Applying domain knowledge and academic information to enhance unknown-item search in OPAC

Peerasak Intarapaiboon and **Chainarong Kesamoon***
Department of Mathematics and Statistics (LC.3 Bld.),
Faculty of Science and Technology,
Thammasat University, Pathum Thani, 12121, THAILAND
e-mail: peerasak@mathstat.sci.tu.ac.th;
chainarong@mathstat.sci.tu.ac.th* (corresponding author)

ABSTRACT

Many students usually use the unknown-item search strategies, including subject and keyword searches, to retrieve books or other materials provided in library catalogs. However, the success rates for unknown-item searching is relatively low compared with the known-item search strategies, i.e., title or author searches. In this paper, a framework for improving the unknown-item search is proposed. The main contributions of our framework concern both user's keywords and book indexing: (i) To enhance a user's keyword, the framework will select other relevant terms in a domain-related ontology; (ii) Topics expressed in course description are used as book indexing. A preliminary experiment shows that the proposed framework gives satisfactory results in terms of the numbers and the precision scores of retrieved books. Furthermore, the proposed interesting-score measure can facilitate to lift the precision levels.

Keywords: Digital library; OPAC; Ontology; Semantic search; Search strategies.

INTRODUCTION

Due to the emergence of new academic disciplines, the numbers of books available in university libraries grow exponentially -undoubtedly, many books are provided for a single academic field. Thus, a freshman student who has less experience in library skills is faced with a list of either insufficient or excessive search results. Consequently, an efficient system for book searching is necessary.

Most of libraries use Online Public Access Cataloging (OPAC) for easy access of books and other materials. Since this study is concerned with books, the explanation of OPAC is necessary in relation to this. There are four basic searching types in OPAC, i.e. Author, Title, Subject, and Keyword searches. We can classify the four search strategies into 2 groups: *known-item search* (including author and title searches) and *unknown-item search* (including keyword and subject searches). Users will use the former group when they have a particular item in mind and they want to determine whether the library holds that item, while the others will do the latter group when they have an interesting subject in mind, but no known title. So far, many statistical reports have indicated that the success rates for known-item search is higher than those of unknown-item search (Antell and Huang 1998; Slone 2000; Hessel and Fransen 2012; Rondeau 2013; Wakeling et al. 2017). One crucial rationale behind the failure of unknown-item search is that an interesting topic in a user's mind does not match with the bibliographic records. More precisely, it might be due to the lack of the user's

https://doi.org/10.22452/mjlis.vol24no1.3

experience to create suitable keywords on one hand and the not up-to-date book indexing on the other. In this work, a novel framework for improving unknown-item search is proposed. The main technical challenges we focus in this work are threefold:

- *How to improve a keyword representing the topic in a user's thought*: If a user is not familiar to book indexing systems, he/she tends to express the information need with unsuitable keywords. In this framework, one module for expanding a user's keyword to other related terms based on a domain-specific ontology is introduced. With those extended terms, the possibility to obtain the relevant books could be increased.
- *How to assign more meaningful indexes to a book:* With regards to this question, academic information is taken into account. Each academic program has its own course catalog in which both description of an individual course and relationship among courses are detailed. When a semester begins, each instructor usually provides a course syllabus to his/her students. Reference textbooks are important content therein. In this work, we utilize topics in course description for textbook indexing.
- *How to select compatible books in the library catalog, when a textbook contains the user's topics of interest, but it is not in the catalog*: In this work, a title-based similarity measure is presented to determine a degree of similarity between books.

In the experiment for retrieving mathematical textbooks, we will see that the proposed framework returns relevant textbooks more than the traditional system (OPAC). In addition, we can refine the results by selecting the book with high scores where the relevant-book scoring is also part of the proposed framework.

LITERATURE REVIEW

So far, many statistical reports have indicated that the success rates for known-item search is higher than those of unknown-item search (Antell and Huang 1998; Slone 2000; Hessel and Fransen 2012; Rondeau 2013; Wakeling et al. 2017). As an evidence, Antell and Huang (1998) analyzed the search transaction log of the University of Oklahoma Libraries OPAC. The report revealed that, among all subject search occurrences, 48.8 percent of them yielded zero results and 10.6 percent yielded more than five hundred results. In the report, searches that yielded either zero results or more than five hundred results were considered to be unsuccessful.

Many rationale behind the relatively low success rates of unknown-item searches are explored. Some of them are: (i) since users, particularly undergraduate students, are not familiar with the subject lists used in the libraries, they cannot match terminologies in their mind with the suitable terminologies providing in the subject heading structures (Long 2000); (ii) when a concept in the user's thought contains multiple terms, the Boolean operators can be used to make a searching term that is semantically closer to the user's thought (Wood, Smigielski and Haynes 2007). However, many users do not understand the Boolean operators well; (iii) the catalog interface does not give users adequate guidance in finding and using LC terms or in revising their searches (Breeding 2007; Martell 2008).

