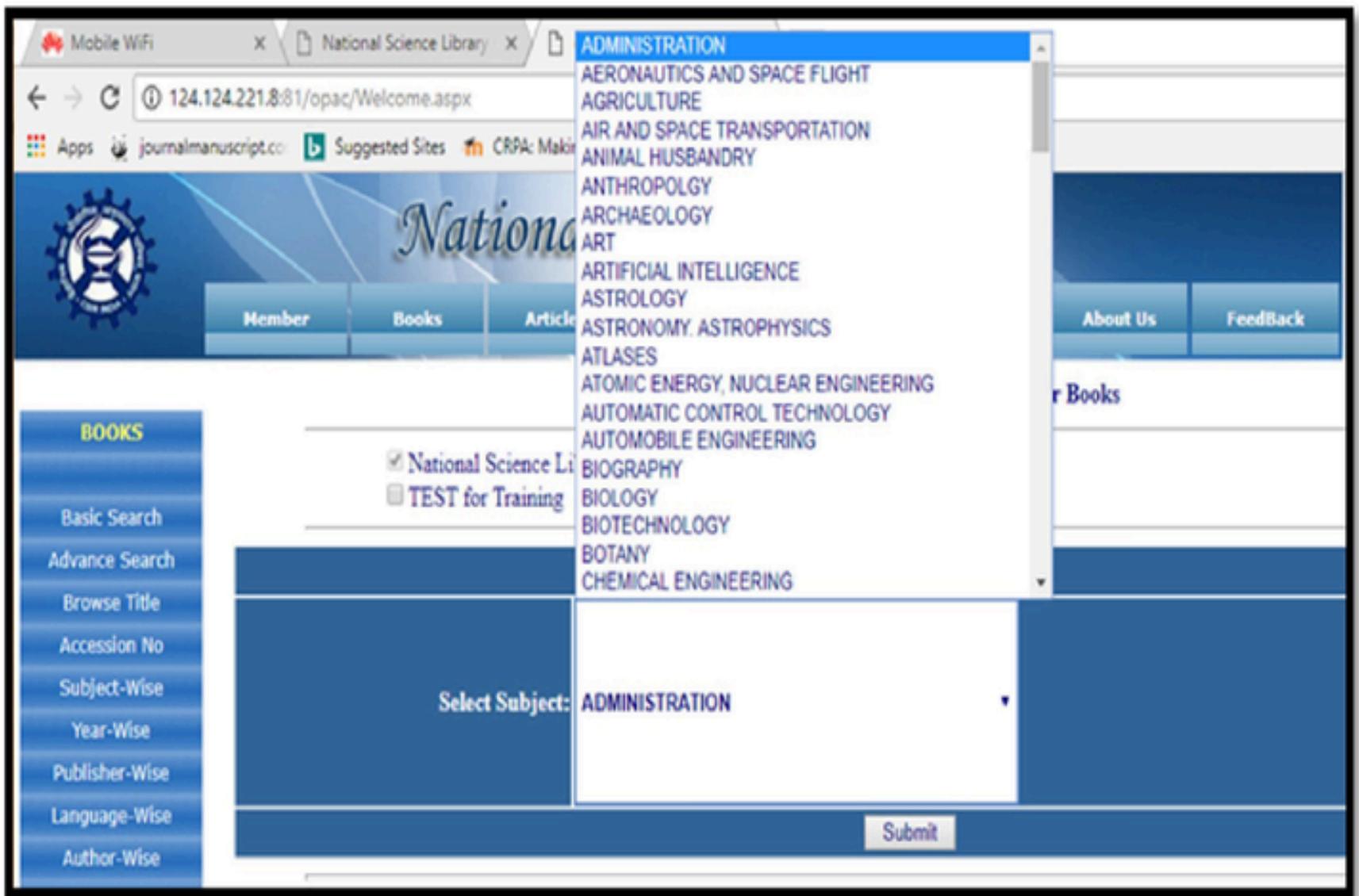
---

## 2. The current trend of subject searching in OPAC

Subject searching is one of the important searching options in the library OPAC system that uses controlled vocabulary which is capable of providing a meaningful descriptor for subject access. In the current MARC-based OPACs, subject searching is carried out by searchers either using a controlled vocabulary (browsing) or keyword approach. In case of controlled vocabulary searching, searcher put keywords and phrases that are closer to or express their information needs. However, it only leads to a successful result when terminology of the searcher matches the terminology of the underlying subject headings. In another way, OPACs offer the searchers alphabetical lists of index terms or precoordinated headings to initiate the search (figure-1). On the other hand, a keyword search approach which is also known as "free-text searching" looks for words that may appear in one or more fields of the record (e.g., title, notes, series statement, subject headings) (figure-2). Keyword searches with more than one words are automatically processed as Boolean 'AND' queries and access documents which contain the specified word(s) anywhere in the record. It allows users to specify the fields in which the searching is to be performed. Both the processes are based on exact-matching technique and for successful results usually requires exact matching between the query terms and subject headings (figure-3). This type of searching diminishes the semantic expressiveness of the LCSHs. Also, LCSH is so designed that it does not work well together with the Boolean searching provided by the current keyword-based OPACs systems. The subject searching problems are described as follows.

