Improving Library Searches Using Word-Correlation Factors and Folksonomies

Maria Soledad Pera
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Computer Sciences Commons

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/etd/1972

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Improving Library Searches Using Word-Correlation Factors and Folksonomies

by

Maria Soledad Pera

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computer Science

Brigham Young University

April 2009
This thesis has been read by each member of the following graduate committee and by the majority vote has been found to be satisfactory.

Date
Dr. Yiu-Kai Dennis Ng, Chair

Date
Dr. Christophe Giraud-Carrier

Date
Dr. Scott Woodfield
As chair of the candidate’s graduate committee, I have read the thesis of Maria Soledad Pera in its final form and have found that (1) its format, citations and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Dr. Yiu-Kai Dennis Ng
Chair, Graduate Committee

Accepted for the Department

Dr. Kent Seamons
Graduate Coordinator

Accepted for the College

Dr. Thomas W. Sederberg
Associate Dean, College of Physical and Mathematical Sciences
ABSTRACT

Improving Library Searches Using Word-Correlation Factors and Folksonomies

Maria Soledad Pera
Department of Computer Science
Master of Science

Libraries, private and public, offer valuable resources to library patrons; however, formulating library queries to retrieve relevant results can be difficult. This occurs because when using a library catalog for library searches, patrons often do not know the exact keywords to be included in a query that match the rigid subject terms (chosen by the Library of Congress) or terms in other fields of a desired library catalog record. These improperly formulated queries often translate into a high percentage of failed searches that retrieve irrelevant results or no results at all. This explains why frustrated library patrons nowadays rely on Web search engines to perform their searches first, and upon obtaining the initial information, such as titles, subject areas, or authors, they query the library catalog. This searching strategy is an evidence of failure of today’s library systems. In solving this problem, we propose an enhanced library system, called EnLibS, which allows partial, similarity matching of (i) tags defined by ordinary users at a folksonomy site which describe the content of books and (ii) keywords in a library query to improve the searches on library catalogs. The proposed library system allows patrons to post a query Q with commonly-used words and ranks the retrieved results according to their degrees of resemblance with Q. Experimental results show that EnLibS
(i) reduces the amount of queries that retrieve no results, (ii) obtains high precision in retrieving and accuracy in ranking relevant results, and (iii) achieves a processing time comparable to existing library catalog search engines.
ACKNOWLEDGMENTS

I want to thank Dr. Christophe Giraud-Carrier for his feedback and suggestions for improving this work. I would also like to thank Dr. Giraud-Carrier and Dr. Scott Woodfield for their willingness to serve as my committee members.

I also want to thank the CS 705 class (Sabra, David and Dr. Embley) for their comments and help in "pithifying" my thesis.

Thanks go also to Bill Lund, for his disposition and help in gathering material from the Harold B. Lee Library.

My sincere appreciation to my mom, JC, the Cutri and Bollinger families, Koko, and Na, for their support and for being there for me every step of the way.

And of course, I want to thank my advisor, Dr. Yiu-Kai Ng, for being an incredible mentor, for his encouragement, patience and help in completing this work and throughout my time at BYU.

To my Yaya.

Sole

December 2008
CONTENTS

INTRODUCTION ......................................................................................................................... 1

RELATED WORK ........................................................................................................................ 5

OUR ENHANCED LIBRARY SEARCH APPROACH ................................................................. 9

3.1 Word Similarity .................................................................................................................. 9
  3.1.1 Word-Correlation Factors ......................................................................................... 10
  3.1.2 Database Records ..................................................................................................... 11

3.2 Using LibraryThing Tags as Document Representation .................................................. 12
  3.2.1 LibraryThing Tags .................................................................................................... 13
  3.2.2 Reduced LibraryThing Tags and Corresponding Tables .......................................... 16

3.3 Subset of Relevant Records .............................................................................................. 17

3.4 Relevance Ranking .......................................................................................................... 23

3.5 Query Processing Time .................................................................................................... 27
  3.5.1 Substring Prefix ........................................................................................................ 28
  3.5.2 Query Processing Time/Memory Allocation for Indexing ........................................ 29

3.6 The Overall Evaluation Process ....................................................................................... 32

EXPERIMENTAL RESULTS ..................................................................................................... 33

4.1 The Dataset ....................................................................................................................... 34

4.2 Controlled Experiments .................................................................................................. 34
  4.2.1 The Cross-over Experiment .................................................................................... 36
  4.2.2 The Number of Participants ................................................................................... 36
  4.2.3 Independent Appraisers ......................................................................................... 39
  4.2.4 Number of Queries ................................................................................................. 41
  4.2.5 The Evaluation Tasks .............................................................................................. 42
4.3 Evaluation Measures .............................................................................................................. 43

4.3.1 Precision .............................................................................................................................. 43

4.3.2 Mean Average Precision ..................................................................................................... 44

4.3.3 Mean Reciprocal Rank ........................................................................................................ 46

4.3.4 Spearman Rank Correlation Coefficient ........................................................................... 47

4.4 Performance Evaluations ........................................................................................................ 50

4.4.1 Queries with Zero-hits ...................................................................................................... 51

4.4.2 Relevance of the Retrieved Results ................................................................................... 52

4.4.3 Ranking of Retrieved Records .......................................................................................... 56

4.4.4 Query Processing Time ...................................................................................................... 59

4.5 Overall Performance of EnLibS ............................................................................................ 60

4.6 Implementation ....................................................................................................................... 61

4.7 Impacts of EnLibS ................................................................................................................... 61

CONCLUSIONS .............................................................................................................................. 63

APPENDIX A .................................................................................................................................... 65

APPENDIX B .................................................................................................................................... 67

APPENDIX C .................................................................................................................................. 73

REFERENCES .................................................................................................................................. 76
List of Figures

Figure 1. HBLL library catalog record $R$ for the book “Mt. McKinley: the Pioneer Climbs” .................................................................................................................................................................................. 13

Figure 2. Information on the book “Mt. McKinley: the Pioneer Climbs” as shown in the corresponding LibraryThing record .................................................................................................................................................................................. 15

Figure 3. Tags and additional information created by different LibraryThing users for the book “Mt. McKinley: the Pioneer Climbs” .................................................................................................................................................................................. 16

Figure 4. Distribution of the word-correlation values in the 13% matrix .............................................................................................................................................................................................................. 20

Figure 5. Number of word pairs in each 13% matrix and the average processing time of queries in HBLL-set .................................................................................................................................................................................. 20

Figure 6. Required memory allocation for prefix indexes and their average processing time on the queries on the HBLL-set .................................................................................................................................................................................. 30

Figure 7. Average processing time in answering queries in HBLL-set using different data and indexing tables .................................................................................................................................................................................. 31

Figure 8. The overall query evaluation process of EnLibS ............................................................................................................................................................................................................... 32

Figure 9. Educational backgrounds of the appraisers who participated in the empirical study ......................................................................................................................................................................................................... 40

Figure 10. Academic majors of the appraisers who participated in our study ........................................................................................................................................................................................................ 40

Figure 11. Percentage of male/female appraisers ............................................................................................................................................................................................................................................. 41

Figure 12. Zero-hits queries in HBLL-log generated by the HBLL system versus EnLibS 52
Figure 13. Average P@1 computed for each of the six queries evaluated by each group in our study based on the top relevant library records retrieved by EnLibS and the HBLL system on each of the test queries, respectively .......................................................... 53

Figure 14. The average 10-Precision values computed for the test queries in our study based on the library records retrieved by EnLibS and the HBLL system, respectively .... 54

Figure 15. Average MRR computed by using the top relevant library records retrieved by EnLibS and the HBLL system for each test query ................................................................. 55

Figure 16. The MAP(r) values for r ∈ {3, 5, 7, 10} based on the query results retrieved by EnLibS and the HBLL system, respectively ................................................................. 56

Figure 17. The Spearman Rank Correlation Coefficient computed for each of the test queries in our study using EnLibS and the HBLL system, respectively ..................... 59

Figure 18. Query processing time for the queries in HBLL-log using EnLibS and the HBLL system ........................................................................................................ 60

Figure 19. An evaluation form used by the (2nd) group(s) of appraisers in our controlled experiment ........................................................................................................... 68
List of Tables

Table 1. The word-correlation factors of the keywords in the query Q: “Climb Alaska”, the title, and (portion of the) subject terms of the library record R shown in Figure 1. ................................. 13

Table 2. A subset of library patron’s queries in HBLL-set and their stemmed, non-stop word versions used for experimentation .......................................................... 21

Table 3. Words that are similar to the query keywords in Q: “Climb Alaska” in the CoM35-13 matrix ........................................................................................................ 22

Table 4. Words that are similar to the query keywords in Q: “Climb Alaska” in the CoM36-13 matrix ........................................................................................................ 22

Table 5. Library records retrieved using the reduced matrices CoM35-13 and CoM36-13, respectively for the query “Climb Alaska” ........................................................................ 23

Table 6. Ten of the catalog records ranked for query Q: “Climb Alaska” using Sim ...... 26

Table 7. Top 10 catalog records ranked for query Q: “Climb Alaska” using LimitedSim. 27

Table 8. Titles of the ten ranked library catalog records shown in Tables 6 and 7 .......... 27

Table 9. Size (in MB) of different indexed tables ........................................................................ 30

Table 10. Different types of queries in HBLL-log as well as the number of zero-hits queries ................................................................................................................................. 35

Table 11. (Some of the) Subject areas and their corresponding number of queries in HBLL-log .................................................................................................................... 35

Table 12. Library records ranked by EnLibS and the average ranking generated by using the rankings created by the appraisers in Group 1 for Query 2 “Disney” along with the (squared) difference among the relative positions ................................................. 50
Table 13. Relevance and ranking evaluations on the library records retrieved for the query “Child Development” by EnLibS that were generated by two different appraisers in Group 1 ............................................................................................... 59

Table 14. Top 10 library records retrieved by EnLibS for the query "African culture" ........................ 70

Table 15. Top 10 library records retrieved by the HBLL system for the query "African culture" ........................................................................................................................................ 72

Table 16. Average ranking provided by the appraisers in Group 1 ............................................. 74

Table 17. Average ranking provided by the appraisers in Group 2 ............................................. 74

Table 18. Average ranking provided by the appraisers in Group 3 ............................................. 75

Table 19. Average number of relevant library records retrieved by EnLibS and the HBLL system, respectively according to the independent appraisers in our study ........................................ 75
Chapter 1

Introduction

Libraries, private and public, provide valuable sources of information, from century-old to the latest publications, which include journals, newspapers, textbooks, and (non-) fiction books in many different languages, in addition to maps, audio and video scripts, etc. Library patrons from different age groups and educational background, with diverse information needs, turn to libraries to locate information through library catalogs. Library catalogs, ranging from the form of card catalogs in the nineteenth century to the digital version used nowadays, have been a place to start searching for information, since they contain essential data (such as title, authors, subject headings, etc.) of each library resource, e.g., books, maps, periodicals, etc. It is imperative to know that materials that are archived only at libraries are not accessible by simply querying Web search engines, and thus the only way is to consult the corresponding library catalog. The library catalog, however, is defined by controlled vocabularies, subject guides, and broad categories rather than commonly-used words, which are rigid and unintuitive to use by ordinary library patrons [Rethlefsen 2007]. As a result, library patrons looking for needed materials available in a library through the library catalog often encounter discouraging results. The problem library patrons must deal with is not the lack of resources, but the inflexibility in locating them through the library catalog. Learning to use the library catalog, however, is a tedious and time-consuming process.

1 The Library of Congress (http://catalog.loc.gov/help/contents.htm) defines the library catalog as a database of records that describe the collection of materials held by a library.
Thus, library patrons who demand easy-to-formulate queries and relevant information to be retrieved in a timely manner look elsewhere to satisfy their information needs. Recently, a common practice adopted by library patrons is to first utilize Web searching tools, such as Google and Google Scholar, in locating the primitive information, such as author, titles, and subject areas, before looking up library materials using library catalogs [Herrera 2007]. These users’ behavior and expectations are influenced by their Web search experiences [Li 2008].

The inadequate online library catalog has been the focal point of criticism [Larson 1991] and is the cause of the library-search phenomenon discussed above. Even after more than a decade, the major design faults of online library catalogs still exist. These design problems include (i) the lack of user’s understanding about the LCSH\(^3\), (ii) difficulty in properly formulating queries using the library catalog (as discussed earlier) compared with using simple keyword queries, (iii) information overload, i.e., searches that return too many results, (iv) search failures, i.e., searches that return no results at all, and (v) irrelevant searches, i.e., searches that return records that do not match the user’s needs. Along with these problems, the lack of relevance ranking on retrieved results is another major concern of library patrons [Novotny 2004]. Even though solutions to problems (iii) - (v) have been integrated into the design of current Web search engines, such as Google and Yahoo, none of the problems (i) - (v) has been thoroughly addressed for improving

\(^2\) [De Rosa 2005] show that 89% of library patrons start their searches using Web search engines and then query library catalogs.

\(^3\) LCSH (Library of Congress Subject Heading) are words or phrases standardized by the US Library of Congress that are adapted for describing library resources [Inouye 2001].
library catalog searches [Yu 2004].