Several methods have been proposed to raise the rate of success in unknown-item search. Cousins (1992) revealed that the percentages of exact matches between users' queries in historic data and the three wildly-used index systems, i.e., Dewey Decimal Classification (DDC), Library of Congress Subject Headings (LCSH) and PRECIS, are ranging from 30 to 94 percent (62.83% on the average). Owning to the low coverage scores, the author claimed that the quality of index systems are inadequate. A new way to enhance index systems was

proposed by indexing bibliography records with selected natural language queries from users in the historical database. The experimental results indicated that natural language enhanced indexing significantly outperform the traditional indexing.

Long (2000) introduced 17 guidelines that can be incorporated into the OPAC systems to help users perform unknown-item searches more efficiently. Those guidelines are: (i) the user's search terms should be highlighted in retrieved records; (ii) users should be told what subject list is used in the library catalog; (iii) the catalog should include search features that incorporate the entire cross-reference structure of subject headings.[1] Based on the guideline, the authors then evaluated the OPAC systems of 31 libraries. The results show that most systems are deficient.

In library science, authority control is a process that organizes bibliographic information. The typical library examples are the set of all books written by an individual author; the set of terms referring to the same object. Utilizing authority files, O'Neill, Kenneth and Kammerer (2014) proposed a two-step prototype to improve subject search. In the first step, the authority file is searched to find the appropriate subject heading to the user keyword. Then, in the second step, the bibliographic records are searched to identify the resources with the selected subject heading.

Rondeau (2013) mentioned that unknown-item search is a much more ambiguous process, often fraught with a certain degree of anxiety followed by an even greater sense of self-satisfaction upon successful retrieval. Since many library users interact with the OPAC in their searches for known and unknown items, the interface plays an essential role in this interaction. The author, then, explored and gave sugestions about the catalogue interface relating to the retrieval of unknown items.

Wood, Smigielski, and Haynes (2007) reported that, with a domain specific subject heading, namely Medical Subject Heading (MeSH), medical students can get to a list of articles that are relevant to their searches. For example, the results of searching using the phrase “sore throat” do not get the same results as searching using the term “pharyngitis”.

Ontologies provide conceptual models for representing and sharing domain knowledge. Generally, an ontology consists of concepts, semantic relations among these concepts. Ontologies have been mentioned and applied to enable the fulfillment of several applications including with digital library. For example, Castro, Giraldo, and Castro (2010) presented a prototype that allows users to annotate content within digital libraries. The annotation is then built upon an ontology. As results, the users can use semantic query to retrieve interesting information. Noah et al. (2010) introduced an ontology-driven framework for more semantic thesis search. By the ontologies focusing on academic thesis, the thesis metadata and content are inserted and populated to a knowledgebase. The ontologies contain not only concepts, e.g. ‘Supervisor’, ‘Contributor’, ‘Thesis’, but also various semantic links, e.g. ‘hasContributor’, ‘superviseBy’. It allows users to apply sophisticated query and searching such as “Find the supervisor of Arifah Alhadi and the title of her thesis.” (This query is hard to be solved by keyword search.) Zaid and Lau (2014) embedded an ontology into an academic information search system at a university in Malaysia to assist inexperienced students for searching academic resources. Khan and Bhatti (2017) conducted interviews with librarians and academicians about semantic web technologies. The thematic analysis indicated that the next-generation digital libraries will use semantic web technologies, which ontologies are thereof, to provide accurate results.

[1] Subject Headings are terms assigned to books to describe the content found within the books.

Moreover, semantic web applications should be included in the library and information science (LIS) curriculum.

THE PROPOSED FRAMEWORK

In this section, we describe our proposed framework for improving unknown-item search of library books searching. **Error! Reference source not found.** shows the proposed framework. When a user submits a keyword, the module Generation of Word List (GWList) will generate a list of terms relating with the user's keyword by using a domain ontology. Then, Generation of Course List (GCList) will search through the curricula in order to retrieve courses relevant to such a word list and create a course list. Based on the course list and a book database, Generation of Book List (GBList) creates a list of books and, finally, each book will be associated with a relevant score. The details of ontology and each component are described in the following sections.