### Figure 1

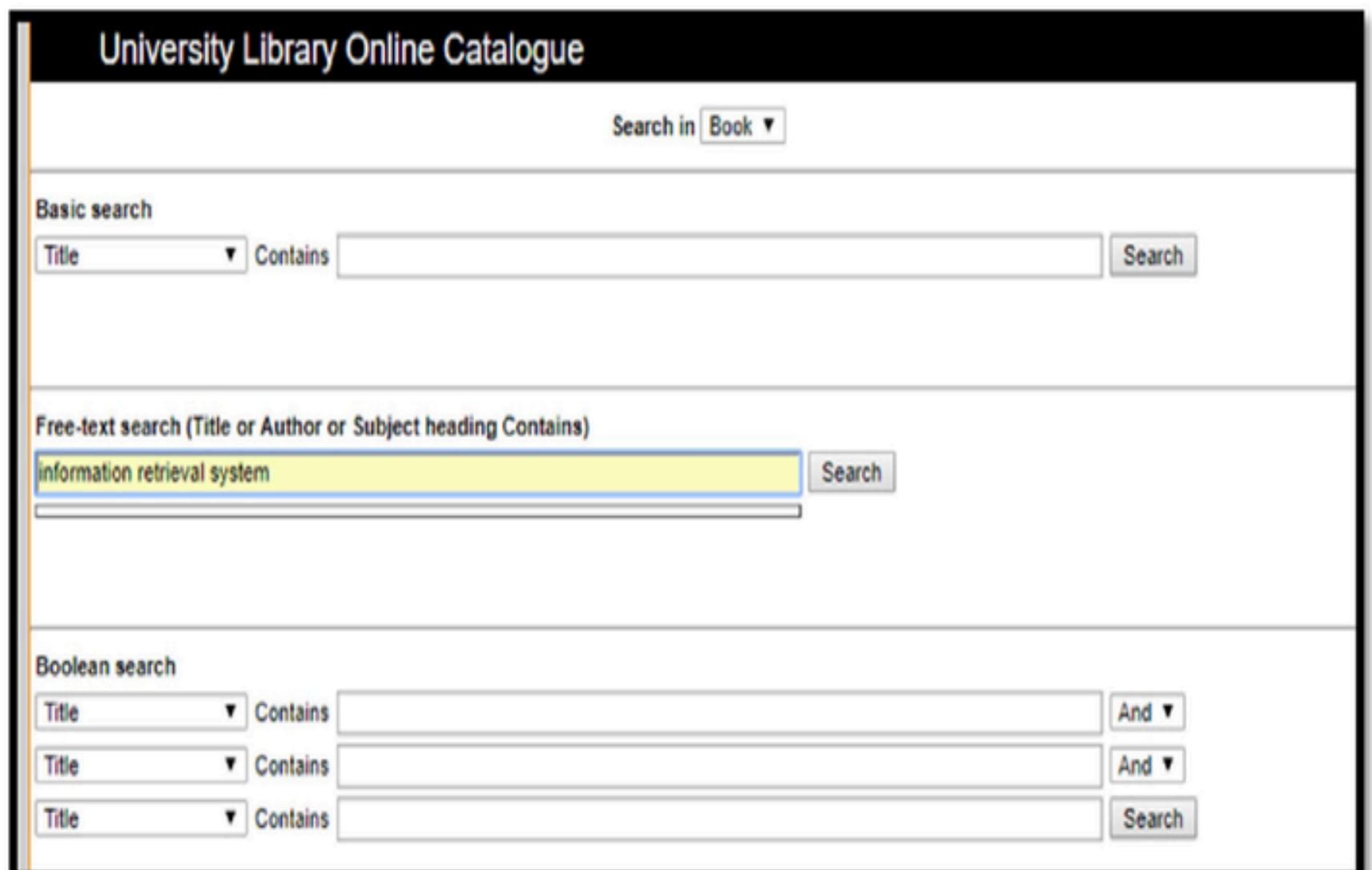
An alphabetical lists of index terms or precoordinated headings to initiate the search



-----

**Figure 2**

Free-text searching looks for words that may appear in one or more fields of the record



-----

**Figure 3**  
Exact-matching technique for successful results

The screenshot shows a library OPAC search interface. At the top, there are navigation tabs: Member, Books, Articles, Journals, Login, Logout, About Us, and Feedback. Below these is a search form with the following fields:

- Material:** All (dropdown), Document (dropdown)
- Search Text:** information retrieval system
- Is:** TITLE (dropdown)
- Operator:** Like (dropdown)
- Order By:** Title (dropdown)
- Sort By:** Ascending (dropdown)

Buttons for 'Submit' and 'Reset' are located below the search form. Below the search form, it indicates 'Total Records: 10'. A table displays the search results:

Sr No	TITLE	View Details
1	Evaluating information retrieval systems, the PBOBE program	<a href="#">View Detail</a>
2	Evaluation of factors affecting document ranking by information retrieval systems	<a href="#">View Detail</a>
3	File management and information retrieval systems	<a href="#">View Detail</a>
4	Information retrieval system	<a href="#">View Detail</a>
5	Information retrieval systems	<a href="#">View Detail</a>
6	Information retrieval systems	<a href="#">View Detail</a>
7	Information retrieval systems	<a href="#">View Detail</a>
8	On line information retrieval systems	<a href="#">View Detail</a>
9	Online information retrieval systems	<a href="#">View Detail</a>
10	Testing comparison and evaluation of information retrieval systems	<a href="#">View Detail</a>

On the left side of the interface, there is a sidebar menu with the following options: BOOKS, Basic Search, Advance Search, Browse Title, Accession No, Subject-Wise, Year-Wise, Publisher-Wise, Language-Wise, Author-Wise, and Recent Arrivals.