Realizing all of these design faults, we have developed an enhanced library system, called EnLibS, which handles all of the design problems listed above. EnLibS focuses on retrieving information using search engines designed for academic institutes, such as university library systems, and it allows novice, as well as expert, users to quickly search for desired information without requiring special skills and advanced training in using the library catalog. EnLibS adapts Information Retrieval (IR) techniques, such as the Fuzzy Set IR Model [Baeza-Yates 1999], which extends the traditional IR models to establish the degree of similarity among (key)words in a query and in a document (i.e., library catalog record in our work). Furthermore, in processing a library query EnLibS consults “folksonomies,” which have been used in Flickr (http://www.flickr.com), YouTube (http://www.youtube.com), and Del.icio.us (http://delicious.com), to effectively describe and identify pictures, videos, and Websites, respectively. Folksonomies, which is also known as social classification or social tagging [Neal 2007], are keywords that describe a particular item, e.g., a picture, book, or video. These keywords, which are not selected from a controlled vocabulary nor a pre-defined taxonomy, are determined by ordinary users, i.e., people who use Flickr, YouTube, LibraryThing, etc.

The design goals of EnLibS include (i) reducing the high percentage of failed catalog searches, (ii) retrieving highly relevant results, (iii) ranking the retrieved results according to their degrees of relevance to the corresponding query, rather than simply ordering the results by the date of publication, date a particular record was added to the
library catalog, or the order in which they are retrieved, (iv) allowing users to perform library catalog searches by using simple keyword queries, and (v) maintaining the query processing time comparable with the processing speed of current library search engines.

To determine the efficiency of our enhanced library system, we have conducted two different performance evaluations. We first evaluate the effectiveness of EnLibS using the Harold B. Lee Library (HBLL) transaction logs (from July 2006 to January 2007), called HBLL query log, in (i) reducing the number of failed searches and (ii) determining the percentage of no-result searches generated by the Harold B. Lee Library system (HBLL system for short) compared with the ones generated by EnLibS. We also compare the HBLL system and EnLibS in terms of the computational time required to retrieve (relevant) results using the HBLL query log. Hereafter, we conduct a controlled experiment and analyze the overall performance of EnLibS in terms of its degree of accuracy in retrieving and ranking relevant results.

The remaining chapters are organized as follows. In Chapter 2, we discuss various reasons that affect the quality of library catalog searches. In Chapter 3, we describe in detail the design of EnLibS for locating relevant library records using the library catalog. In Chapter 4, we present the experimental results, which verify the accuracy of our library catalog search approach in enhancing the quality of the retrieved results and reducing the number of no-result queries when using the library catalog. In Chapter 5, we give a conclusion and include future directions for our work.
Chapter 2

Related Work

In designing an online library system, developers have considered design issues that include (i) representing and formatting user’s queries, (ii) processing the queries, and (iii) presenting query results. In this section, we discuss existing design problems that library systems are dealing with in terms of (i) performing online searches to retrieve relevant information from library catalogs, (ii) dealing with failed library searches, (iii) ranking retrieved results, and (iv) evaluating the performance of a library system.

As demonstrated by various studies [Larson 1992], [Yu 2004], and [Lau 2006], between 10% to 50% of online library catalog searches retrieve no results. This high percentage of empty results is mostly due to user’s inability in properly expressing Boolean queries (i.e., unaware of the existence of and unable to properly expressing a query using the logical Boolean operators AND, OR, and NOT) and most importantly due to their lack of understanding in using the Library of Congress Subject Headings to perform a search [Larson 1992]. According to Larson et al. [Larson 1992], the use of Boolean queries and LCSH are two of the most common, yet not properly addressed, problems in performing a library search to retrieve information from a library catalog. With respect to Boolean searches, Lau et. al., [Lau 2006] indicate that a majority of the existing online library systems, including the Nanyang Technological University (NTU) online public access catalog, offer the option of using Boolean operators when performing a search; however, this feature is considered to be rather complex, and as a result it is not frequently used.
by library patrons [Mead 2000]. Furthermore, not all the existing online library systems, including MELVYL, the University of California’s online catalog system [Larson 1991], offer a cross-reference to actual LCSH terms to aid users in locating the proper terms to perform a catalog search. Thus, users are required to either know the controlled vocabularies provided by the Library of Congress or deal with failures at the time to perform a search.

Another common challenge shared by most of the existing online library systems is the exact keyword matching which requires keywords as specified (in specific fields such as title, author, etc.) in the machine readable card (MARC4) associated with a particular library record to be matched exactly with the keywords in a user query in order to locate relevant library catalog records [Larson 1991]. Unfortunately, the exact match constraint affects the quality and quantity of retrieved results, i.e., irrelevant results or no results at all.

As opposed to most of the library systems that return chronologically ordered results to users, the Endeca-powered catalog at the North Caroline State University Library [Antelman 2006] ranks retrieved results of a query $Q$ by relevance. The ranking strategy (i) places higher the results that share exact keywords in $Q$ and in the catalog records, (ii) assigns more weight to results in which the keywords in the query match the keywords in the title rather than other fields, i.e., author, subject, etc., and (iii) uses the term frequency/inverse document frequency (TF-IDF) of keywords in $Q$.

---

4 MARC (http://www.loc.gov/marc/faq.html#definition) is a data format defined by the Library of Congress which allows exchanging, using, and interpreting bibliographic information among computers.
Other library catalogs, such as the one at the University of Toronto, offer the users the choice of performing title searching (i.e., considering only the title of each record in a library catalog) in a keyword or browse mode. In the keyword mode, titles that contain the keywords in a query are retrieved, which according to [Arsenault 2007] can be inadequate when only one keyword is used, since many irrelevant results could be retrieved. In the browse mode, only the catalog records that have a title that exactly begins with the keywords in the query are retrieved, which can be problematic if the user is not aware of whether stopwords (see definition in Footnote 5) are removed by the searching algorithm, which is not always the case [Arsenault 2007].

A viewpoint in evaluating the performance of a library system is that neither non-empty results nor “zero-hits” results, i.e., no returned results, are solid indicators of a successful or failed catalog search. According to [Cooper 2001], a catalog search can be considered a success not only by conducting controlled experiments or by considering whether library patrons save, print, email, or download a retrieved record, but also by analyzing the behavior of a patron during a particular session. By performing this analysis, such as the time a user spends examining a particular record, the length of a session, the number of (title/subject) searches performed in a session, [Cooper 2001] determine the percentage of sessions based on the 905,970 sessions from the University of California’s Melvyl online catalog used for experimentation to be considered successful in using a particular library system. Farajpahlou et al., [Farajpahlou 1999] claim that the success of a library system should not be determined by the success in the performed searches only. Instead, other features, such as the simplicity of the system, the
response rate, and the ability to coexist with other library processes, should also be considered. In the study, [Farajpahlou 1999] propose to use a 26-item scale criteria for measuring the success of a library system. Although this measure appears to be valid and reliable, a larger number of library systems must be evaluated using the proposed measure before its applicability can be confirmed.

Unlike the searching and ranking approaches mentioned above, we intend to demonstrate that by (i) allowing (in)exact word matches, (ii) detecting semantically similar keywords, and (iii) using representative keywords (e.g., LibraryThing tags) as opposed to LCSH to describe a book, EnLibS can reduce almost completely, the relatively high percentage of searches that generate no result or irrelevant results and improve the quality of the retrieved results.
Chapter 3

Our Enhanced Library Search Approach

In this chapter, we detail the design of EnLibS and present the pre-processing and evaluation strategy of EnLibS in answering library patron’s queries.

3.1 Word Similarity

During the process of evaluating a library patron’s query $Q$, we determine the degree of resemblance of $Q$ and the representation of a library catalog record $R$, which is calculated by using the pre-computed degrees of similarity among the keywords in $Q$ and $R$. These degrees of similarity, which are the word-correlation factors in the word-correlation matrix $M$ [Koberstein 2006], were generated by using a set of approximately 880,000 Wikipedia documents (downloaded from http://www.wikipedia.org/), and each factor indicates the degree of similarity of two words\(^5\) based on their (i) frequency of co-occurrence and (ii) relative distance in each Wikipedia document. Wikipedia documents were chosen to construct $M$, since they were written by more than 89,000 authors with different writing styles and terminology that cover a wide range of topics. Thus, the Wikipedia documents are diverse in word usage and content. Furthermore, the words in $M$ are common words in the English language that appear in various online English dictionaries, such as 12dicts-4.0 (http://prdownloads.sourceforge.net/wordlist/12dicts-

---

\(^5\) Words in the Wikipedia documents were stemmed (i.e., reduced to their grammatical roots) after all the stop words, i.e., words with little meaning such as articles, conjunctions, prepositions, etc., were removed which minimize the number of (key)words to be considered. From now on, unless stated otherwise, (key)words refer to non-stop, stemmed words.
4.0.zip), Ispell (http://www.cs.ucla.edu/geoff/ispell.html), and BigDict (http://packetstormsecurity.nl/Crackers/bigdict.gz).

3.1.1 Word-Correlation Factors

The word-correlation matrix is a 57,908 x 57,908 symmetric matrix, since its word-correlation factors $C(i, j)$ and $C(j, i)$ are equal, where $i$ and $j$ are any two given words, and $C(i, j)$ reflects how closely related $i$ and $j$ are, and is defined as

$$C(i, j) = \frac{\sum_{w_i \in V(i)} \sum_{w_j \in V(j)} \frac{1}{d(w_i, w_j)}}{|V(i)| \times |V(j)|}$$

where

- $d(w_i, w_j)$ denotes the distance (i.e., the number of words in) between $w_i$ and $w_j$ plus one;
- $V(i)$ ($V(j)$, respectively) denotes the set of words that includes $i$ ($j$, respectively) and its stem variations;
- $|V(i)| \times |V(j)|$ is the normalization factor.

Compared with synonyms and related words compiled by WordNet (http://wordnet.princeton.edu), in which pairs of words are not assigned similarity weights, word-correlation factors provide a more accurate measure of word similarity that are computed by the appearance of any two words in a huge set of documents.

Due to the size of the word-correlation matrix $M$, which sums up to 6.0 GB, accessing such a huge matrix for computing the degree of resemblance of a query $Q$ and the representation of a catalog record $R$ using the word-correlation factors could increase
the processing time of $Q$. Hence, we consider a reduced version of $M$, which contains 13% of the most frequently-occurring words (based on their frequencies of occurrence in the Wikipedia documents), and for the remaining 87% of the less-frequently-occurring words only exact-matched correlation factors (i.e., 1.0) are considered. The reduced word-correlation matrix is further minimized to yield the $\text{CoM}57-13$ matrix, which retains in the 13% matrix those pairs of words that have a correlation value of at least $5 \times 10^{-7}$ [Gustafson 08]. The further reduced matrix does not affect the accuracy of computing the degree of resemblance of $Q$ and $R$, since it contains the top 7,300 most frequently-occurred words that appear in 90% of the Wikipedia documents. In the $\text{CoM}57-13$ matrix, except for the word-correlation factors that reflect an exact match, i.e., factors assigned the value of 1.0, the remaining word-correlation factors in the matrix are in the range of $10^{-4}$ and $10^{-9}$. (See Section 3.3 for an in-depth discussion on the values of the word-correlation factors in the reduced 13% matrix.)

3.1.2 Database Records

In order to facilitate the storage structure and query processing techniques offered by existing relational database management systems (RDBMSs), such as query optimization, query execution, scalability, and indexing, we convert the $\text{CoM}57-13$ word-correlation matrix into a table, called $\text{correlation}57$, in MySQL, which is a three column, 25 MB table that consists of 688,994 tuples. Each tuple is of the form $<w_1, w_2,$

---

*From now on, the notation $\text{CoMYZ-13}$ denotes a reduced 13% matrix in which the word-correlation factors are $Y \times 10^{-Z}$ or higher.*
corrValue>, where \( w_1 \) and \( w_2 \) are words, and \( corrValue \) is the correlation factor of \( w_1 \) and \( w_2 \).

In correlation, \( w_1 \) and \( w_2 \), which compose the primary key, are ordered alphabetically.

Example 1. Using the HBLL system (http://catalog.lib.byu.edu/) to create the query \( Q: \) “Climb Alaska” and perform a search against its library catalog, we retrieved no results. Figure 1 shows the HBLL catalog record \( R \) for the book “Mt. McKinley: the Pioneer Climbs”, which is one of the records that should have been retrieved with respect to \( Q \), since \( R \) describes climbing experiences in Mt. McKinley in Alaska. However, due to the “exact matching” evaluation criteria, the book is not retrieved by the HBLL system, which is a major design fault of the library system, as well as other existing library systems. Table 1 shows the word-correlation factors among the keywords in \( Q \) and some of the keywords that appear in the title and subject term of \( R \). Clearly, the non-zero word-correlation factors indicate that keywords in \( Q \) are related to most of the keywords in \( R \). Thus, considering the correlation factors of the words in \( Q \) and \( R \), as opposed to exact matches only, it is anticipated that more relevant library catalog records can be retrieved with respect to \( Q \).