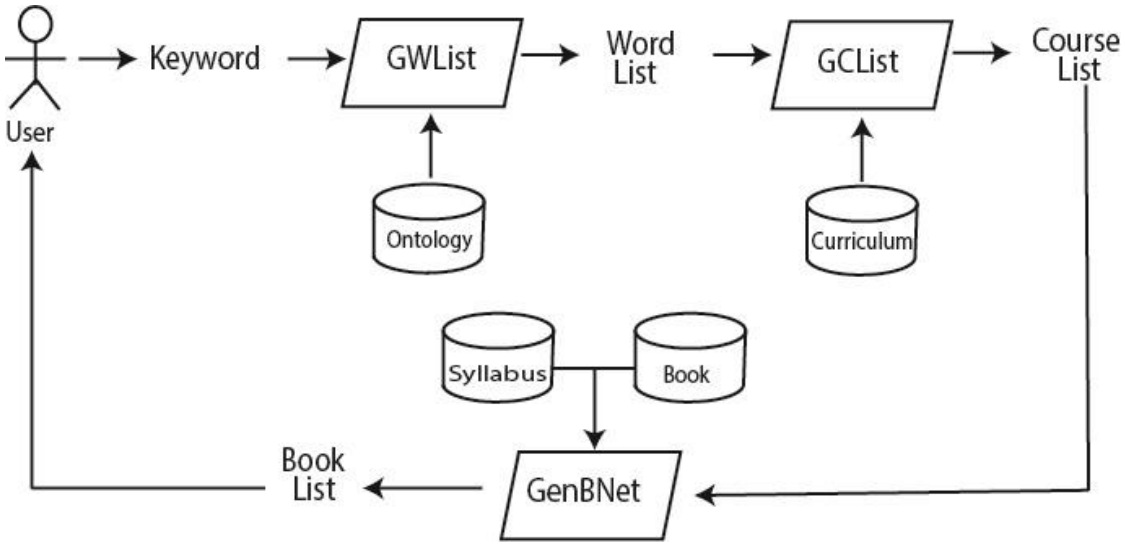

Figure 1: The Proposed Framework for Improving Unknown-Item Search of Library Books

Ontology

OntoMathPro[2] is the ontology used in this work. It is an OWL ontology that is geared to be the hub of mathematical knowledge in the Web of Data. There are 3,449 mathematical concepts. Each concept relates to others by 4 relation types, i.e. “belong-to”, “defined-by”, “see-also”, and “solved-by”.

Figure 1 shows parts of the ontology as a tree. A node represents a mathematical concept where a link represents a “belong-to” relation meaning that a lower node “belongs to” the upper node. For instance, ‘Analytical Geometry’ belongs to ‘Geometry’.

[2] The full description of this ontology, OntoMathPro, is appeared in (Nevzorova, Zhiltsov, Kirillovich, & Lipachev, 2014).

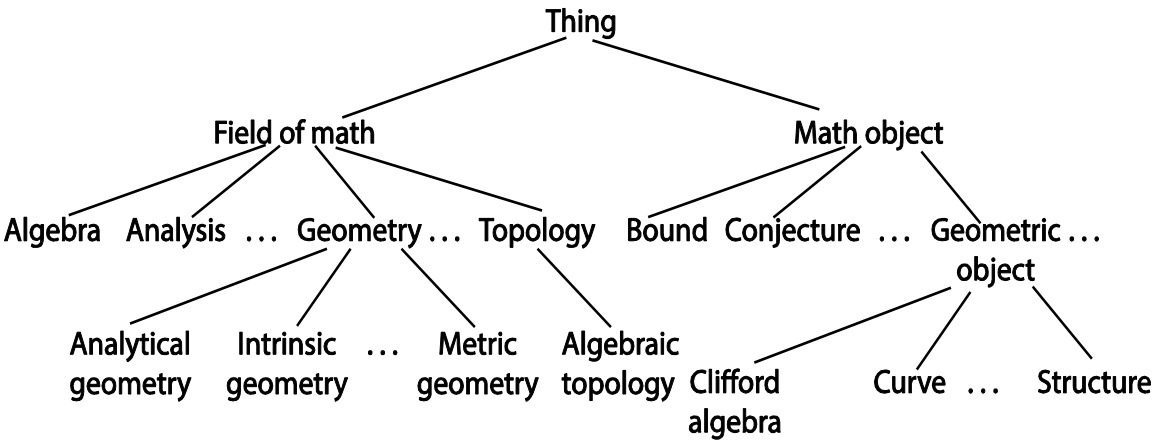

Figure 1: A Part of an Ontology

GWList: Generation of Word List

The output of this process is a list of words associated with degrees of similarity to the user's keyword. The similarity level is determined relying on the distance in the ontology. Intuitively, the more two concepts are close in terms of their structural properties, the more they are similar. For example, in

Figure 1, `Analytical geometry' is 2 steps away from `Metric geometry' (i.e., `Analytical geometry' → `Geometry' → `Metric geometry'), while 4 steps away from `Algebraic topology' (i.e., `Analytical geometry' →`Geometry' → `Field of Math' →`topology' →`Algebraic topology'). Then, the similarity between `Analytical geometry' and `Metric geometry' should be higher than that between `Analytical geometry' and `Algebraic topology.' The details of the process is discussed below:

1. Let w_0 be a user's keyword, and W be the set of all words (or concepts) in the ontology in use.
2. Calculate the similarity degree between w_0 and each concept c in the ontology by the following formula (Batet, Sánchez, & Valls, 2011):

$$S(w_0, c) = \frac{2D_{LCA}}{D_{w_0} + D_c}, \quad (1)$$

where D_{LCA}, D_{w_0} and D_c are the distances from the root node to the least common ancestor of w_0 and c, that to w_0 and that to c, respectively. The distance between two nodes is the number of edges linking the nodes. As an example, the distance between `Analytical geometry' and `Algebraic topology' is 4.