### 3. Subject Searching Problems

There is a good number of studies on OPAC use and searching problems. These studies revealed that subject searching is the most problematic area of online catalogues (O'Brien, 1994). Searchers are not very successful in matching their subject terms with the catalogue's controlled subject vocabulary (Markey, 1984). Even, they do not understand the structure of authorized subject headings and therefore have difficulty in identifying the correct subject terms to complete a comprehensive subject search (Drabenstott et al., 1999, Olson & Boll, 2001). The main reason behind ineffectiveness of subject searching may be users' inability to match their terminology with the subject system's terminology. Users are not familiar with the subject lists used in the catalogue such as LCSH (Library of Congress Subject headings) searches contain spelling and typographical errors (Yee, 1991). Apart from this other subject searching problems faced by users while OPAC searching are i) one-third of subject queries fail to produce results; ii) large retrievals (high recall) discourage users from scanning the results and iii) users are discouraged by subject access and look for alternative approaches (Drabenstott & Marjorie 1996). M.S. Sridhar (2004) identified different subject searching problems are: i) too many failed searches or no records retrieved; ii) retrieve unmanageably large number of records; iii) navigational frustrations; iv) failure to match the system's subject vocabulary; v) inadequate / lack of cross-references; vi) too many or too few bibliographic records linked to a subject heading; vii) lack of user perseverance. Studies also found that the design of traditional subject browsing OPAC interfaces is inadequate for subject browsing (Borgman, 1996, Papadakis et al., 2008). Users often feel frustrated and dissatisfied with either too many results or none when they use this type of search. Some users do not even know the existence of subject terms (Antell & Huang 2008). Therefore, the subject search is the most difficult and least chosen search

method in OPAC (Guo & Huang, 2011). The present MARC-based OPACs are relying only on the syntactic matching of words instead of considering its semantic meaning. Therefore, the precision rate is low, and the false detecting rate is high. Another problem is, searching becomes difficult when users do not know the appropriate searching terminologies. From the above discussion, it is clear that there is a need for a user-friendly subject searching system for OPAC. Thus, this paper presents an ontology-based searching framework for library OPAC system to overcome these problems.