3.2 Using LibraryThing Tags as Document Representation

Instead of considering the keywords in the LCSH of a library catalog record \( R \) in evaluating a library patron’s query \( Q \), we use existing folksonomies, which include collaborative tagging that describes the content of a given object, such as a Web page, a picture, a book, etc. We have chosen the folksonomies defined in LibraryThing
(http://www.librarything.com), since to the best of our knowledge LibraryThing is the most popular social application that was set up solely for cataloging books.

Figure 1. HBLL library catalog record R for the book “Mt. McKinley: the Pioneer Climbs”

<table>
<thead>
<tr>
<th></th>
<th>Alaska</th>
<th>climb</th>
<th>history</th>
<th>McKinley</th>
<th>mount</th>
<th>mountaineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>climb</td>
<td>7.5x10^-5</td>
<td>1.0</td>
<td>9.3x10^-3</td>
<td>0.0</td>
<td>6.2x10^-5</td>
<td>2.5x10^-6</td>
</tr>
<tr>
<td>Alaska</td>
<td>1.0</td>
<td>7.5x10^-5</td>
<td>3.7x10^-3</td>
<td>0.0</td>
<td>4.5x10^-5</td>
<td>1.4x10^-5</td>
</tr>
</tbody>
</table>

Table 1. The word-correlation factors of the keywords in the query Q: “Climb Alaska”, the title, and (portion of the) subject terms of the library record R shown in Figure 1

As of November 2008, LibraryThing archives 3,792,392 unique records (or books), and approximately 550,000 users have added more than 42 million tags to different book records at LibraryThing, according to the Zeitgeist Overview (http://www.librarything.com/zeitgeist), which provides official statistical data of LibraryThing.

3.2.1 LibraryThing Tags

LibraryThing was founded in 2006 for aiding users in cataloging and referencing books.
A LibraryThing user can create an account for rating and reviewing books, as well as adding labels, i.e., *tags*, which describe the content of books in his/her online personal library catalog. By looking up tags in LibraryThing, a library patron can locate books using *commonly-used* and *intuitive* words (i.e., words included in a simple and unrestricted vocabulary), rather than the rigidly controlled vocabulary used in LCSH. Besides serving as a robust cataloging tool, LibraryThing provides a mean of communication among users to share personal library catalogs and/or discuss the content of different books, in addition to making book recommendations to others. Furthermore, LibraryThing uses collective intelligence strategies to suggest books that may (not) be of interest to the users [Starr 2007].

Recall that Figure 1 shows the HBLL catalog record for the book “Mt. McKinley: the Pioneer Climbs”, whereas Figure 2 shows the corresponding LibraryThing record on the book. While both records share common information, which include title, author, and subjects (terms), the LibraryThing record incorporates additional information (as shown in Figure 3), such as reviews, rating, book recommendations, and most importantly a set of *tags* and their respective frequency counts. (Each count of a tag is the total number of users who suggested the tag after reviewing the corresponding book.)
Figure 2. Information on the book “Mt. McKinley: the Pioneer Climbs” as shown in the corresponding LibraryThing record

Since LibraryThing imposes no restriction on the number of tags that can be used to describe a particular book, the number of tags assigned to a given book ranges from 1 to thousands. However, a significant portion of those tags are (i) personalized and used as a reminder, such as “read,” “want to read,” or “borrowed,” as well as (ii) stop words, which provide little meaning in identifying a library catalog record. These tags are not considered during query processing. Eventually, we use a subset of LibraryThing tags (that are selectively chosen according to their frequency of occurrence as discussed in Section 3.2.2) to identify each book in a library catalog.
Figure 3. Tags and additional information created by different LibraryThing users for the book “Mt. McKinley: the Pioneer Climbs”

3.2.2 Reduced LibraryThing Tags and Corresponding Tables

Due to the huge number of tags available on LibraryThing, i.e., 42,793,821 in total as of November 19, 2008, we reduce the number of tags to be considered during the query evaluation process by choosing only the top-\(n\) \((n \geq 1)\) tags describing a particular book \(B\), where \(n\) is the top \(n^{th}\) frequency counts of tags for \(B\) in LibraryThing. The ideal number of \(n\) can be defined according to (i) its accuracy in retrieving relevant library records as well as (ii) the processing time in establishing the degree of resemblance between a query and a library catalog record. In determining the proper value for \(n\), we have conducted an empirical study on a range of different values for \(n\) and chose the one that satisfies the
two criteria listed above. As shown by the conducted experiments (see Chapter 4), the appropriate value for \( n \) is three.

We include in our MySQL database the *tags* that identify the content of each library catalog record by creating a table `idtag(id, tag)`, where `id` is a unique identifier of a particular library record \( R \), and `tag` is one of the top *three* non-stop, stemmed, non-personalized keywords that represent the content of \( R \). `Idtag` is ordered by `id` number, which facilitates the process of locating the descriptive data, i.e., title and author, of a library catalog record\(^7\), which can be retrieved from another MySQL table `catalog(id, title, author)`, where `id` is as defined in the `idtag` table. The `idtag` table is 107 MB in size and contains 1,036,708 tuples.

### 3.3 Subset of Relevant Records

The `catalog` table, which contains the library catalog records that match the records found in LibraryThing, can be huge. (As of the year 2007, the HBLL catalog includes 521,517 unique records that also appear in LibraryThing.) It is impractical to evaluate each catalog record against a library patron’s query \( Q \) sequentially. Thus, prior to computing the degrees of resemblance between \( Q \) and the catalog records, each of which is represented by the *three* most frequent, user-recommended tags, we choose a *subset* of catalog records that are highly likely relevant to \( Q \), which have a tag that is either the same as one of the query keywords in \( Q \) or their word-correlation factor is at least \( 3 \times 10^{-5} \). This filtering technique is used for reducing the potentially very large number of

\(^7\) Title and Author are the fundamental data provided to a library patron when the latter performs a search.
comparisons to be conducted between records [Christen 2008], i.e., library records in our case, and is generally known as blocking [Kelley 1984], which can significantly minimize the computational costs [Yan 2007].

In order to facilitate the search of those records that have tags that are the same or highly similar to at least one of the keywords in $Q$, we construct another table in MySQL, i.e., $tagid(tag, id)$, which is ordered alphabetically by $tags$ (as opposed to the $idtag$ table which contains the same information but is ordered by $id$.)

Using $3 \times 10^{-5}$ as the cut-off value of the word pairs in the CoM35-13 matrix, one of the reduced 13% matrices, we select a subset of catalog records to reduce the query processing time without affecting the accuracy of retrieving relevant library records. Figure 4 shows the distribution of the word-correlation values among different word pairs in the 13% matrix, whereas Figure 5 shows the number of word pairs included in various 13% matrices that can be used in the pre-processing step for selecting a subset of library records. Note that, as shown in Figure 4, most of the word-correlation factors are not exact matches and are in the range of $1 \times 10^{-4}$ and $1 \times 10^{-9}$, and thus, pairs of words with a word-correlation factor closer to $1 \times 10^{-4}$ are treated as highly similar, whereas word pairs with lower word-correlation factors are treated as less similar.

Figure 5 also shows the average query processing time for selecting the corresponding subset of library records for answering each of the queries in the $HBLL-set$, which consists of a 100 library patron’s queries (as partially shown in Table 2) randomly selected from the 2007 $HBLL$ query log. Clearly, it is unacceptable to use the 13%-matrix,
since it requires an average of 217 seconds in pre-processing each of the queries in HBLL-set. Even though the CoM34-13 matrix decreases the query processing time to an average of 3.72 seconds, the size of the matrix is reduced by only 620 word pairs compared with the CoM35-13 matrix, which on the average requires 4.18 seconds at the pre-processing step. Since the average pre-processing time between the two matrices is insignificant and using a matrix with a larger number of word pairs can only enhance the accuracy of retrieving relevant results, we use the CoM35-13 matrix, which contains 58,532 word pairs. (The CoM35-13 matrix is stored as the table correlation35 in a MySQL database.)

On the average, each word in the CoM35-13 matrix is paired with another 1.01 words. Since the average number of keywords included in a user query is 2.35 [Hoscher 2000], it implies that an average of only three query keywords are evaluated during the pre-processing step, and the involved processing time ranges between 2 and 5 seconds. Initial experimental results using queries in HBLL-set of different sizes (1-5 words) show that the top-10 results, which are often what the users view [Hoscher 2000], are the same when using the CoM35-13 matrix compared with using the other matrices as shown in Figure 5 which further verifies the effectiveness, in terms of accuracy and processing time, of using the CoM35-13 matrix.
Figure 4. Distribution of the word-correlation values in the 13% matrix

Figure 5. Number of word pairs in each 13% matrix and the average processing time of queries in HBLL-set
Table 2. A subset of library patron’s queries in HBLL-set and their stemmed, non-stop word versions used for experimentation

Example 2. Consider the query $Q$: “Climb Alaska” again. We select the library catalog records against $Q$ that have at least one tag that is similar to the query keywords in the similar word (i.e., $w_2$) column of the CoM35-13 matrix as shown in Table 3 such that their word-correlation factors are at least $3 \times 10^{-5}$. This selection step reduces the number of possible library records to be considered from 381 to 37. Note that by using the CoM35-13 matrix, as opposed to the CoM36-13 matrix (as shown in Table 4), we (i) reduce to one-third the total number of similar words to be considered, and (ii) reduce the time required to identify the subset of library records from 19 to 4 seconds without affecting the retrieval of the top relevant library catalog records with respect to $Q$, as shown in Table 5.
### Table 3. Words that are similar to the query keywords in Q: “Climb Alaska” in the CoM35-13 matrix

<table>
<thead>
<tr>
<th>Word</th>
<th>Similar Word</th>
<th>Correlation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>alaska</td>
<td>alaska</td>
<td>1.0</td>
</tr>
<tr>
<td>alaska</td>
<td>climb</td>
<td>7.5x10^{-5}</td>
</tr>
<tr>
<td>alasta</td>
<td>mount</td>
<td>4.5x10^{-5}</td>
</tr>
<tr>
<td>climb</td>
<td>climb</td>
<td>1.0</td>
</tr>
<tr>
<td>climb</td>
<td>alaska</td>
<td>7.5x10^{-5}</td>
</tr>
<tr>
<td>climb</td>
<td>mount</td>
<td>6.2x10^{-5}</td>
</tr>
</tbody>
</table>

### Table 4. Words that are similar to the query keywords in Q: “Climb Alaska” in the CoM36-13 matrix

<table>
<thead>
<tr>
<th>Word</th>
<th>Similar Word</th>
<th>Correlation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>alaska</td>
<td>alaska</td>
<td>1.0</td>
</tr>
<tr>
<td>alaska</td>
<td>borough</td>
<td>3.0x10^{-6}</td>
</tr>
<tr>
<td>alaska</td>
<td>pipeline</td>
<td>3.7x10^{-6}</td>
</tr>
<tr>
<td>alaska</td>
<td>yukon</td>
<td>1.1x10^{-5}</td>
</tr>
<tr>
<td>alaska</td>
<td>mountaineering</td>
<td>1.4x10^{-5}</td>
</tr>
<tr>
<td>alaska</td>
<td>climb</td>
<td>7.5x10^{-5}</td>
</tr>
<tr>
<td>alasta</td>
<td>mount</td>
<td>4.5x10^{-5}</td>
</tr>
<tr>
<td>climb</td>
<td>boulder</td>
<td>3.5x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>ceil</td>
<td>7.9x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>climb</td>
<td>1.0</td>
</tr>
<tr>
<td>climb</td>
<td>hike</td>
<td>3.2x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>ladder</td>
<td>6.7x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>min</td>
<td>1.3x10^{-5}</td>
</tr>
<tr>
<td>climb</td>
<td>rope</td>
<td>5.2x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>stair</td>
<td>4.0x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>steep</td>
<td>4.7x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>summit</td>
<td>3.0x10^{-6}</td>
</tr>
<tr>
<td>climb</td>
<td>mount</td>
<td>6.2x10^{-5}</td>
</tr>
<tr>
<td>climb</td>
<td>alaska</td>
<td>7.5x10^{-5}</td>
</tr>
</tbody>
</table>
Table 5. Library records retrieved using the reduced matrices CoM35-13 and CoM36-13, respectively for the query “Climb Alaska”

3.4 Relevance Ranking

Having selected the subset of library records with respect to a library patron’s query $Q$, we compute the degree of resemblance between $Q$ and each selected library catalog record $R$, which is calculated by adding the correlation factors (in the CoM57-13 matrix) between each of the keywords in $Q$ and tags (i.e., keywords) associated with $R$. The CoM57-13 matrix is used, as opposed to the CoM35-13 matrix considered in the pre-processing step, since the former contains the word-correlation factors ($\geq 5 \times 10^{-7}$) in the 13% matrix, as well as the exact matches for the remaining 87% (as discussed in Section 3.1.1), which provides more accurate similarity measure between $Q$ and $R$ than the more selective CoM35-13 matrix. The degree of resemblance between $Q$ and $R$ is defined as

$$Sim(Q, R) = \sum_{j=1}^{n} \sum_{i=1}^{m} C(q_i, r_j)$$

where

- $n$ ($m$, respectively) denotes the number of keywords in $R$ ($Q$, respectively);
• $q_i$ ($r_j$, respectively) is a keyword in $Q$ ($R$, respectively);

• $C(q_i, r_j)$ is the correlation factor between $q_i$ and $r_j$ in the CoM57-13 matrix, as defined in Equation 1.