3. By a pre-specified threshold α, a list of relevant words with their similarity levels is generated as follows:

$$LW_{w_0}^{\alpha} = \{(w, S(w_0, w)) | w \in W, S(w_0, w) \geq \alpha\}. \quad (2)$$

It is noteworthy that the threshold α indicates the level similarity between extended keywords w and the original keyword w_0. As the result of the threshold, the higher value of α we set, the more relevant textbooks we get.

Example 1.
Consider the ontology in
Figure 1

$$S(\text{Analytical geometry}, \text{Metric geometry}) = \frac{2D_{\text{Geometry}}}{D_{\text{Analytical geometry}} + D_{\text{Metric geometry}}} = \frac{2\times 2}{3+3} = \frac{2}{3},$$

while

$$S(\text{Analytical geometry}, \text{Algebraic topology}) = \frac{2D_{\text{Field of Math}}}{D_{\text{Analytical geometry}} + D_{\text{Algebraic topology}}} = \frac{2\times 1}{3+3} = \frac{1}{3}.$$

The results are satisfactory to our intuition mentioned above that the similarity between "Analytical geometry" and "Metric geometry" should be higher than that between "Analytical geometry" and "Algebraic topology."

GCList: Generation of Course List
Given a set of ordered pairs relating with words and their similarity degrees to the user's keyword, w_0,

$$LW_{w_0}^{\alpha} = \{(w, S(w_0, w)) | w \in W, S(w_0, w) \geq \alpha\}.$$

In this process, every course in the course catalogs that its description contains at least one word in $LW_{w_0}^{\alpha}$ is retrieved. Then, the interest scores for the selected courses are determined. The formal steps in this process are details as follows:

1. Denoted by $Des_k = \{d_{k,1}, d_{k,2}, \dots, d_{k,j_k}\}$ the set of contents in the course description of subject sub_k where j_k is the number of contents in the course description for sub_k.
2. If there exits $d_{k,m} \in Des_k$ such that $d_{k,m} = w_i$ for some $w_i \in LW_{w_0}^{\alpha}$, then the subject sub_k is retrieved.
3. Denoted by $Score_{w_0}(sub_k)$ the interest score for the retrieved subject sub_k with respect to the keyword w_0, where

$$Score_{w_0}(sub_k) = \frac{\sum_i^{|LW_{w_0}^{\alpha}|} \alpha_i \chi_{Des_k}(w_i)}{\sum_i^{|LW_{w_0}^{\alpha}|} \chi_{Des_k}(w_i)}, \quad (3)$$

$$\chi_{Des_k}(w_i) = \begin{cases} 1, & w_i \in Des_k \\ 0, & w_i \notin Des_k \end{cases}. \quad (4)$$

Roughly speaking, the score for a course is the average of the relevant scores corresponding to the terms in $LW_{w_0}^{\alpha}$ that appear in the course description.

GBList: Generation of Book List
For each sub_k selected from the previous process, the textbook's title for that course syllabus is then retrieved. Moreover, we associate such a book with the interest score which is equal to that score of its corresponding course. It means that if $Score_{w_0}(sub_k) = \beta$, then the interest score of the course textbooks is β. Denoted by

$$BK_{w_0} = \{(b_1, \beta_1), (b_2, \beta_2), \dots, (b_p, \beta_p)\},$$

the set of books and their interest scores with respect to the user's keyword w_0 when the first and the second entries of each ordered pair are a book title and an interest score, respectively.

Based on the book titles and their call numbers, we will extend BK_{w_0} to discover more interesting books. The book-extension process is shown in Figure 2.

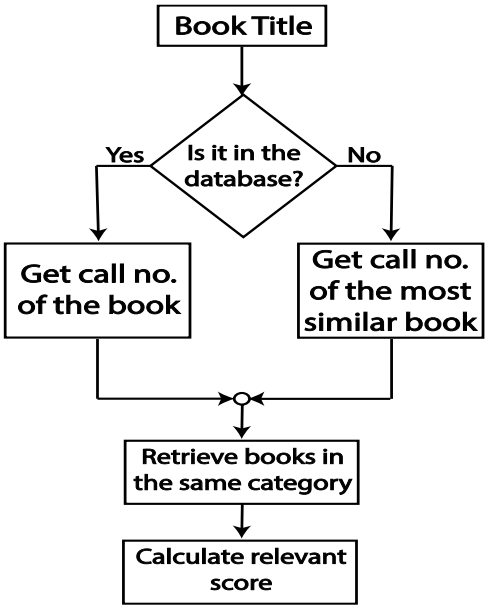

Figure 2: The Flow of Book Extension

1. The title of each book in BK_{w_0} is submitted to the book database.
    a. If the book is in the database, then its call number is extracted.
    b. If the book is not in the database, then the call number of the most similar book is extracted. (Presented in the next section is one title-based method for measuring how close two books are.)
2. Other books classified in the same category are retrieved. More precisely, other books whose call numbers are in the same groups of books obtained from Step 1 are retrieved. For example, the books whose call numbers are "QA371.N24 1996" and "QA371.B68 2013" are in the same category of QA371.
3. Finally, every obtained book is associated with an interesting score whose formula will be expressed latter.