---

## 4. Ontology-based IR

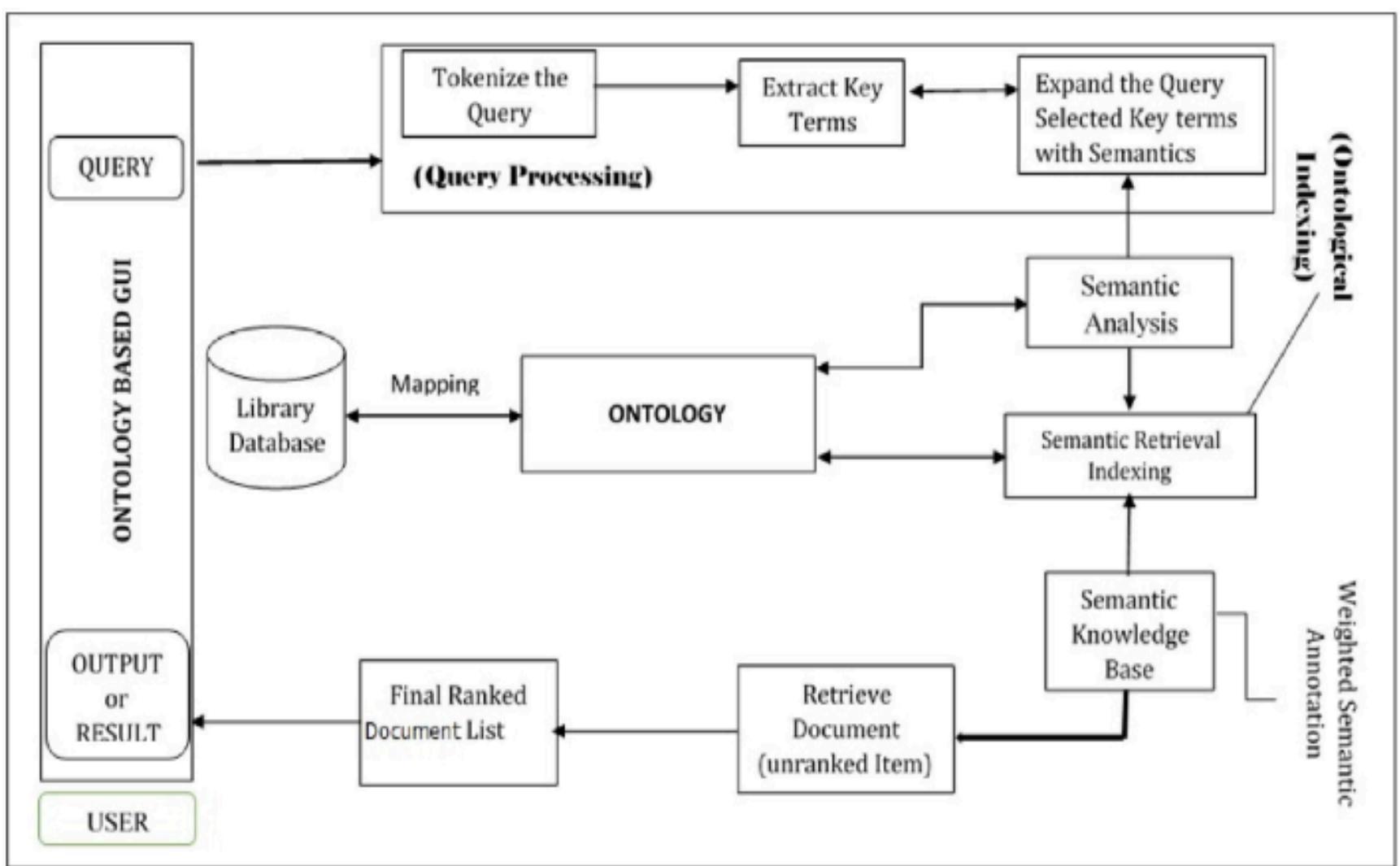
The main problem of keyword-based OPAC system is indexing of documents which occurs according to words rather than meanings. It lacks knowledge representation and semantic processing capabilities. Therefore, to improve searching and knowledge processing capabilities of the present OPAC system, it is necessary to add semantic information. Hence, an ontology-based searching framework is proposed here for this purpose. Ontology is a rich and formal logic-based language that specify meaning and relation of the concepts. It represents knowledge in a conceptual manner that can be distributed among various applications (Vanjulavalli & Kovalan, 2012). In case of information retrieval, it understands user's requirement and maps it into information resources. It describes the information in the semantic level and matches concepts by comparing the concepts in a logical structure. Ontology provides a good concept hierarchy and logical reasoning support for information retrieval (Vallet et al., 2005). An important aspect of the ontology-based retrieval system is its approach to the use of conceptual representations of content instead of plain keywords. In this proposed model, ontology is used to define relationships among the concepts of LCSH and represent them in the form of class, sub-class hierarchical relationship. It is useful in bridging of the gap that exists between the metadata provided by the indexer and the concepts presented by a searcher (Bechhofer & Goble, 2001). It also helps in query expansion and allows users to reformulate their query.

---

## 5. Proposed Searching framework

The proposed model is an ontology-based searching framework for library OPAC system. In this model, subject headings are mapped into ontology for semantic retrieval and improve subject discovery. It maps subject headings into an ontology-based class-subclass hierarchy and displays multi-directional relationship among them (Murakami et al., 2015). Thus, users can able to visualize, expand and reformulate their query. The ontology-based OPAC model architecture given in (Figure-4), has the following components.