The exact matches (with word-correlation value of 1.0) carry a (much) higher weight than other inexact-matched word pairs which are assigned word-correlation factors as low as $5 \times 10^{-7}$ in the CoM57-13 matrix, and thus the $Sim$ value of $Q$ and $R$ is equal to $N$ plus a small value, where $N$ denotes the number of exact matches between $Q$ and $R$. The $Sim$ function assigns higher degree of resemblance to records including tag(s) that match(es) exactly one or more keywords in $Q$. As a consequence, if $R$ includes a tag that matches exactly with one of the keywords in $Q$ and has low similarity with most of the remaining keywords in $Q$, then $R$ is ranked higher than a record including tags that are similar (but not exact match) to most of the keywords in $Q$, which could yield a bias in terms of ranking.

Realizing the shortcomings of $Sim$, we propose another resemblance measure so that if $R$ includes tags highly similar to most (if not all) of the keywords in $Q$, then $R$ should be ranked higher than another record in which only one of its tags is highly similar with only a few of the keywords (or matches exactly one keyword) in $Q$. An alternative measure of the degree of resemblance between $Q$ and $R$ is defined as

$$LimitedSim(Q,R) = \sum_{j=1}^{n} \min\left( \sum_{i=1}^{m} C(q_i, r_j), 1 \right)$$

(3)
where $Q$, $R$, $n$, $m$, $C$, $q_i$, and $r_j$ are as defined in Equation 2.

By using the $Min$ function in Equation 3, we impose a constraint on summing up the correlation factors of keywords in $Q$ and $R$. Even if a tag $T$ in $R$ (i) matches exactly one of the keywords in $Q$ and (ii) is similar to some of the remaining keywords in $Q$ (which would yield a value greater than 1.0, the word-correlation factor of an exact match), we limit the cumulative word-correlation factors of $T$ to 1.0. This constraint ensures that if $R$ contains a dominant tag $T$, i.e., $T$ is similar to (or the same as) a few keywords in $Q$, $T$ alone cannot significantly impact the resemblance value of $R$ and $Q$, whereas if $R$ contains a number of tags that are similar to most of the keywords in $Q$, then $R$ is assigned a higher degree of resemblance due to its diversity in matching keywords in $Q$.

Example 3. Consider the query $Q$ defined in Example 1. Table 6 (Table 7, respectively) shows 10 (out of the 37) retrieved catalog records and their degrees of resemblances with respect to $Q$ computed by using the $Sim$ ($LimitedSim$, respectively) measure. Table 8 shows the titles of the records in Tables 6 and 7. By restricting the sum of the word correlation factors between a tag in $R$ and all the keywords in $Q$ to 1.0 using Equation 3, a comparatively higher degree of resemblance (i.e., $LimitedSim$ value) is assigned to library catalog records which include tags that match most of the keywords in $Q$. Even though Record 3 in Table 7 has a lower similarity value (computed by using Equation 3) with respect to the same record in Table 6 (computed by using Equation 2), Record 3 is ranked higher, i.e., fourth position, in Table 7, since its
keywords are similar to both keywords in \( Q \) (i.e., climb and Alaska), which indeed is more relevant in terms of its content than the fourth ranked record, i.e., Record 6, in Table 6, which is similar to only one of the keywords in \( Q \) (i.e., Alaska), and the contents of Records 3 and 6 have been verified manually. Moreover, records that are related (in term of their contents with respect) to \( Q \), such as Records 1, 2, and 4, are ranked higher (i.e., at positions 9, 7, and 8, respectively) by LimitedSim, whereas the same records are ranked lower (17, 16, and 15, respectively) by Sim.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Record #</th>
<th>Tags</th>
<th>Degree of Resemblance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>alaska, climbing, mountaineering</td>
<td>2.0000166960</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>alaska, climbing, mountaineering</td>
<td>2.0000166960</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>alaska, climbing, hunting</td>
<td>2.0000154134</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>alaska, nature, history</td>
<td>1.0000756434</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>alaska, aviation, mountaineering</td>
<td>1.0000123297</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>alaska, mountain, memoir</td>
<td>1.0000121143</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>mountaineering, climbing, adventure</td>
<td>1.0000094294</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>photography, climbing, mountaineering</td>
<td>1.0000094088</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>climbing, mountaineering, history</td>
<td>1.0000092849</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>alaska, photography, travel</td>
<td>1.0000003964</td>
</tr>
</tbody>
</table>

Table 6. Ten of the catalog records ranked for query Q: “Climb Alaska” using Sim
Table 7. Top 10 catalog records ranked for query Q: “Climb Alaska” using LimitedSim

<table>
<thead>
<tr>
<th>Rank</th>
<th>Record #</th>
<th>Tags</th>
<th>Degree of Resemblance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>alaska, climbing, mountaineering</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>alaska, climbing, mountaineering</td>
<td>2.0</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>alaska, climbing, hunting</td>
<td>2.0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>mountaineering, climbing, adventure</td>
<td>1.0000060541</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>alaska, aviation, mountaineering</td>
<td>1.0000045620</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>alaska, mountain, memoir</td>
<td>1.0000042120</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>climbing, mountaineering, history</td>
<td>1.0000014629</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>photography, climbing, mountaineiring</td>
<td>1.0000014579</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>alaska, photography, travel</td>
<td>1.0000013700</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>alaska, nature, history</td>
<td>1.0000001068</td>
</tr>
</tbody>
</table>

Table 8. Titles of the ten ranked library catalog records shown in Tables 6 and 7

<table>
<thead>
<tr>
<th>Record #</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alaska: Images of the Country</td>
</tr>
<tr>
<td>2</td>
<td>Climbing in North America</td>
</tr>
<tr>
<td>3</td>
<td>Facing the Extreme: One Woman's Story of True Courage, Death-defying Survival, and Her Quest for the Summit</td>
</tr>
<tr>
<td>4</td>
<td>Galen Rowell's Vision: the Art of Adventure Photography</td>
</tr>
<tr>
<td>5</td>
<td>In the Shadows of Denali</td>
</tr>
<tr>
<td>6</td>
<td>More Readings from One Man's Wilderness: the Journals of Richard L. Proenneke, 1974-1980</td>
</tr>
<tr>
<td>7</td>
<td>Mt. McKinley: the Pioneer Climbs</td>
</tr>
<tr>
<td>8</td>
<td>The Ascent of Denali: A Narrative of the First Complete Ascent of the Highest Peak in North America</td>
</tr>
<tr>
<td>9</td>
<td>Wager with the Wind: the Don Sheldon Story</td>
</tr>
<tr>
<td>10</td>
<td>Wilderness of Denali Explorations of a Hunter-Naturalist in Northern Alaska</td>
</tr>
</tbody>
</table>

3.5 Query Processing Time

As stated earlier, one of the design goals of EnLibS is to process user queries with processing time compatible with existing library catalog search engines. In an attempt to reduce the query processing time, we have constructed sophisticated data/file structures for storing (i) general information about library catalog records, (ii) LibraryThing’s tags describing the content of library catalog records, and (iii) the reduced word-correlation
matrices (i.e., the correlation35 and correlation57 MySQL tables), besides using the InnoDB storage engine of MySQL database, which is designed for maximizing the performance in processing large data volumes and has a CPU efficiency that is not matched by other disk-based relational database engines (see http://dev.mysql.com/doc/refman/5.1/en/innodb-overview.html).

3.5.1 Substring Prefix

Since significant query processing time is allocated for selecting the proper subset of library catalog records with respect to a user’s query, we have implemented a prefix indexes on correlation35, correlation57, and tagid tables. In creating the prefix indexes on these tables, we follow the recommendations made by [Dubois 2005] who claims that (i) since shorter values are compared more quickly, implementing prefix indexes on smaller index values as opposed to indexing the entire column allows faster lookups, (ii) smaller indexes require less disk access, and (iii) by considering shorter indexing values, MySQL can hold more keys in the cache memory, which translates into less index blocks swapping from disks in performing a search. Hence, we use a pre-determined prefix length in defining a prefix index for the corresponding columns in correlation35, correlation57, and tagid tables, instead of indexing the entire columns in the corresponding tables.

In creating an index on a string column, [Dubois 2005] suggests indexing 15% of the entire length of the column. Based on these recommendations and since the tag (word, respectively) in the tagid (correlation35 and correlation57, respectively) table is between 15
and 20 characters long, we define a prefix index on the string prefix of length 3 in the “word” column in the correlation35 and correlation57 tables, and the “tag” column in the tagid table.

3.5.2 Query Processing Time/Memory Allocation for Indexing

We have verified the appropriateness of choosing the three-character prefix strings as the prefix index values. Figure 6 shows (i) the average time (in seconds) for processing the queries in HBLL-set using prefix indexes of different prefix sizes, i.e., 3, 5, and 8 characters as well as (ii) the memory space required for these prefix indexes. Although the difference between the average query processing time when using the prefix index of size 3 instead of size 5 is not significant (7.0 versus 7.8 seconds), the required memory space is reduced (from 195.4 MB to 181.3 MB), which further confirms the ideal choice of using the three-character prefix indexes. Furthermore, the subset of catalog records chosen at the pre-processing step does not change when prefix indexes of different sizes are implemented. Hence, the accuracy of the retrieval is not compromised when using shorter (instead of longer) prefix indexes.
Table 9 shows the size (in MB) of each indexed table in our MySQL database, as well as the size (in MB) of the corresponding prefix indexes, whereas Figure 7 shows the average processing time required to answer a query, with and without using the (top-3) tags and prefix indexes on the queries, in HBLL-set. Due to the significant processing time reduction (from 429 to 7 seconds), the choice of using the (i) secondary indices, (ii) the CoM35-13 matrix, and (iii) top-three LibraryThing tags is obvious.

<table>
<thead>
<tr>
<th>Table</th>
<th>Size in Memory</th>
<th>Prefix Index Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>tagid</td>
<td>107</td>
<td>69.2</td>
</tr>
<tr>
<td>correlation5en7</td>
<td>25</td>
<td>14.7</td>
</tr>
<tr>
<td>correlation3en5</td>
<td>4.5</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Table 9. Size (in MB) of different indexed tables
Furthermore, in our pre-processing step the subset of selected library catalog records contain tags that match exactly or are highly similar to the keywords in a user query. The number of highly similar keywords in the records (with respect to the keywords in a user’s query) determines the number of records to be further ranked. Moreover, the more records retrieved, the higher the number of records to be evaluated in order to determine their degrees of resemblance with respect to a user’s query, and the longer query processing time is required. By using the CoM35-13 matrix in processing the queries in HBLL-set, it has been shown that the average number of similar query keywords and the original query keywords to be compared with LibraryThing tags is 9, as opposed to 200, if the CoM57-13 matrix is used instead. More importantly, the reduced number of keywords to be compared does not affect the quality of the retrieved results.
3.6 The Overall Evaluation Process

Figure 8 shows the entire query evaluation process of EnLibS, which illustrates that when a library patron submits a query $Q$, keywords in $Q$ are first reduced to their grammatical roots and stopwords are eliminated, i.e., step (i). Using the set of non-stop, stemmed keywords $K$ and the correlation table, we retrieve the set of correlated keywords in the table, including the keywords from $K$, i.e., step (ii), which are matched with the tags that describe each of the library catalog records in the tagid table, and the matching yields the subset $S$ of library catalog records that are highly likely relevant to $Q$, i.e., step (iii). Hereafter, using the idtag table we identify each record in $S$ and based on their tags (i.e., the top-three LibraryThing’s tags associated with a particular record based on their frequency count) along with the word-correlation factors from the correlation table, we rank the retrieved records in $S$ according to their degrees of resemblance with $Q$, i.e., step (iv).

![Diagram](image)

Figure 8. The overall query evaluation process of EnLibS
Chapter 4

Experimental Results

In this section, we assess the overall performance of EnLibS in terms of its (i) accuracy in retrieving relevant results, (ii) reliability in providing proper rankings on retrieved library records, (iii) capability in reducing the percentage of queries that yield no results, and (iv) compatibility with the HBLL system with regard to the query processing time. We first describe the dataset (in Section 4.1) used for the experiments in verifying the anticipated properties listed in (i) - (iv). Hereafter, we detail the controlled experiments conducted by independent appraisers on the relevance and ranking of the query results generated by EnLibS and the HBLL system separately, which provide an objective and unbiased measure on the performance of both querying systems (in Section 4.2). The performance analysis is based on a number of evaluation strategies to be introduced (in Section 4.3), which will be followed by the detailed discussion on (i) the percentage of library searches performed by EnLibS that yield zero-hits (in Section 4.4.1), i.e., no results, (ii) the effectiveness of using our similarity matching approach in retrieving highly relevant query results (in Section 4.4.2), and (iii) the accuracy of our ranking approach (in Section 4.4.3). Afterward, we discuss the query processing time of EnLibS in retrieving and ranking relevant library records (in Section 4.4.4). During the evaluation process, we compare the performance of EnLibS and the HBLL system.
4.1 The Dataset

To evaluate the performance of EnLibS in querying library catalogs, we used queries in the HBLL query log, which (as mentioned in Chapter 1) was created between July 2006 and January 2007, a file that is 144 MB in size and the average size of each entry in the log is 180 bytes. Each entry in the log includes (i) a query, (ii) the date and time when the query was formulated, and (iii) the corresponding number of library records retrieved with respect to the query. Due to the large number of queries in the log (with approximately one million queries), we randomly selected 500 of them, which constitute the test dataset, denoted HBLL-log. The chosen 500 queries are sampled from queries used for searching the HBLL library catalog, since they cover various subject areas and include different n-gram (1 ≤ n ≤ 5) queries. (See Table 10 for the types of queries in HBLL-log and Table 11 for some of the subject areas covered by the HBLL-log queries.) Due to the (i) randomness of the queries in HBLL-log, (ii) the diversity of users who formulated the queries, and (iii) the general subject areas covered in the queries, HBLL-log is an ideal dataset for our empirical study.