A Title-based Similarity Measure

In this part, one similarity measure between two books using their own titles is presented. Given $T_1 = \{w_{11}, w_{12}, \dots, w_{1k}\}$, and $T_2 = \{w_{21}, w_{22}, \dots, w_{2m}\}$, are the sets of stems (root words),[3] excluding stop words (e.g. 'a', 'the', 'of'), from the titles of books B_1 and B_2, respectively. Based on the Jaccard's similarity measure (Jaccard 1901), the similarity degree of the two books is defined as:

[3] A stem is the form of a word before any inflectional affixes are added. For example, the words connect, connected, connecting, connections all can be stemmed to the word "connect".

$$Sim(B_1, B_2) = \frac{|T_1 \cap T_2|}{|T_1 \cup T_2|}. \quad (5)$$

Example 2.
To measure similarity between the books $B_1$and B_2 titled "Data Structures and Algorithms" and "Introduction to Algorithms", by the method explained above, we have

$$T_1 = \{Data, Structure, Algorithm\},$$
$$T_2 = \{Introduction, Algorithm\}.$$

Then,

$$Sim(B_1, B_2) = \frac{|\{\text{Algorithm}\}|}{|\{\text{Data, Structure, Algorithm, Introduction }\}|} = \frac{1}{4} = 0.25.$$

Book Interesting Score Calculation

Recall that $BK_{w_0} = \{(b_1, \beta_1), (b_2, \beta_2), \dots, (b_p, \beta_p)\}$, is the set of books and their interest scores with respect to the user's keyword w_0 when b_i and β_i are a book title and an interest score, respectively. By the book-extension process described above:

- If b_i is in the book database, then its interesting score is set as β_i.
- If b_i is not in the database and, among the books in the database, d is the closest book to b_i, then the interesting score of d (not b_i) is $sim(b_i, d)$, where the function sim is defined in Eq. (5).
- For each of the other books obtained from Step 2 of the process, its score is as equal as the score of its seed book.

Example 3
This section shows an example of the proposed framework. We use the book catalog provided in the library of Thammasat University. Suppose the user's keyword is "Laplace Transformation (LT)". To save the space, we collect a part of OntoMathPro related to the keyword 'LT' as shown in Figure 3. By Eq. (1), we have

- $Sim(\text{LaplaceTransformation, LaplaceTransformation}) = 1,$
- $Sim(\text{LaplaceTransformation, BorelTransformation}) = 0.75,$
- $Sim(\text{LaplaceTransformation, IntegralTransformation}) = 0.86,$
- $Sim(\text{LaplaceTransformation, LinearOperation}) = 0.57.$

If the predefined threshold is 0.7, then the word list generated by GWList is

$$LW_{LT}^{0.7} = \{(\text{LaplaceTransformation}, 1), (\text{ BorelTransformation}, 0.75),$$
$$(\text{LinearOperation}, 0.86)\ \}.$$

After searching through the course catalog of Department of Mathematics and Statistics, Thammasat University, four courses whose descriptions contain at least one term in $LW_{LT}^{0.7}$ are found and shown in Table 1,[4] where the last column depicts the relevant scores of the four courses.

[4] The third column shows only the terms in $LW_{LT}^{0.7}$ that appear in the course descriptions, not all topics in the course descriptions.

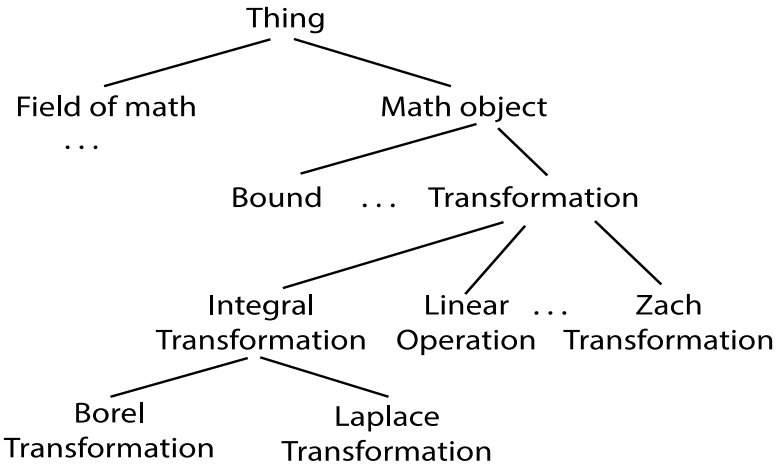

Figure 3: A Part of OntoMathPro for Example 3

Table 1: Courses containing related terms to the keyword "Laplace Transformation"

Course ID	Course Name	Covered Topics	Score
MA214	Differential Equations	Laplace transformation	1
MA313	Ordinary Differential Equations	Laplace transformation	1
MA318	Partial Differential Equations	Laplace transformation	1
MA 646	Applied Analysis	Integral transformation	0.86

From the course syllabi for the four relevant subjects, the details of the textbooks are shown in Table 2. Following GBList, all of the textbooks, except *Advanced Engineering Mathematics*, are available in the library, where the last column of the table presents the call numbers thereof.