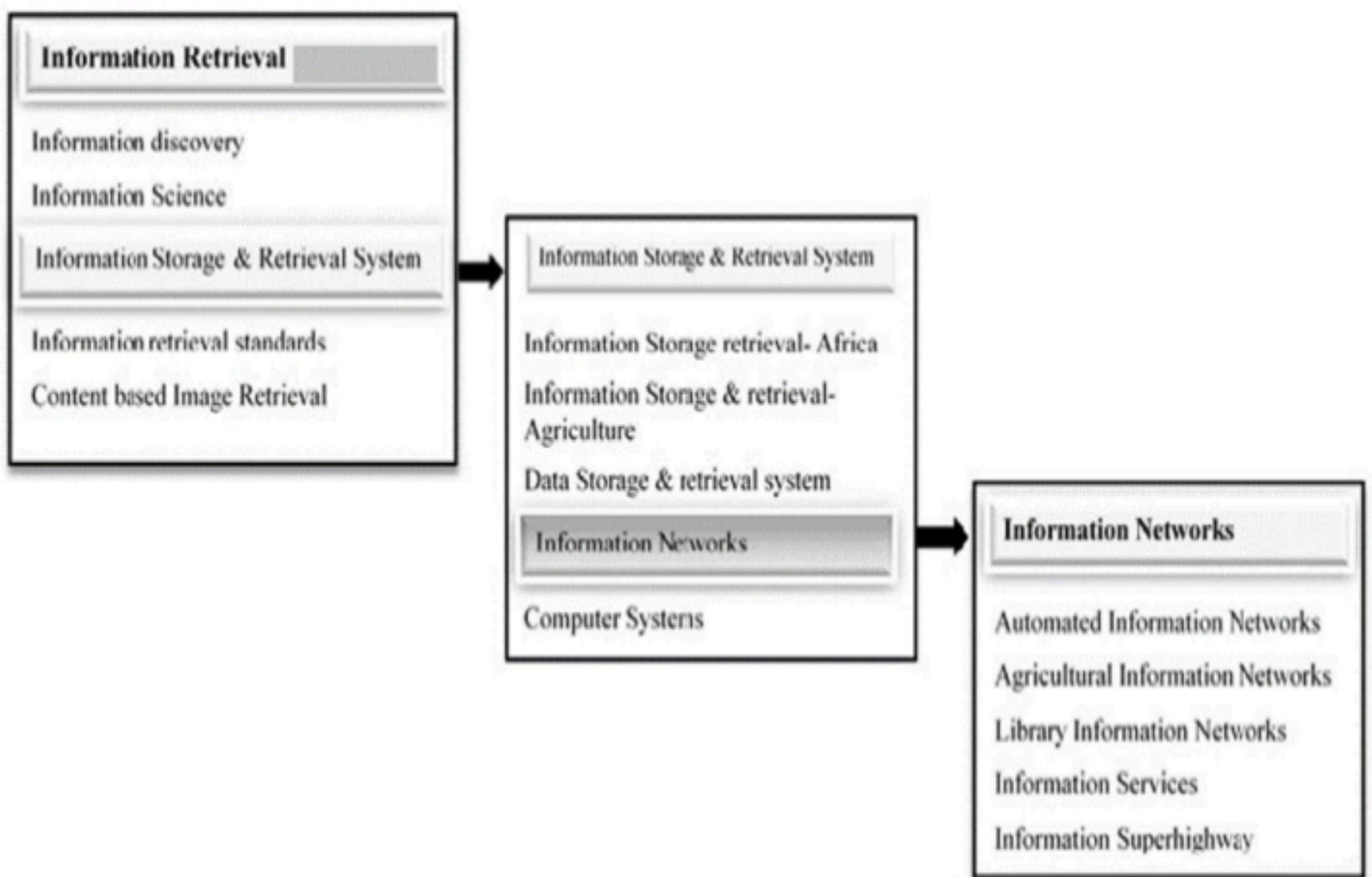
**Figure 4**  
Ontology-based OPAC model



## 5.1. Ontology-based GUI module

According to the prototype model, the OPAC system consists of an interactive GUI with an auto-suggestion search box. The users interact with the OPAC through the auto-suggestion search box (figure-5) which helps to modify or expand their query by selecting alternate terms that best match with their information needs. It also removes the problem of spelling mistakes and provides relevant result quickly. Through this interface, a user will input a query and the query request will be regulated using an underlying ontology. The ontology is developed from a library of Congress subject heading(LCSH) in order to make subject searching effective and easy. It also enables the user to visualize related subject heading structure and allows him to choose appropriate heading by moving from one heading to another heading through a semantic-based process.

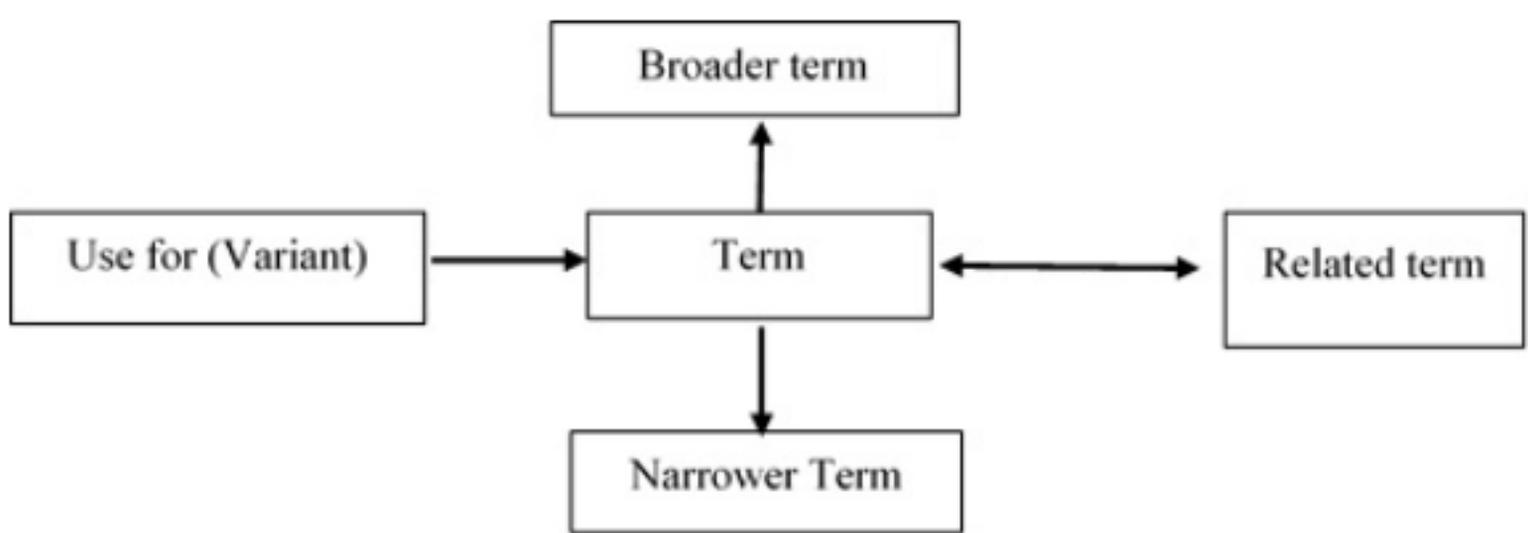
**Figure 5**  
Auto-suggestion box for query expansion