4.2 Controlled Experiments

In order to objectively assess the performance of EnLibS, we conducted a controlled experiment in which each of the participants (i.e., appraisers) involved in the empirical study was asked to evaluate a set of randomly selected queries from the HBLL query log.

---

8 To the best of our knowledge, there are no existing benchmark datasets/measurements that can be used for evaluating the retrieval and ranking performance of any library system.

9 Appraisers who participated in the performance evaluation were volunteers.
log\textsuperscript{10}. (See Appendix A for the list of queries used.) The purpose of this controlled experiment is to determine the gold-standard that will be used for comparing the performance of EnLibS and the HBLL system.

<table>
<thead>
<tr>
<th>Type of Query</th>
<th>Total Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-word (unigram) queries</td>
<td>58</td>
<td>11.6%</td>
</tr>
<tr>
<td>2-word (bigram) queries</td>
<td>201</td>
<td>40.2%</td>
</tr>
<tr>
<td>3-word (trigram) queries</td>
<td>198</td>
<td>39.6%</td>
</tr>
<tr>
<td>4-word (4-gram) queries</td>
<td>29</td>
<td>5.8%</td>
</tr>
<tr>
<td>5-word (5-gram) queries</td>
<td>14</td>
<td>2.8%</td>
</tr>
<tr>
<td>Number of queries that retrieved no results</td>
<td>81</td>
<td>16.2%</td>
</tr>
</tbody>
</table>

Table 10. Different types of queries in HBLL-log as well as the number of zero-hits queries

<table>
<thead>
<tr>
<th>Subject Areas</th>
<th>Number of Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting (12)</td>
<td>(22)</td>
</tr>
<tr>
<td>Arts (19)</td>
<td>(22)</td>
</tr>
<tr>
<td>Biology (12)</td>
<td>(20)</td>
</tr>
<tr>
<td>Business (36)</td>
<td>(19)</td>
</tr>
<tr>
<td>Chemistry (18)</td>
<td>(21)</td>
</tr>
<tr>
<td>Computer Science (27)</td>
<td>(37)</td>
</tr>
<tr>
<td>Economy (9)</td>
<td>(27)</td>
</tr>
<tr>
<td>Education (11)</td>
<td>(7)</td>
</tr>
<tr>
<td>Epidemiology (3)</td>
<td>(16)</td>
</tr>
<tr>
<td>Foreign Languages (22)</td>
<td></td>
</tr>
<tr>
<td>General Culture (30)</td>
<td></td>
</tr>
<tr>
<td>History (19)</td>
<td></td>
</tr>
<tr>
<td>Law (21)</td>
<td></td>
</tr>
<tr>
<td>Literature (37)</td>
<td></td>
</tr>
<tr>
<td>Medicine (27)</td>
<td></td>
</tr>
<tr>
<td>Movies (7)</td>
<td></td>
</tr>
<tr>
<td>Music (16)</td>
<td></td>
</tr>
<tr>
<td>Physics (12)</td>
<td></td>
</tr>
<tr>
<td>Poetry (4)</td>
<td></td>
</tr>
<tr>
<td>Politics (25)</td>
<td></td>
</tr>
<tr>
<td>Religion (31)</td>
<td></td>
</tr>
<tr>
<td>Science (32)</td>
<td></td>
</tr>
<tr>
<td>Sociology (7)</td>
<td></td>
</tr>
<tr>
<td>Sports (9)</td>
<td></td>
</tr>
<tr>
<td>Statistics (24)</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. (Some of the) Subject areas and their corresponding number of queries in HBLL-log

For each retrieved result of a test query, the appraisers determined whether the result was relevant. The appraisers were also asked to order (to the best of their knowledge), i.e., rank, the retrieved results in terms of their degrees of relevance with respect to the corresponding query.

\textsuperscript{10} None of the queries used in the controlled experiment were the queries in HBLL-log, even though they were randomly selected from the same HBLL query log.
4.2.1 The Cross-over Experiment

In designing a controlled experiment, we adapted the cross-over experiment [Jones 2003] which requires each subject, i.e., appraiser, to conduct a sequence of experiments individually. A cross-over experiment that evaluates two alternatives is called two-period two-treated design [Jones 2003], which requires each subject to evaluate two different alternatives: A, which is the set of query results generated by EnLibS in our empirical study and B, which is the set of query results generated by the HBLL system in the study. Half of the subjects were given A first and then B, whereas the other half received B first and then A. According to [Jones 2003], cross-over experiments can be used for comparing measures obtained by the same subject on different alternatives, which are the set of query results generated by EnLibS and the HBLL system, respectively.

4.2.2 The Number of Participants

Having adapted the cross-over experiment, we must determine the number of subjects who should participate in our empirical study so that the experimental results are reliable and objective. We depended on well-known statistical methods to define the appropriate number of participants.

In statistics, two different types of errors, Type I and Type II errors, are defined [Jones 2003]. Type I errors, also known as $\alpha$ errors or false positives, are the mistakes of rejecting a null hypothesis when it is true, whereas Type II errors, also known as $\beta$ errors or false negatives, are the mistakes of accepting a null hypothesis when it is in fact false. We consider Type I and II errors, since the probabilities of Types I and II errors have a direct
effect on determining the number of participants that should be involved in a controlled experiment so that the experimental results are statistically sound. It turns out that the number of participants can be defined according to the chosen percentage of Types I and II errors. We applied the following formula in [Jones 2003] to determine the number of participants, \( n \), which is dictated by the probabilities associated with errors of Types I and II:

\[
n = \left( \frac{Z_{\alpha/2} + Z_{\beta}}{\Delta^2} \right)^2 \cdot 2\sigma^2 + \frac{Z_{\alpha/2}^2}{2}
\]

(4)

where

- \( \Delta \) is the \textit{minimal expected difference}, which is 2 in our study;
- \( \sigma^2 \) is the \textit{variance} of the data, which is 3.551 in our study;
- \( \alpha \) denotes the probability of making a Type I error, which is 0.05 in our study;
- \( \beta \) denotes the probability of making a Type II error, which is 0.20 in our study.

Based on the value of \( \beta \), we can determine the probability of a false null hypothesis to be correctly rejected, i.e., \( 1 - \beta \) [Greene 2000];

- \( Z \) is the value associated with the standard \textit{normal distribution}. According to the standard normal distribution, when \( \alpha = 0.05 \), \( Z_{\alpha/2} = 1.96 \), whereas when \( \beta = 0.20 \), \( Z_{\beta} = 0.84 \).

To determine the values of \( \Delta \) and \( \sigma^2 \) in Equation 4, we conducted an experiment in May 2008, using the randomly sampled 100 queries in \textit{HBLL-set}, a subset of \textit{HBLL query log} (as discussed in Section 3.3) and manually evaluated the \textit{relevance} of the retrieved results of
each. Based on the analysis conducted on the results retrieved for each one of the 100 queries, we observed that the minimal expected difference for comparing EnLibS and the HBLL system is at least 2 (i.e., the $\Delta$ value). The value of $\Delta=2$ implies that if the HBLL system retrieves one relevant document, then we expect EnLibS to retrieve at least three so that we can actually claim that the performance of EnLibS is better.

In computing the variance\(^{11}\), i.e., $\sigma^2$, we calculated the mean among the number of relevant library records retrieved by EnLibS using the queries in HBLL-set, which was 6. Averaging the sum of the square difference between the mean, i.e., 6, and the actual number of relevant records retrieved for each one of the 100 queries, we obtained 3.551, which is the value of $\sigma^2$.

The values of $\alpha$ and $\beta$ are set to be 0.05 and 0.20, respectively, which imply that we have 95% confidence on the correctness of our analysis and that the power (i.e., probability of avoiding false negatives) of our statistical study is 80%. According to [Kazmier 2003] and as stated in [Hinton 2004], 0.05 is the commonly-used value for $\alpha$, whereas 0.80 is a conventional value for $(1 - \beta)$ and a test with $\beta \leq 0.20$ is considered to be statistically powerful. Nota that alternative statistical analysis can be adopted to further enhance the estimation of the proper number of appraisers required in the study.

Based on the values assigned to the variables in Equation 4, we determined that the number of participants (i.e., appraisers) required in our study should be 16, as calculated below.

\(^{11}\) Variance is generally used in statistics, along with the standard deviation measure (which is the square root of the variance), to measure the average dispersion of the scores in a distribution [Urdan 2005].
\[ n = \frac{ (1.96 + 0.84)^2 \times 2 \times 3.551 }{ 2^2 } + \frac{ 1.96^2 }{ 2 } \approx 16 \]

Note that the values of \( \alpha, \beta, \sigma^2, \) and \( \Delta \) directly influence the size of \( n \). Furthermore, the results collected from the \( n \) participants in the study are expected to be comparable with the results that could be obtained by the actual population, i.e., library patrons looking for information using the library catalog.

### 4.2.3 Independent Appraisers

To recruit appraisers, we advertised our empirical study in CS 100 (An Introduction to Computing and Multimedia), EXSC 106 (Badminton, Beginning), EXSC 182 (Tennis, Beginning), EXSC 186 (Volleyball, Beginnings), and EXSC 187 (Volleyball, Intermediate), in addition to recruiting college students and faculty members from various church groups in between October 6, 2008 and October 24, 2008. Since these participants were in different majors and academic standing, they formed a diverse group of appraisers. Figures 9, 10 and 11 provide the demographic information of the appraisers involved in our study.
Figure 9. Educational backgrounds of the appraisers who participated in the empirical study

Figure 10. Academic majors of the appraisers who participated in our study
4.2.4 Number of Queries

In determining the ideal number of queries to be included in the controlled experiment, we relied on two different variables: the average attention span of an adult and the average number of search queries that a person often creates when using a search engine. As mentioned in [Schell 1996 and Rozakis 2002], the average attention span of an adult is between twenty to thirty minutes. Furthermore, Jansen et al. [Jansen 2000], who have evaluated Web users’ behavior specially on (i) the amount of time Web users spend on a Web search engine, (ii) the average size of user’s queries, and (iii) the average number of queries submitted by a user, estimate that the average number of queries created by each user on a Web search engine in one session is approximately 2.8. Therefore, in our study each appraiser was asked to evaluate the retrieved results of six queries, three of which with results retrieved by using EnLibS, whereas the results of the remaining three queries were retrieved by the HBLL system. We considered six to be the proper number.
of queries to be evaluated, since prior to establishing the number of queries in our controlled experiment, we observed that in practice, evaluating six queries could take between thirty to forty minutes, which we believe is the amount of time appraisers will be willing to dedicate for our controlled experiment.

4.2.5 The Evaluation Tasks

Quite often Web search engine users view only the first 10 retrieved results when performing a search [Hoscher 2000]. Thus, we considered only the top (i.e., first) 10 retrieved records (if they exist) for each test query. In other words, for each test query, each appraiser was asked to (i) determine the (non-)relevance of each one of the top 10 retrieved results and (ii) rank the relevant ones in the order of relevance (“1” denotes the most relevant and “10” denotes the least relevant) to the best of their knowledge. (See Figure 19 in Appendix B for a sample evaluation form provided to each of the appraisers. Appendix B also includes details with regard to the information provided to the appraisers in the controlled experiment for each query, in addition to a sample of the results retrieved by EnLibS and the HBLL system.)

To show that the decisions on (non-)relevant retrieved library records and ranking results provided by the appraisers were unbiased, we repeated the study three times. As a result, 48 appraisers were recruited and randomly assigned to three different groups, and 18 distinct queries were arbitrarily partitioned into three sets as shown in Appendix A.
4.3 Evaluation Measures

In assessing the performance of EnLibS, we considered measures commonly used for determining the effectiveness of information retrieval systems, which are Precision, Mean Average Precision, Mean Reciprocal Rank, and Spearman Rank Correlation Coefficient.