Table 2: Textbooks for the Four Courses in Table 1

Course ID	Textbook Title	Author(s)	Call Number
MA214	Elementary Differential Equations and Boundary Value Problems	W. E. Boyce & R. C. DiPrima	QA371.B68 2013
	Advanced Engineering Mathematics	A. Jeffrey	Not Available
MA313	Fundamentals of Differential Equations	R. K. Nagle & E. B. Saff	QA371.N24 1996
	Elementary Differential Equations and Boundary Value Problems	W. E. Boyce & R. C. DiPrima	QA371.B68 2013
MA318	Elementary Applied Partial Differential Equations	R. Haberman	QA377.H3 1998
MA 646	Introduction to Functional Analysis with Applications	E. Kreyszig	QA320.K74 1989

By Eq. (5), the most similar textbooks to *Advanced Engineering Mathematics by A. Jeffrey* are *Advanced Engineering Mathematics by D. M. Greenberg* and *Advanced Engineering Mathematics by C. R. Wylie and L. C. Barrett.*, where the respective call numbers are TA330 .G6 1997 and QA401.W9 1995. It is clear that the similarity degree is 1 because of the identical title (In fact, there are more than two books, titled *Advanced Engineering Mathematics*, but they are classified into the groups TA330 or QA401). In Table 3, we refer each book by its call number instead of the title. Then, the set of books related to the keyword is retrived and denoted as BK_{LT}.

After searching books in the same classification of those in the set BK_{LT}, i.e. QA320, QA371, QA377, QA401, and TA330, are selected and evaluated. The results are revealed in Table 3 where the first column is the book classification, the second and the third show the numbers of Thai and English books in each group, and the last indicate the percentage of the books related to the user's keyword. To interpret, for example, 71 books are selected from those in group QA371 by our framework (26 of them are Thai books and 45 are English).

Table 3: Number of Retrieved Textbooks

Call no. group	Number of textbooks[5]	
	Thai	English
QA320	0	34
QA371	26	45
QA377	1	24
QA401	3	25
TA330	12	28

$$BK_{LT} = \{(\text{QA371. B68 2013}, 1), (\text{QA371 . N24 1996}, 1), (\text{QA371 . B68 2013}, 1), (\text{QA377 . H31998}, 1), (\text{QA320 . K74 1989}, 0.86), (\text{QA401 . W9 1995}, 1), (\text{TA330. G6 1997}, 1)\}.$$

Comparing to the traditional book searching, with the same keyword "Laplace Transformation", 4 items are returned (3 of them are English and 1 is Thai.) On further investigation, we found that: (i) One book is common with the results of our frameworks. (ii) Three books are in classification group QA432, which is not in the results of our frameworks.

EXPERIMENTAL RESULTS

In this section, more experiments are presented. Eight students enrolled in mathematical courses were selected. Each of them was asked to present one mathematical keyword. The eight terms are listed in the first column of Table 4. For our framework, the threshold α in Eq. (2) was set to 0.7 and books of which the interest scores α are greater than 0.8 were selected. In order to evaluate the performance of searching systems, we used the numbers of retrieved items and precision, which is the ratio of relevant items to retrieved items.

After applying the eight keywords to the OPAC system of Thammasat University (TU)[6] using the keyword-search option and to the proposed framework, the results is shown in Table 4 when 'OPAC' is referred to the results of the traditional researching system and 'OPAC+' is referred to those of our framework. Moreover, the last column of this table presents the numbers of common books between the two methods. One can see, for example, that the keyword "Laplace Transform" OPAC and OPAC+ returned 32 and 71 items, respectively. Among such retrieved items, 81.25 percent and 91.55 percent were classified as the relevant books. The two item sets contained 15 common books. The table reveals that, on the average, the number of retrieved items from OPAC are significantly less than those from OPAC+. By considering the precision values, when the cases of small retrieved books, i.e., "Cauchy residue theorem," "Differentiable function," "Hausdorff space," and "Projective

[5]We compare only two languages, i.e. Thai and English, because all mathematical books in our libraries are these languages.

[6]From the TU library database, there are about 1039 textbooks related to Mathematics.

plane," OPAC obtained 100 percent. For the other keywords, it obtained variant scores ranging from about 58 percent to 81 percent. In contrast, OPAC+ owned the relatively low variant scores ranging from about 81 percent to 93 percent. On the average, the precision of OPAC+ is slightly better than that of OPAC, i.e., 86.69 percent with 3.60 percent of standard deviation (SD) versus 84.33 percent with 16.82 percent of SD.