## 5.2. Ontology Construction

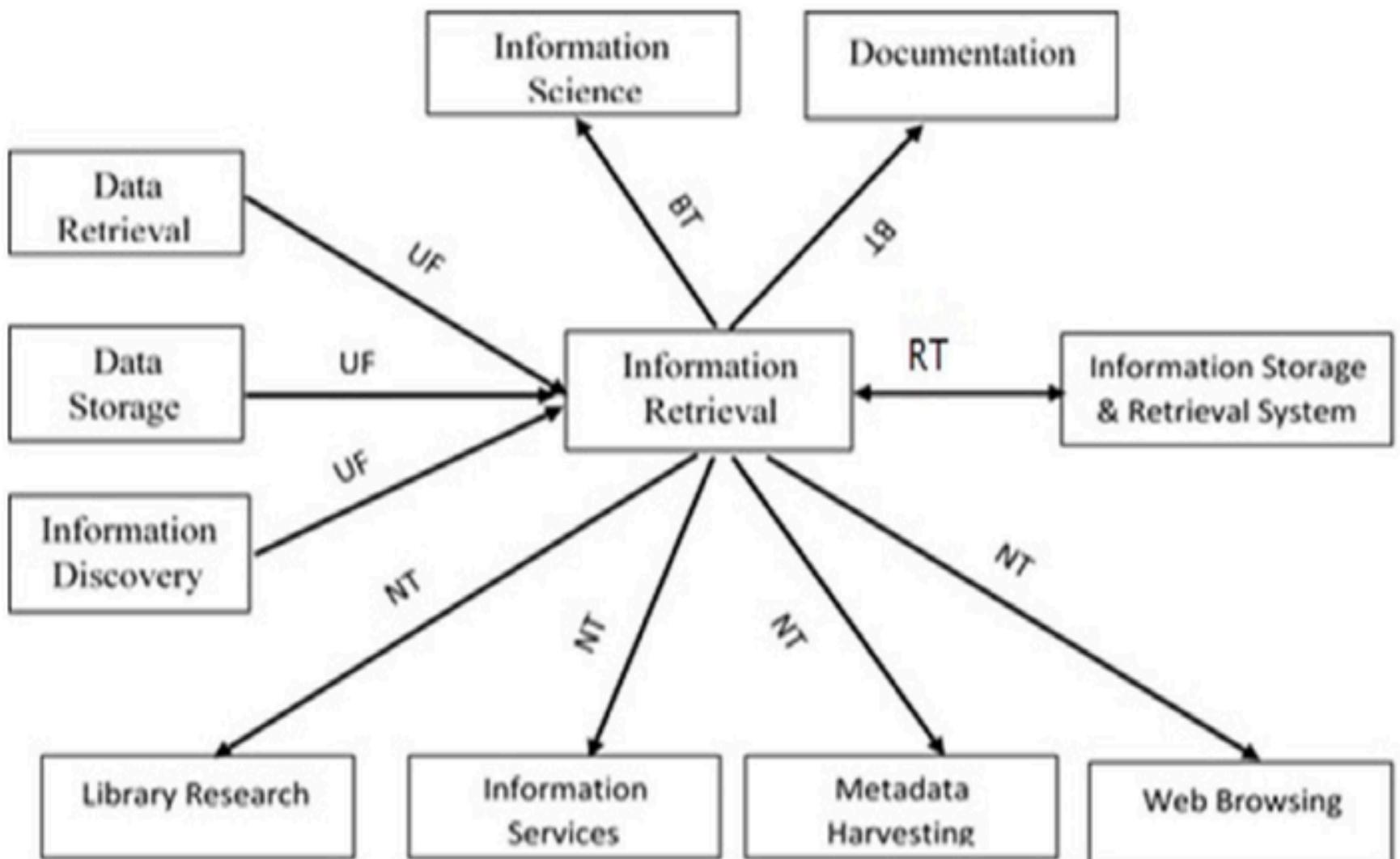
The key module of the prototype model is ontology construction. With the help of the proposed ontology schema (figure-6), LCSHs used within the OPACs are mapped into the ontology. The ontology allows conceptual representations of the contents and provides enough information about the property and relationship that the concepts may have with other terms. It defines classes, class hierarchy, and features of the ideas which collectively develop a structure for semantic interpretation and retrieval. The expressiveness of information within ontologies will help in providing more accurate results. Therefore, to make subject searching more effective, this model develops an ontology from LCSH. The ontology is composed of main topical headings and possible subdivisions along with scope/context notes. It includes broader terms (BT) as superclasses, narrower terms (NT) as subclasses and related terms (RT) excluded from it. USE/UF terms are accepted as semi-synonyms or synonyms of established terms (Murakami et al., 2015). A class has many super or subclasses as per the LCSH taxonomy. The hierarchical navigation through the class hierarchy will assist users in query expansion and eventually direct them towards desired information resources. Based on the schema an example is presented in (figure-7) where a term 'information Retrieval' chosen from the LCSH and transformed into ontology. According to the ontology schema, the broader terms (BT), narrower terms(NT), related terms (RT) and USE/UF terms of 'Information Retrieval' are arranged and presented.