4.3.1 Precision

Precision determines the fraction of retrieved results that are relevant. In our study, precision quantifies the set of (relevant) library records retrieved by EnLibS as well as the set of library records retrieved by the HBLL system for each one of the 18 test queries\(^{12}\). Since in our study the appraisers were asked to evaluate only the top 10 retrieved results (a common practiced as claimed in [Hoscher 2000] and discussed in Section 3.3), we adapted two variations of the precision measure that are widely used: precision @ 1 and 10-precision. Precision @ 1 [Cong 2008], denoted P@1, yields the number of queries, \(n\), in a set of queries such that the first retrieved record for each one of the \(n\) queries is relevant to its corresponding query, whereas the 10-Precision value [Goncalves 2004] quantifies the top 10 retrieved results in terms of their relevance with respect to its corresponding query, and is defined as

\[
10\text{-Precision} = \frac{\text{Number of Retrieved Relevant Records}}{10}
\]

\(^{(5)}\)

\(^{12}\) The dataset used in our empirical study, i.e., library records in the HBLL query log (and thus the HBLL-log), have not been previously labeled as (ir)relevant with respect to each query in its set, and hence the recall ratio that measures the fraction of relevant records actually retrieved cannot be determined in this study, which is not as significant as precision in measuring the top-ranked retrieved records.
where *Number of Retrieved Relevant Records* denotes the number of relevant records in the top 10 retrieved results.

### 4.3.2 Mean Average Precision

Besides the precision measures, we also consider the *Mean Average Precision* (*MAP*) [Aslam 2006] to provide additional performance evaluation of *EnLibS*. *MAP* is defined as

\[
MAP(r) = \frac{1}{Q} \times \sum_{q=1}^{Q} \frac{r}{t(q)}
\]  

(6)

where

- *Q* is the total number of queries, i.e., the number of queries assigned to each appraiser (group) in our study, which is *six*;
- *r* is the number of relevant documents to be considered, which is either 3, 5, 7, or 10 in our study;
- *t(q)* is the total number of records retrieved when the *r*th relevant record on the *q*th (1 ≤ *q* ≤ *Q*) query is encountered.

The ideal value of *MAP(r)* is 1, which denotes that all the top *r* retrieved records are relevant, and the closer *MAP(r)* is to 1, the better the performance of the corresponding IR system is. By using *MAP*, we not only evaluate the average precision of the retrieved results, but at the same time we can determine the effectiveness of our ranking approach which should position higher in the rank the library records with higher degrees of relevance with respect to the corresponding query [Baeza-Yates 1999].
Since we evaluate only the top 10 retrieved results, there are possible cases in which the \( r^{th} \) relevant record as defined in \( \text{MAP}(r) \) is not among the top 10 retrieved library records. To address this issue we have considered the Inferred Average Precision (infAP) [Yilmaz 2006], which estimates the precision value of the \( r^{th} \) relevant document based on the (available) probability of the given documents being relevant, i.e., it predicts the average precision measure of the \( r^{th} \) relevant document, which is not in the top 10 ranking. infAP is defined as

\[
\text{infAP} = \frac{1}{k} + \frac{k - 1}{k} \left( \frac{|\text{judgedDoc}|}{|\text{rel}|} \right) \times \frac{|\text{rel}|}{|\text{rel}| + |\text{nonrel}|}
\]

(7)

where

- \( k \) is the number of relevant documents (library records in our case) of interest, i.e., \( r \) in \( \text{MAP}(r) \);
- \(|\text{rel}|\) is the number of relevant documents in a given set of judged documents;
- \(|\text{nonrel}|\) is the number of irrelevant documents in a given set of judged documents;
- \(|\text{judgedDoc}|\) is the number of judged documents.

Note that the given set of judged documents is a set of documents previously labeled as relevant or irrelevant with respect to a query. In our case, the (number of) documents in

---

\(^{13}\text{JudgedDoc} \) is “d100” in the original equation in [Yilmaz 2006]. We change the name of the variable so that the new variable name is more meaningful than the original one.
the set of judged documents are the same as the (number of) documents shown to the appraisers, i.e., 10.

In Equation 7, 1/k denotes the probability of the current document, i.e., in position k, being relevant, and \( \frac{k-1}{k} \times \left( \frac{1}{k-1} \times \frac{|rel| + |nonrel|}{|rel|} \right) \) estimates the probability of the relevance within the (k-1) documents.

Yilmaz et al. [Yilmaz 2006] state that the average and the estimated precisions are the same, when the \( r^{th} \) relevant records are available, and the estimated precision is a reasonable approximation of the average precision when the \( r^{th} \) relevant record is not available. Hence, \( \text{infAP} \) is an appropriate metric for estimating the (mean) average precision of the \( r^{th} \) record when there is insufficient information available to compute \( \text{MAP}(r) \), i.e., when the number of possible relevant records to be considered or retrieved is less than \( r \).

Example 4. One of the appraisers in Group 1 claimed that only 7 library records were relevant to Query 1. Hence, determining the \( \frac{r}{t(q)} \) in Equation 7 for computing \( \text{MAP}(10) \) is not possible. Instead, we replace the \( \frac{r}{t(q)} \) value by its estimated value, i.e., \( \text{infAP} = \frac{1}{10} + \frac{10-1}{10} \left( \frac{10}{10-1} \times \frac{7}{7+3} \right) = 0.80. \)

4.3.3 Mean Reciprocal Rank

Another measure we have adapted for further verifying the effectiveness of EnLibS in retrieving relevant results is the Mean Reciprocal Rank (MRR) [Cong 2008, Voorhees 1999]. As stated in [Cong 2008], MRR determines the reciprocal rankings averaged over
a set of $n$ queries, which is defined in [Zelkowitz 2005] as follows:

$$MRR = \frac{\sum_{q=1}^{n} \frac{1}{rank_q}}{n}$$

(8)

where

- $rank_q$ is the position (in the ranking) of the first relevant document retrieved by the $q^{th}$ query;

- $n$ is the total number of queries evaluated.

Using $MRR$ we can establish the number of library records a library patron must examine in the ranked list of retrieved results before finding the first relevant library record. As mentioned in [Zelkowitz 2005], the closer to 1 $MRR$ is, the better the corresponding retrieval system performs.

### 4.3.4 Spearman Rank Correlation Coefficient

Having defined the metrics for measuring the effectiveness of retrieving relevant library records retrieved using $EnLibS$, as well as the HBLL system, we proceed to consider the ideal metric for measuring the appropriateness of the order, i.e., ranking, of the retrieved results generated by $EnLibS$ (as well as the HBLL system). We determine the ranking accuracy of the retrieved results of a test query generated by $EnLibS$ by comparing it with the ranking on the same set of retrieved results created by independent appraisers (which serve as the gold-standard) using the Spearman Rank Correlation Coefficient [Callan 2001].
According to Callan et al. [Callan 2001], the Spearman Rank Correlation Coefficient is widely-used for comparing two different orderings, i.e., rankings in our study. The coefficient is a value between -1 and 1, where 1 indicates that the two given orderings are identical, 0 means that the two orderings are not related, and -1 implies that the two orderings are in the reversed order. Hence, using the Spearman Rank Correlation Coefficient, we can measure whether our ranking strategy, which is based on the similarity measure between a given query and the retrieved library records, is indeed ideal for ranking retrieved library records. Using the coefficient, we determine how closely matched each ranking generated by EnLibS (the HBLL library system, respectively) and the corresponding ranking anticipated by library patrons (through the individual appraisers) is.

Prior to computing the Spearman Rank Correlation Coefficient, we combine the ranking positions of the retrieved results of each test query $Q$ determined by a group of appraisers in our study by averaging the positions assigned to each retrieved library record by the appraisers, which yields a unique ranking of library records retrieved for $Q$. To perform a comparable comparison between the ranking generated by the appraisers and the ranking generated by EnLibS (the HBLL system, respectively), we order the average ranking positions provided by the appraisers and assign them a value in the range from 1 to 10 (with “1” being the most relevant record and “10” being the least relevant), which is also the range of which the library records retrieved by EnLibS (HBLL system, respectively) are ordered.
We follow the same assumption given by [Callan 2001] on Spearman Rank Correlation Coefficient, which allows two or more elements (i.e., library records in our study) in the ranking to share the same ranking position. Hence, given any two different rankings generated for a query, one created by EnLibS and the other by averaging the rankings of the corresponding set of retrieved library records determined by the independent appraisers, the Spearman Rank Correlation Coefficient, i.e., $R$, (defined below) is used for verifying the accuracy of our ranking approach.

$$R = 1 - \frac{6}{n^3 - n} \left( \frac{\sum d_i^2 + \frac{1}{15} \sum f_k^3 - f_k}{\sqrt{1 - \frac{\sum f_k^2 - f_k}{n^3 - n}}} \right)$$  \hspace{1cm} (9)$$

where

- $d_i$ is the difference between the two rankings for the same document $i$ in terms of their relative positions;
- $n$ is the number of documents ranked, i.e., the top 10 library records in our study;
- $f_k$ is the number of ties in the $k^{\text{th}}$ ($\geq 1$) group of ties in the ranking computed by the independent appraisers. The group of ties shows the number of possible ties in the corresponding position in the ranking.

Example 5. The average ranking provided by the appraisers in Group 1 for Query 2 is shown in Table 12. The appraisers in Group 1 assigned two different library records, i.e., library records retrieved and ranked as the 1\textsuperscript{st} and 7\textsuperscript{th} records by EnLibS, to the 3\textsuperscript{rd} position. The Spearman Rank Correlation
Coefficient, i.e., $R$, between the ranking generated by EnLibS and the average ranking provided by the appraisers of Group 1 on Query 2 is

$$R = 1 - \frac{6}{10^3 - 10} \left( \frac{(4 + 1 + 25 + 25 + 4 + 16 + 16 + 9 + 9 + 0) + \frac{1}{12} (2^3 - 2)}{\sqrt{1 - \frac{2^3 - 2}{10^3 - 10}}} \right)$$

$$= 0.34$$

where $f_1 = f_3 = 2$, which is the number of library records tied in the third position, and 0.34 is the Spearman Rank Correlation Coefficient between the ranking provided by EnLibS and the (averaged) one provided by the involved appraisers.

| EnLibS Ranking | Average Appraiser Ranking | $|d|$ | $|d|^2$ |
|----------------|---------------------------|------|-------|
| 1              | 3                         | 2    | 4     |
| 2              | 1                         | 1    | 1     |
| 3              | 8                         | 5    | 25    |
| 4              | 9                         | 5    | 25    |
| 5              | 7                         | 2    | 4     |
| 6              | 2                         | 4    | 16    |
| 7              | 3                         | 4    | 16    |
| 8              | 5                         | 3    | 9     |
| 9              | 6                         | 3    | 9     |
| 10             | 10                        | 0    | 0     |

Table 12. Library records ranked by EnLibS and the average ranking generated by using the rankings created by the appraisers in Group 1 for Query 2 “Disney” along with the (squared) difference among the relative positions

4.4 Performance Evaluations
Having introduced the performance evaluation measures used in assessing the performance of EnLibS, we first discuss the empty query results generated by using HBLL-log (in Section 4.4.1). Hereafter, we present the degree of relevance of the results retrieved by EnLibS (in Section 4.4.2) as well as the effectiveness of the ranking approach of EnLibS (in Section 4.4.3) using the assessments provided by the appraisers in our controlled experiment. We also compare the performance of EnLibS with the HBLL system in terms of the query processing time (in Section 4.4.4). In each subsection, we compare performance measure of EnLibS with the corresponding one of the HBLL system.

4.4.1 Queries with Zero-hits

As discussed in the Related Work section, a shortcoming of existing library systems is the generation of a large percentage of zero-hits, i.e., no library records were retrieved for the corresponding library patron’s queries. With that in mind, one of the design goals of EnLibS is to minimize the number of zero-hits and improve the quantity of relevant library records retrieved using word-similarity matching and folksonomies from LibraryThing (as discussed in Chapter 3). To verify that this design goal is achieved, we compared the numbers of queries in HBLL-log that yield zero-hits as a result of retrieval using the HBLL system and EnLibS, respectively.

According to the retrieved results of the 500 queries in HBLL-log, the percentage of zero-hits searches is reduced from 16.2% (using the HBLL system) to 1% (using EnLibS), i.e., from 81 to 5 zero-hits, which is a significant improvement (see Figure 12). Most
importantly, the (top 10) results retrieved by *EnLibS* for each of the queries for which the HBLL library system retrieved no results at all were manually examined, and the examination showed that *EnLibS* retrieved *relevant results* for queries that the HBLL library system yielded zero-hits.

![Figure 12. Zero-hits queries in HBLL-log generated by the HBLL system versus EnLibS](image)

### 4.4.2 Relevance of the Retrieved Results

We have computed the precision values on the retrieved library records using Precision @ 1, 10-Precision, Mean Reciprocal Rank (*MRR*), and Mean Average Precision (*MAP*).

**Precision@1 (P@1)**

We computed the P@1 value based on the assessments performed by the appraisers on relevant library records retrieved at the top position for each test query. As shown in Figure 13, *EnLibS* achieves a 0.84 P@1 value, which means that on an average 84% of the queries processed by *EnLibS* yield a relevant result at the *top* position in the ranking of
the retrieved library records, whereas the HBLL system yields a 0.61 P@1 value, i.e., 61% of the queries processed by the HBLL system come up with a relevant result positioned first in its ranking. Hence, we conclude that EnLibS is capable of retrieving a relevant result that ranks first more often than the HBLL system.