Table 4: Experimental Results

Keyword	**No. of retrieved books**		**Precision (%)**		No. of common books
	OPAC	**OPAC+**	**OPAC**	**OPAC+**	
Cauchy residue theorem	1	13	100.00	84.62	0
Differentiable function	1	21	100.00	85.71	1
Hausdorff space	1	14	100.00	92.86	1
Laplace transform	32	71	81.25	91.55	15
Lie algebra	11	27	63.64	85.19	6
Maclaurin series	7	42	71.43	80.95	5
Maxwell's equation	12	111	58.33	85.59	8
Projective plane	2	31	100.00	87.10	2
Average	**8**	**41**	**84.33**	**86.69**	**5**
Standard deviation			**16.82**	**3.60**	

To gain insight the results, the book interesting scores from OPAC+ were also investigated. For this examination, the retrieved books were ranked by descending order of their interesting scores. Table 5 shows the results of OPAC+ when the books ranked in the top N percent of their interesting scores were selected to users. The table reveals that our framework reached the 100 percent of precision for all queries when the top 10% and 20% items were selected. Not surprisingly, the more items are selected, the less precision scores are obtained. However, even N is increased up to 50, the precision degrees for all queries are satisfactory. The degrees are 98.08 percent, 92.89 percent, and 90.31 percent, when N's are 30, 40, and 50, respectively. Figure 4 shows the average precision over the eight keywords for each N. The graph indicates that when N is greater than 30 the precision relatively low.

Table 5: The Precision of OPAC+ when the Top N% of the Ranked Items were Retrieved.

Keyword	**Precision (%)**				
	N= 10	**N = 20**	**N = 30**	**N = 40**	**N = 50**
Cauchy residue theorem	100.00	100.00	100.00	100.00	85.71
Differentiable function	100.00	100.00	100.00	88.89	90.91
Hausdorff space	100.00	100.00	100.00	100.00	100.00
Laplace transform	100.00	100.00	100.00	93.10	91.67
Lie algebra	100.00	100.00	100.00	90.91	92.86
Maclaurin series	100.00	100.00	84.62	82.35	80.95
Maxwell's equation	100.00	100.00	100.00	95.56	92.86
Projective plane	100.00	100.00	100.00	92.31	87.50
Average	**100.00**	**100.00**	**98.08**	**92.89**	**90.31**
Standard deviation	**0.00**	**0.00**	**5.09**	**5.47**	**5.31**

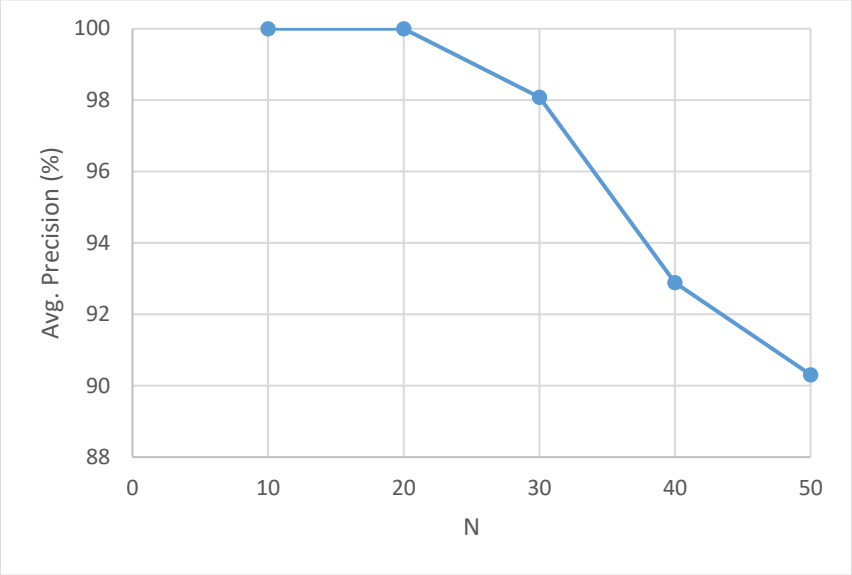

Figure 4: The Average Precision Value of each N in Table 4.

DISCUSSION

On further consideration to each module in the proposed framework we found that GWList can generate sets of keywords which are semantically close to the users' keywords. Unfortunately, the precision of our framework is lower than that of the traditional system (e.g. the cases of "Cauchy residue theorem" and "Differentiable function " in Table 4) resulting from GBList where the text books from couse syllabuses are not in the library database (see Figure 3). In the case, our system retrived the books based on their title. However, two textbooks, whose their titles have some common words, occasionally contain few common topics. Even this the module owns up to this low precision, it is helpful when the textbooks for courses relevant to the users' keywords are just published. For this situation, we suggest to filter the retrived book collection by selecting only books in top 10-30 percent ranking by relevant scores.

CONCLUSION

This work introduces a framework to improve the success rates of unknown-item search from library catalogs. There are three main components in the framework: (i) GWList for expanding a user's keyword to other related terms based on domain-specific ontology; (ii) GCList for selecting the courses whose contents relate to the expanding keyword set; and (iii) GBList for retrieving books relevant to the selected courses.

From the experiments, we notice that GWList can produce a set of keywords semantically relating to a user's keyword. By the keyword set, GCList is able to link the obtained keyword set to related courses. Finally, GBList can generate relevant books, since the precision values are high. Compared to OPAC, the proposed framework can retrieve relevant books better than the traditional keyword search in OPAC. Moreover, the numbers of returned book for our framework are also more than those for OPAC. It means that the introduced framework

provides both quality and quantity of returned books. In addition, the proposed interesting-score measurement can facilitate to lift the precision levels up (The more score is set, the more precision is obtain).