**Figure 6**  
Ontology schema for LCSH



**Figure 7**

Example based on the ontology schema



### 5.3. Query processing module

Query processing module is also known as semantic analysis module which carries out query processing work by expanding user's query semantically. Through the GUI user will input a query and the query related concepts extracted from the built-in ontology with the help of ontology-based matching algorithm (Dan & Hui-Lin, 2006). This process is more important since the given query will pass through the parser and semantically analyzed words are obtained. It exploits the ontology to understand the semantic relation of the query, which means to identify the different relationship between the concepts in the ontology and the user's query. These words are then matched with the concepts contained in the ontology to get a set of more related keywords. Thus, by the semantic expansion, this module will generate a new search expression/words and allow users to reformulate their query. (Dan & Hui-Lin, 2006, Ruban, 2014).

## 5.4. Semantic retrieval module

This module carries out indexing function and creates the ontological index. It proposes to use Apache Lucene for indexing and retrieval tool (Fernández et al., 2011) to achieve high performance and scalability. It extends traditional keyword-search with extracted and inferred information using developed ontology. This module is also based on free-text searching and helps in achieving semantic retrieval by implementing a custom ranking for Lucene indices so that documents containing ontological information get higher rates.

## 5.5. Document Processing

This module is mainly responsible for the indexing of information resources of the library and finally gets the semantic knowledge library. It is the process of knowledge acquisition by transforming or mapping subject headings into ontology classes. It extracts subject headings from the bibliographic database which are stored in metadata format, then encoded it and stored them in a semantic metadata database.

## 5.6. Output/search results Module

In the ontology-based search framework, the ontology-aided index structure is the basis for the retrieval of the semantic knowledge database. Based on the ontology data structure users formulate their query, then the query is executed, matched documents are retrieved, and output will be sent to the users. At first, the query terms are matched with the metadata stored in the semantic database and retrieves a set of most relevant documents (Fernández et al., 2011). Then the retrieved documents are pre-processed using the pre-processing module, and the results are prioritized according to sorting algorithm and provided to the users through the GUI (Papadakis et al., 2008).

---

## 6. Process

The prototype system has been developed to demonstrate the ontology-based search framework. In this process, user input query through ontology-based GUI, then query was processed with the help of developed ontology (LCSH data store structure). The query interface is designed to match the hierarchical structure of the ontology datastore in the form of classes/subclasses relationship. It represents keywords/sub-keywords in the user query interface. In other words, the user's query is semantically expanded with related concepts on the basis of developed ontology (Papadakis et al., 2008). Then the semantically enhanced query sent to users to review and after that the user selects particular term/concept. Unless and until the user was satisfied with the provided options, the process will be repeated till the user is not picking a specific query out of the expanded terms. After that, the query is executed and searches for documents in the semantic knowledge library (i.e., semantically indexed cluster of documents) to retrieve the relevant documents. Once the query is processed, matched documents are retrieved, and output will be sent to the users to review. If the desired result is not returned, the user repeats the query, and the process continue until the desired result is obtained. The developed hierarchical structure based LCSH ontology would function as a dynamic query interface to assist users in formulating their query easily and quickly. This way the problem of query formulation of not knowing the right keywords for searching can be solved. By this system, the searcher will no more to search the database on the basis of keywords that were included in the index of the database. Instead they can formulate their query based on the ontology of keyword that has been identified. That means searcher's own query determine how the search should be conducted. Thus it will make the searching process and the model more efficient and effective.

---

## 7. Conclusion

This study has introduced ontology-based navigation for library OPAC. As per the proposed

model, users can formulate their query with the help of interactive auto suggestion box and search the library collections semantically. In case of OPAC searching, usually, users face problem in query formation. They are unaware of the formal terminology or subject heading that is used as index term (i.e., represent the content of the document). Hence, the proposed interface assists the user to express his query efficiently with minimal effort. The model describes LCSH in ontology form where concepts are linked together through implicit and explicit relations (Papadakis et al., 2008). It expands user's query by visualizing related headings and allows to reformulate their query by choosing an option which represents their need appropriately (Kiryakov et al., 2004). The ontology-based model makes retrieval process fast and provides multi-dimensional navigation paths for searching by interlinking meta-information contained in the subject heading ontology. The ontology-based searching framework would replace the problems of current keyword-based OPAC systems by creating an inverted index based on semantic entities (meanings) associated with the documents. The keyword-based retrieval mainly relies upon exact matching using a statistical algorithm instead of exposing the expressiveness of the LCSH-based strings to the end user. This system can be easily integrated to any existing OPAC system. The study can be further extended by enabling the system to include user defined tags (folksonomies) for the OPAC subject searching. These tags after translated into ontology will offer a dynamic query interface, that will help users to express their needs more specifically and appropriately.