Figure 13. Average P@1 computed for each of the six queries evaluated by each group in our study based on the top relevant library records retrieved by EnLibS and the HBLL system on each of the test queries, respectively.

10-Precision

Another evaluation measure we have considered is the 10-Precision, which estimates the number of relevant library records appeared in the top 10 retrieved results generated by the HBLL system and EnLibS on each test query. Figure 14 shows that regardless of which particular appraiser group, the 10-Precision values of the retrieved results generated by EnLibS are consistently higher than the 10-Precision values of the results generated by using the HBLL system. What is more, EnLibS achieves an average of 0.67
10-Precision as opposed to an average of 0.58 10-Precision obtained by the HBLL system, which indicates that on the average 67% of the top 10 results retrieved and ranked by EnLibS are relevant with respect to a patron’s query, compared with the 58% achieved by the HBLL system.

Figure 14. The average 10-Precision values computed for the test queries in our study based on the library records retrieved by EnLibS and the HBLL system, respectively

**MRR Measures**

The reader is likely interested in the number of library records retrieved by EnLibS (the HBLL system, respectively) that a library patron is required to scan through before locating the first relevant one. The number can be determined by the Mean Reciprocal Rank (MRR) measure. As shown in Figure 15 (on the average) the first library record retrieved by EnLibS for a query is relevant (since its MRR value is 1.13), whereas when using the HBLL system, two library records are expected to be retrieved before locating the first relevant one (since its MRR value is 2.09). Based on the MRR value, it is clear
that the EnLibS is more effective in locating relevant results earlier in the ranking than the HBLL system does.

![Figure 15. Average MRR computed by using the top relevant library records retrieved by EnLibS and the HBLL system for each test query](image)

**MAP Measures**

We have also evaluated the performance of EnLibS in terms of the MAP measure. In comparing the MAP values generated by the HBLL system and EnLibS, we set \( r = 3, 5, 7, \) and 10 sequentially, i.e., evaluating the top-3, top-5, top-7, and top-10 relevant records retrieved by using the 18 queries in our study, respectively. As shown in Figure 16, the MAP values obtained by EnLibS are higher than the corresponding ones obtained by the HBLL system. A higher MAP value means that less library records are expected to be accessed or examined by library patrons in finding the desired number (i.e., \( r \)) of relevant records. According to the experimental results, on the average EnLibS locates
the $r \in \{3, 5, 7, 10\}$ desired relevant records between the $r^{th}$ and $r + 3^{th}$ record, whereas the HBLL system requires an average between the $r + 2^{th}$ and $r + 7^{th}$ record.

![Graph showing MAP(r) values for r ∈ {3, 5, 7, 10} based on the query results retrieved by EnLibS and the HBLL system, respectively.]

Figure 16. The MAP(r) values for $r \in \{3, 5, 7, 10\}$ based on the query results retrieved by EnLibS and the HBLL system, respectively

### 4.4.3 Ranking of Retrieved Records

Besides analyzing the number of relevant library records retrieved by EnLibS, we have also evaluated the effectiveness of our ranking approach in presenting to library patrons the retrieved library records ordered according to their degrees of relevance with respect to a query. In verifying the accuracy on ranking of the retrieved results, Figure 17 shows the Spearman Rank Correlation Coefficient computed by using the results of each query in our study. According to the experimental results, thirteen out of the eighteen queries processed by EnLibS yield a ranking that is more related to, i.e., alike, the average ranking generated by the appraisers involved in this study than the ranking yielded by the HBLL. Based on the observed results, this tendency of the rankings of being opposite to
the ones provided by the HBLL system is due to its ranking approach, i.e., ordered by publication or alphabetically, that the HBLL system use for presenting library catalog search results to the users. As shown in Figure 17, the coefficients computed for EnLibS on the eighteen queries are higher than coefficients computed for the HBLL system, except for Queries 9, 10, 11, 16, and 17. We observed that the HBLL system ranked the retrieved results more often in an order that tends to be the opposite to the one suggested by the appraisers, since in five out of the eighteen queries, the HBLL system yields a Spearman Rank Correlation Coefficient lower than 0, whereas as shown in Figure 17, none of the coefficients achieved by EnLibS is lower than 0. Furthermore, for queries 9 and 10, i.e., Q9 and Q10, the Spearman Rank Correlation Coefficient for the HBLL system is close to 1, in which only 3 library records were retrieved, as opposed to the 10 retrieved by EnLibS, which explains the differences in the ranking. If we were to consider only the top-3 results retrieved by EnLibS and ranked by the appraisers, the corresponding Spearman Rank Correlation Coefficient would also be closer to 1. Regarding Queries 11, 16, and 17, in each of these three cases EnLibS retrieved more relevant results than the HBLL system (as shown in Table 19 in Appendix C); however, the ranking provided by the independent appraisers tended to be more similar to the one provided by the HBLL system than the one provided by EnLibS. This is due to the considerably different rankings created by the different appraisers for the same queries.

While it is possible to compare the ranking computed by the appraisers in our study with respect to the rankings generated by EnLibS (the HBLL system, respectively), it is not possible to compare the Spearman Rank Correlation Coefficient of EnLibS with respect to
the coefficient of the HBLL system, since most of the retrieved library records were
different, i.e., the top 10 library records retrieved using EnLibS may not necessarily be
the same ten library records retrieved by the HBLL system.

It is worth mentioning that based on the analyzed data in our study, the rankings
provided by different appraisers are directly affected by their (educational) background.
While different appraisers tend to agree in the (non-)relevancy of a library record with
respect to a test query, the orders in which they expect the library records to be retrieved
usually differ, which clearly affects the Spearman Rank Correlation Coefficient values
computed for both EnLibS and the HBLL system. Table 13 shows six (out of the 10)
retrieved records for the query “Child Development”. Note that while the appraisers’
relevance rating of the results is identical, the ranking provided by Appraiser A, whose
educational background is on Exercise Science, is very different to the ranking provided
by Appraiser B, whose background is on Education.

However, due to the randomness in which the participants were assigned to different
groups, the reported evaluations in this section should not be compromised by the
correlation between the education backgrounds of the appraisers and the ranking of the
results.

In Appendix C, we include a summary of the results computed for each of the test
queries in each of the three groups in our study in terms of its average ranking and the
relevance of its retrieved results.
Figure 17. The Spearman Rank Correlation Coefficient computed for each of the test queries in our study using EnLibS and the HBLL system, respectively.

Table 13. Relevance and ranking evaluations on the library records retrieved for the query “Child Development” by EnLibS that were generated by two different appraisers in Group 1.

<table>
<thead>
<tr>
<th>Library Record</th>
<th>Relevance</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Appraiser A</td>
<td>Appraiser B</td>
</tr>
<tr>
<td>Introduction to the anatomy and physiology of</td>
<td>relevant</td>
<td>relevant</td>
</tr>
<tr>
<td>Bears, bears everywhere! : supporting children’s</td>
<td>relevant</td>
<td>relevant</td>
</tr>
<tr>
<td>emotional health in the classroom</td>
<td>relevant</td>
<td>relevant</td>
</tr>
<tr>
<td>What’s the harm? : does legalizing same-sex marriage</td>
<td>relevant</td>
<td>relevant</td>
</tr>
<tr>
<td>really harm individuals, families, or society?</td>
<td>relevant</td>
<td>relevant</td>
</tr>
<tr>
<td>The handbook of children, media, and development</td>
<td>irrelevant</td>
<td>irrelevant</td>
</tr>
<tr>
<td>Only-child experience and adulthood</td>
<td>irrelevant</td>
<td>irrelevant</td>
</tr>
<tr>
<td>The handbook of social policy</td>
<td>irrelevant</td>
<td>irrelevant</td>
</tr>
</tbody>
</table>

4.4.4 Query Processing Time

We have assessed the performance of our enhanced library system in terms of processing time by measuring on the average the amount of time required to evaluate each query in the HBLL-log, as shown in Figure 18. When processing the 500 queries in HBLL-log, the
average time required for processing each one of the queries using the HBLL system is 6.1 seconds, whereas by using EnLibS the average time is 7.0 seconds.

Although the query processing time of EnLibS is higher than the query processing time of the HBLL system, the difference, which is less than one second, is not significant, especially when the results retrieved by EnLibS are more accurate, in terms of relevancy and ranking (as discussed in Sections 4.4.1 - 4.4.3.) than the results generated by the HBLL system.

![Graph showing query processing time for EnLibS and HBLL](image)

Figure 18. Query processing time for the queries in HBLL-log using EnLibS and the HBLL system

4.5 Overall Performance of EnLibS

We have detailed the performance evaluations of EnLibS using the metrics described in Section 4.3 and have observed that consistently EnLibS outperforms the HBLL system. By (i) representing the content of library records using LibraryThing tags, as opposed to the rigid LCSH in the HBLL system, (ii) allowing partial, similarity matching, i.e.,
relaxing the exact-match constraint imposed by the HBLL system, and (iii) ranking the retrieved results according to their relevance to a query as opposed to simply listing *most recent publications* first or listing publications *alphabetically*, *EnLibS* retrieves *more* relevant results and *minimizes* the number of queries that retrieved no results, which is one of the main concerns in locating information when using the library catalog.

**4.6 Implementation**

The original word-correlation matrix [Koberstein 2006] was implemented using the C language on an Intel Centrino Duo workstation with dual 2.66 GHz processors, 3 GB RAM, and a hard disk of 320 GB running under the Linux (Ubuntu) operating system.

We installed version 6.0 of MySQL and implemented *EnLibS* using the PHP programming language on an Intel Dual Core workstation with dual 2.66 GHz processors, 3 GB RAM size, and a hard disk of 300 GB running under the Windows XP operating system.

**4.7 Impacts of *EnLibS***

Searches performed by using the HBLL catalog are powered by *SirsiDynix’s Unicorn*, which is installed on most library systems at different places around the world, e.g., Carnegie Mellon University, Arizona State Library, Kansas City Public Library, Pennsylvania State University, Princeton City Schools, Natural Resources Canada Library, Supreme Court of Canada Library, Gribskov Community Library-Denmark, to name a few (see *Unicorn’s* official Web site [http://www.sirsidynix.com/Solutions/Products/integratedsystems.php](http://www.sirsidynix.com/Solutions/Products/integratedsystems.php)). While *Unicorn*
includes necessary features such as modules for circulation, acquisitions, outreach, materials booking, reserves, etc., it still lacks the accuracy in retrieving and ranking relevant library catalog records. By incorporating (i) the use of LibraryThing tags, (ii) similarity matching, and (iii) relevance ranking, we enrich the catalog searches powered by Unicorn and hence the library systems used by many private and public libraries.
Chapter 5

Conclusions

A library catalog offers library patrons a mean to locate the extensive resources available in public and private libraries. Unfortunately, due to the high percentage of searches that yield irrelevant or no results and the lack of relevance ranking, in addition to the difficulty in formulating queries using the rigid and unintuitive keywords defined in the Library of Congress Subject Heading to perform a search, library patrons have been turning to Web search engines to locate the initial information, such as titles, authors, or subject areas, which yield the primitive information that library patrons can later use for querying the library catalog, a tedious and inefficient searching strategy.

In order to improve existing library searches, we have proposed to use word-correlation factors and folksonomies to perform similarity matches between keywords in a library patron’s query and the user-generated tags from LibraryThing, which describe the contents of library books in the library catalog using commonly-used, intuitive words. Experimental results show that the proposed library system, EnLibS, (i) reduces zero-hits query results and (ii) ranks highly relevant library records high by using our similarity matching approach, while maintaining the processing time comparable with existing library catalog search engines. EnLibS outperforms, and can be adapted for enhancing, existing library systems powered by SirsiDynix’s Unicorn, a widely-used search engine at private and public libraries these days.
Regarding future work, we would like to further enhance the accuracy and types of queries that can be handled by EnLibS. We plan to incorporate a Fuzzy Set IR model component on EnLibS to handle Boolean queries, i.e., EnLibS users can formulate their queries using Boolean operators, such as AND, NOT and OR. By using these Boolean operators, EnLibS users can formulate a more sophisticated query, which allows the specification of inclusion, exclusion, and alternation of keywords as an advanced search option (which is available these days among popular Web and library search engines) that can further enhance the expressive power of EnLibS.

Besides enhancing the expressive power of EnLibS, we will consider scaling the values of the word-correlation factors used for computing the degree of similarity among the keywords in a query and the tags that describe a particular library record, since in replacing most of the currently-used, word-correlation factors, which are in the range of $3 \times 10^5$ or lower, for their corresponding scaled values between 0% and 100%, we can provide a more representative, i.e., easier to understand, similarity value for determining the probability that any two words share the same semantic meaning. Using the scaled word-correlation factors, the accuracy of EnLibS should not be affected.
Appendix A

Queries Used in the Controlled Experiments

The following lists include all the queries that were evaluated by independent appraisers in the controlled experiments: Queries 1 to 6 were evaluated by the appraisers in Group 1, Queries 7 to 12 were evaluated by the appraisers in Group 2, and the remaining queries were evaluated by the appraisers in Group 3.