Even though the proposed framework can retrieve more relevant books, it has some limitations. In order to apply the framework in another domain, it requires a domain ontology which may not be available. To the best of our knowledge, there are few domains that ontologies relating to keywords in course descriptions of subjects about the domains are published. Those domains are biology, engineering, and chemistry. [7]

As future works, we will extend this framework by using other pieces of academic information such as taxonomy of course catalog. We expect that prerequisite course information facilitate in personal book suggestion. More experiments on users' satisfaction will also be taken into account. To apply the proposed framework into the traditional search system, we will implement a web application on-top the OPAC system. We will embed GWList, GCList, and GBList into the web application. A primitive list of books from course syllabuses—the output of the implemented application—will be automatically passed to the OPAC system for searching more interesting books in the library database. Finally, the obtained books will be ranked by using the proposed interesting score measure.

ACKNOWLEDGEMENT

This research is supported by Faculty of Science and Technology Fund, Thammasat University, Contract No. 15-08-2559.

REFERENCES

Antell, K. and Huang, J. 1998. Subject Searching Success Transaction Logs, Patron Perceptions, and Implications for Library Instruction. *Information Storage and Retrieval,* Vol. 48, no. 1: 69-76.

Batet, M., Sánchez, D., and Valls, A. 2011. An Ontology-Based Measure to Compute Semantic Similarity in Biomedicine. *Journal of Biomedical Informatics,* Vol. 44, no. 1: 118-125.

Breeding, M. 2007. Next-generation library catalogs. *Library Technology,* Vol. 43, no. 4: 5-14.

Castro, L. J., Giraldo, O. X., and Castro, A. G. 2010. Using the Annotation Ontology in Semantic Digital Libraries. *Proceedings of the 2010 International Conference on Posters & Demonstrations Track*, (pp. 153-156). Shanghai, China.

Cousins, S. A. 1992. Enhancing Subject Access to OPACS: Controlled Vocabulary VS Natural Language. *Journal of Documentation,* Vol. *48*, no. 3: 291-309.

[7] We search through several ontology portals e.g. https://onki.fi/en/browser and https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library.

Hessel, H. and Fransen, J. 2012. Resource discovery: comparative survey results on two catalog interfaces. *Information Technology & Librarie,* Vol. 31, no. 2: 21-44.

Jaccard, P. 1901. Étude comparative de la distribuition florale dans une portion des Alpes et des Jura. *Bulletin de la Société vaudoise des sciences naturelles,* Vol. *37*, no. 142: 547-579.

Khan, S. and Bhatti, R. 2017. Semantic Web and ontology-based applications for digital libraries: An investigation from LIS professionals in Pakistan. *The Electronic Library*. doi:10.1108/EL-08-2017-0168

Long, C. E. 2000. Improving Subject Searching in Web-Based OPACs. *Journal Of Internet Cataloging,* Vol. 2, no. 3-4: 158-186.

Martell, C. 2008. The Absent User: Physical Use of Academic Library Collections and Services Continues to Decline 1995–2006. *The Journal of Academic Librarianship,* Vol. 34, no. 5: 400-407.

Nevzorova, O., Zhiltsov, N., Kirillovich, A., and Lipachev, E. 2014. OntoMathPro Ontology: A Linked Data Hub for Mathematics. *Lecture Notes in Computer Science (Knowledge Engineering and the Semantic Web),* Vol. 468: 105-119.

Noah, S. A., Alias, N. A., Osman, N. A., Abdullah, Z., Omar, N., Yahya, Y., and Yusof, M. M. 2010. Ontology-driven semantic digital library. *Lecture Notes in Computer Science (Asia Information Retrieval Societies Conference),* Vol. 6458: 141-150.

O'Neill, E. T., Bennett, R., and Kammerer, K. 2014. Using Authorities to Improve Subject Searches. *Cataloging & Classification Quarterly,* Vol. 52, no.1: 6-19.

Rondeau, W. 2013. Making Use of Existing Tools for Unknown Item Needs: Improving Subject Retrieval in Online Catalogues. *Library Literature & Information Science,* Vol. 59, no. 4: 30-32.

Slone, D. J. 2000. Encounters with the OPAC: On-Line Searchingin Public Libraries. *Journal of The American Society for Information Science,* Vol. 51, no. 8: 757-773.

Wakeling, S., Clough, P., Silipigni Connaway, L., Sen, B., and Tomás, D. 2017. Users and uses of a global union catalog: A mixed-methods study of WorldCat. org. *Journal of the Association for Information Science and Technology,* Vol. 68, no. 9: 2166-2181.

Wood, J. A., Smigielski, E. M., and Haynes, G. 2007. Case-based approach for improving student MEDLINE searches. *Medical Education,* Vol. 41, no. 5: 510-511.

Zaid, N., and Lau, S. 2014. Emerging of Academic Information search system with ontology-base approach. *Procedia-Social and Behavioral Sciences,* Vol. 116: 132-138.