---

## References

- Antell, K., & Huang, J. (2008). Subject searching success: Transaction logs, patron perception, and implications for library instruction, *Reference & User Services Quarterly*, 48(1) 68-76
- Bechhofer, S., & Goble, C. (2001). Thesaurus construction through knowledge representation, *Data & Knowledge Engineering*, 37(1) 25-45.
- Borgman, C L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7), 493-503.
- Dan, W. U., & Hui-lin, W. A. N. G. (2006). Role of ontology in information retrieval. *Journal of Electronic Science and Technology*, 4(2), 148-154.
- Drabenstott, Karen M. & Marjorie, S Weller. (1996). Failure analysis of subject searches in a test of a new design for subject access to online catalogues, *Journal of the American Society for Information Science* 47(7) 519-537.
- Drabenstott, K M., Simcox, S., & Fenton, E G. (1999). End-user understanding of subject headings in library catalogs. *Library Resources & Technical Services*, 43(3) 140- 160.
- Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E., (2011). Semantically enhanced information retrieval: An ontology-based approach. *Web semantics: Science, services and agents on the world wide web*, 9(4) 434-452.
- Guo, J., & Huang, J. (2011). Subject headings and subject search: A comparative study, *Chinese librarianship: An international electronic journal*, 31,1-17. Retrieved from <http://www.iclc.us/cliej/cl31GH.pdf>.
- Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2(1), 49-79. Retrieved from [sciencedirect.com/science/article/pii/S1570826804000162](http://www.sciencedirect.com/science/article/pii/S1570826804000162).
- Markey, Karen. (1984). Subject searching in library catalogs: before and after the introduction of online catalogs. (OCLC Library, Information, and Computer Science Series; Dublin, Ohio).
- Mi, J., & Weng, C. (2008). Revitalizing the library OPAC: Interface, searching, and display challenges, *Information technology and libraries*, 27(1) 5-22.
- Murakami, H., Tang, Z., & Kurihara, A. (2015). Design and implementation of system for exploring subject headings. *Proceedings of the Association for Information Science and Technology*, 52(1)1-4. Retrieved from <http://>:

onlinelibrary.wiley.com/doi/10.1002/pa2.2015.145052010086/pdf.

O'Brien, A. (1994). Online catalogs: Enhancements and developments, in Martha E. Williams (Ed.) Annual review of information science and technology, (29) 219-242.

Olson, H A., & Boll J J. (2001). Subject analysis in online catalogs, 2nd edn (CO: Libraries Unlimited;Englewood). Retrieved from <https://books.google.co.in/books?isbn=1563088002>.

Papadakis, I., Stefanidakis, M., & Tzali, A. (2008). Visualizing OPAC subject headings. LibraryHiTech,26(1),19-23. Retrieve from [www.emeraldinsight.com/doi/abs/10.1108/07378830810857762](http://www.emeraldinsight.com/doi/abs/10.1108/07378830810857762).

S, Ruban., Tendolka, Kedar., Rodrigues, Austin Pete., & Niriksha, Shetty. (2014). An ontology-based information retrieval model for domesticated plants, *International Journal of Innovative Research in Computer and Communication Engineering*, 2(5) 202-213. Retrieved from at <https://www.rroij.com/.../an-ontologybased-information-retrievalmodel-for-domesticatedplants>.

Schallier, W. (2005). Subject retrieval in OPAC's: a study of three interfaces. 557-567. Retrieved from <http://www.bd.ub.edu/isko2005/schallier.pdf>.

Sridhar, M S. (2004). Subject searching in the OPAC of a special library: problems and issues, OCLC Systems & Services: International digital library perspectives, 20(4) 183-191.

Vanjulavalli, N., & Kovalan, A. (2012). Ontology based Semantic Search Engine, *International Journal of Computer Science & Engineering Technology*, 2(8) 1349-1353.

Vallet, D., Fernández, M., & Castells, P. (2005). An ontology-based information retrieval model, *In European Semantic Web Conference*, Berlin, Heidelberg May 2005, 455-470.

Yee, M M. (1991). System design and cataloguing meet the user: user interface to online public access catalogs, *Journal of the American Society for Information Science and Technology*, 42 (2) 78-98.

Yu, H., & Young, M. (2004). The impact of web search engines on subject searching in OPAC, *Information technology and libraries*, 23(4) 168-180.

Zumer, M. (2007). Amazon: competition or complement to OPACs. *Bid: textos universitaris de, Biblioteconomia i Documentació*, (19) Retrieved from <http://www.bid.ub.edu/pdf/19zumer2.pdf>.

---

1. Research Scholar at University of Calcutta and working as a Librarian at Malda Women's College, Malda, West Bengal, India-732101.  
Email- [rosalien222@gmail.com](mailto:rosalien222@gmail.com)

---

Revista ESPACIOS. ISSN 0798 1015  
Vol. 39 (Nº 44) Year 2018

[Index]

[In case you find any errors on this site, please send e-mail to [webmaster](mailto:webmaster)]

©2018. revistaESPACIOS.com • ®Rights Reserved