Group 1:

1. Mormon Missionaries
2. Disney
3. Comic book
4. Poisson regression
5. Child development
6. Shakespeare Sonnets

Group 2:

7. African culture
8. Divorce and separation
9. Uganda children
10. Geography Costa Rica
11. Ecology
12. Lost continent
Group 3:

13. Poetry American anthology
14. Pride and prejudice
15. In cold blood
16. Women and authority
17. Desserts
18. Basic statistics
Appendix B

The Evaluation Form

Figure 19 shows a sample evaluation form that each appraiser was asked to complete. Note that while Figure 19 includes only a single query, the actual form provided to the appraisers includes all of the six queries and their corresponding retrieved results presented on the screen of a computer monitor for which each appraiser was asked to evaluate.

Table 14 (Table 15, respectively) shows the top 10 library records retrieved by EnLibS (the HBLL system, respectively) for the query “African culture”. As shown in the tables, the appraisers were given a short description of each library record to aid them in making decisions on the relevancy and ranking of the retrieved results. Most of the descriptions were obtained from the HBLL catalog. When a short description was not available through the HBLL catalog, we extracted the descriptions from either Amazon.com (http://www.amazon.com) or LibraryThing.com (http://www.librarything.com).
Figure 19. An evaluation form used by the (2nd) group(s) of appraisers in our controlled experiment.
<table>
<thead>
<tr>
<th>Record #</th>
<th>Title</th>
<th>Author</th>
<th>Description</th>
<th>ISBN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Complexion of Race: Categories of Difference in Eighteenth-Century British Culture</td>
<td>Roxanne Wheeler</td>
<td>At the beginning of the eighteenth century skin color was far less important and far more diffusely related to group identity than were designations based upon religion and civility. Wheeler identifies the 1770s as the pivotal decade in this century's process of British racialization. From that point to the end of the century race became more rigidly defined and progressively more significant as a marker of British difference, especially as between Britons and Africans.</td>
<td>812217225</td>
</tr>
<tr>
<td>2</td>
<td>Color of Justice</td>
<td>Gary Hardwick</td>
<td>Raised in the inner city, Detroit homicide detective Danny Cavanaugh speaks and acts with the unmistakable attitude of a black man. But the savage murders of affluent African Americans are plunging him into the urban heart of terror, where he learns firsthand how powerful, inviolate--and deadly--the color line truly is</td>
<td>380818841</td>
</tr>
<tr>
<td>3</td>
<td>The Golden Road</td>
<td>Caille Millner</td>
<td>California saved Caille Millner’s parents, or at least saved them from lives of poverty and oppression as black Americans growing up in racially benighted backwaters. It provided them with a free education and opportunities for advancement into the solid middle class and even beyond</td>
<td>1594201099</td>
</tr>
<tr>
<td>4</td>
<td>Deals With the Devil: And Other Reasons to Riot</td>
<td>Pearl Cleage</td>
<td>Third-generation Black Nationalist writing to help herself understand the full effects of being black and female in a culture that is both racist and sexist. There is no garden club, country club, or investment club talk. There is, however, &quot;Basic Training&quot; talk: women--and men--are told the sexist and racist facts of life.</td>
<td>345481119</td>
</tr>
<tr>
<td>5</td>
<td>Saturday at the New You</td>
<td>Barbara E. Barber</td>
<td>Shauna, a young African American girl, wishes she could do more to help Momma with the customers at her beauty salon. Then one day she gets her chance</td>
<td>1880000431</td>
</tr>
<tr>
<td>6</td>
<td>African Americans and the culture of pain</td>
<td>Debra W. King</td>
<td>In this compelling new study, Debra Walker King considers fragments of experience recorded in oral histories and newspapers as well as those produced in twentieth-century novels, films, and television that reveal how the black body in pain functions as a rhetorical device and as political strategy</td>
<td>9780813926803</td>
</tr>
<tr>
<td>7</td>
<td>Encyclopedia of the African Diaspora: origins, experiences, and culture</td>
<td>Carole Boyce Davies</td>
<td>A three-volume encyclopedia set packed with over 500 entries, Encyclopedia of the African Diaspora discusses the global history of how people from Africa have spread out to other continents - some by choice, others through abduction and the slave trade. The resulting mix of global cultures would never be the same.</td>
<td>9781851097005</td>
</tr>
<tr>
<td>8</td>
<td>The jazz trope: a theory of African American literary and vernacular culture</td>
<td>Alfonso W. Hawkins</td>
<td>The author looks at the ways in which African American music such as jazz blues and spirituals reflect the African American experience and how scholars and writers have related music to African American life through the decades</td>
<td>9780810861268</td>
</tr>
<tr>
<td>9</td>
<td>African Americans and the culture of pain</td>
<td>Debra W. King</td>
<td>In this compelling new study, Debra Walker King considers fragments of experience recorded in oral histories and newspapers as well as those produced in twentieth-century novels, films, and television that reveal how the black body in pain functions as a rhetorical device and as political strategy</td>
<td>813926807</td>
</tr>
<tr>
<td>10</td>
<td>Black fascisms: African American literature and culture between the wars</td>
<td>Mark C. Thompson</td>
<td>In this provocative new book, Mark Christian Thompson addresses the startling fact that many African American intellectuals in the 1930s sympathized with fascism, seeing in its ideology a means of envisioning new modes of African American political resistance.</td>
<td>978081393926704</td>
</tr>
</tbody>
</table>

Table 14. Top 10 library records retrieved by EnLibS for the query "African culture"
<table>
<thead>
<tr>
<th>Record #</th>
<th>Title</th>
<th>Author</th>
<th>Description</th>
<th>ISBN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Transnationalism in southern African literature: modernists, realists, and the inequality of print culture</td>
<td>Stefan Helgesson</td>
<td>Considering the growing interest in South African Literature at the moment, this study looks at both the Anglophone literature of South Africa and the lusophone literature of Angola and Mozambique</td>
<td>4001586488</td>
</tr>
<tr>
<td>2</td>
<td>Encyclopedia of the African Diaspora : origins, experiences, and culture</td>
<td>Carole Boyce Davies</td>
<td>A three-volume encyclopedia set packed with over 500 entries, Encyclopedia of the African Diaspora discusses the global history of how people from Africa have spread out to other continents - some by choice, others through abduction and the slave trade. The resulting mix of global cultures would never be the same.</td>
<td>9781851097005</td>
</tr>
<tr>
<td>3</td>
<td>The jazz trope : a theory of African American literary and vernacular culture</td>
<td>Alfonso W. Hawkins</td>
<td>The author looks at the ways in which African American music such as jazz blues and spirituals reflect the African American experience and how scholars and writers have related music to African American life through the decades.</td>
<td>9780810861268</td>
</tr>
<tr>
<td>4</td>
<td>African Americans and the culture of pain</td>
<td>Debra W. King</td>
<td>In this compelling new study, Debra Walker King considers fragments of experience recorded in oral histories and newspapers as well as those produced in twentieth-century novels, films, and television that reveal how the black body in pain functions as a rhetorical device and as political strategy.</td>
<td>9780813926803</td>
</tr>
<tr>
<td>5</td>
<td>Black fascisms : African American literature and culture between the wars</td>
<td>Mark C. Thompson</td>
<td>In this provocative new book, Mark Christian Thompson addresses the startling fact that many African American intellectuals in the 1930s sympathized with fascism, seeing in its ideology a means of envisioning new modes of African American political resistance.</td>
<td>9780813926704</td>
</tr>
<tr>
<td>6</td>
<td>Solidarity, a principle of sociality</td>
<td>Sylvanus I. Nnoruka</td>
<td>Phenomenological-hermeneutical approach in the context of the philosophy of Alfred Schutz and an African culture</td>
<td>9783889398635</td>
</tr>
<tr>
<td>7</td>
<td>All bound up together : the woman question in African American public culture, 1830-1900</td>
<td>Martha S. Jones</td>
<td>This volume explores the roles black women played in their communities' social movements and the consequences of elevating women into positions of visibility and leadership. Martha Jones reveals how, throughout the 19th century, the &quot;woman question&quot; was at</td>
<td>9780807831526</td>
</tr>
</tbody>
</table>
the core of movements against slavery and for civil rights.


| 9 | Ordering the African imagination : essays on culture and literature | Tanure Ojaide | This collection of his essays and lectures from over the past decade, addresses issues of culture and literature from a personal African perspective. The focus of this book is African culture and its imaginative productions in the arts, especially in literature. | 9789780232047 |

| 10 | N.A. | N.A. | N.A. | N.A. |

Table 15. Top 10 library records retrieved by the HBLL system for the query "African culture"
Appendix C

Summary of the Relevance and Ranking of the Retrieved Results

Tables 16, 17, and 18 show the average ranking provided by the appraisers for each of the queries processed by their respective group (i.e., Group 1, 2, or 3). Table 19, on the other hand, shows the average number of relevant records marked by the appraisers in different groups. In the first column of Tables 16 - 18 (i.e., EnLibS/HBLL Ranking), it shows the corresponding ranking of the library records retrieved by EnLibS and the HBLL system, respectively. For Query 1 in Table 16, while the library record that EnLibS positioned first in the ranking, the appraisers in Group 1 positioned it in the third place, whereas for the library record that the HBLL positioned first, the appraisers in Group 1 ranked the library record in the 8th position.

Moreover, Table 19 details the average number of library records that were labeled as relevant by the independent appraisers in our study. Note that among 77%, i.e., 14, of the total number of queries, i.e., 18, EnLibS retrieved more relevant library records than the HBLL system, and the overall averages of the 18 queries are 6.7 relevant (out of 10 retrieved) library records (using EnLibS) versus 5.8 relevant (out of the 10 retrieved) library records (using the HBLL system).
Table 16. Average ranking provided by the appraisers in Group 1

<table>
<thead>
<tr>
<th>EnLibS/HBLL Ranking</th>
<th>Appraisers in Group 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Query 1</td>
<td>Query 2</td>
<td>Query 3</td>
<td>Query 4</td>
<td>Query 5</td>
<td>Query 6</td>
<td></td>
</tr>
<tr>
<td>EnLibS/HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 17. Average ranking provided by the appraisers in Group 2

<table>
<thead>
<tr>
<th>EnLibS/HBLL Ranking</th>
<th>Appraisers in Group 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Query 7</td>
<td>Query 8</td>
<td>Query 9</td>
<td>Query 10</td>
<td>Query 11</td>
<td>Query 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EnLibS/HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
<td>EnLibS</td>
<td>HBLL</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
<td>NA</td>
<td>5</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>4</td>
<td>NA</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>NA</td>
<td>5</td>
<td>NA</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>NA</td>
<td>5</td>
<td>NA</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>NA</td>
<td>5</td>
<td>NA</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>NA</td>
<td>9</td>
<td>NA</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>NA</td>
<td>6</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>
### Table 18. Average ranking provided by the appraisers in Group 3

<table>
<thead>
<tr>
<th>EnLibS/HBLL Ranking</th>
<th>Appraisers in Group 3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Query 13</td>
<td>Query 14</td>
<td>Query 15</td>
<td>Query 15</td>
<td>Query 17</td>
<td>Query 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EnLibS</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>HBLL</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>EnLibS</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>HBLL</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>EnLibS</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>HBLL</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>EnLibS</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>HBLL</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>EnLibS</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>HBLL</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Overall Average</td>
<td>6.7</td>
<td>5.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 19. Average number of relevant library records retrieved by EnLibS and the HBLL system, respectively according to the independent appraisers in our study

<table>
<thead>
<tr>
<th>Query</th>
<th>EnLibS</th>
<th>HBLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.0</td>
<td>5.9</td>
</tr>
<tr>
<td>2</td>
<td>3.5</td>
<td>5.0</td>
</tr>
<tr>
<td>3</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>4.3</td>
<td>6.5</td>
</tr>
<tr>
<td>6</td>
<td>7.5</td>
<td>7.3</td>
</tr>
<tr>
<td>7</td>
<td>6.9</td>
<td>5.6</td>
</tr>
<tr>
<td>8</td>
<td>9.3</td>
<td>7.5</td>
</tr>
<tr>
<td>9</td>
<td>6.0</td>
<td>2.5</td>
</tr>
<tr>
<td>10</td>
<td>5.6</td>
<td>2.4</td>
</tr>
<tr>
<td>11</td>
<td>8.3</td>
<td>7.6</td>
</tr>
<tr>
<td>12</td>
<td>6.1</td>
<td>7.4</td>
</tr>
<tr>
<td>13</td>
<td>8.5</td>
<td>8.0</td>
</tr>
<tr>
<td>14</td>
<td>7.5</td>
<td>7.4</td>
</tr>
<tr>
<td>15</td>
<td>3.5</td>
<td>4.7</td>
</tr>
<tr>
<td>16</td>
<td>7.0</td>
<td>6.9</td>
</tr>
<tr>
<td>17</td>
<td>9.6</td>
<td>6.6</td>
</tr>
<tr>
<td>18</td>
<td>7.6</td>
<td>7.4</td>
</tr>
<tr>
<td>Overall Average</td>
<td>6.7</td>
<td>5.8</td>
</tr>
</tbody>
</table>
References


[Novotny 2004] E. Novotny. I Don’t Think I Click: A Protocol Analysis Study of Use of a


