

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

10-20-2021

Evaluation of Scholarly Information Retrieval Using Precision and Recall

Kiran Butt

Deputy Manager, ISD, Riphah International University, Islamabad, Pakistan, littlescholar189@gmail.com

Abid Hussain

PhD Student at Department of Library and Information Science, Allama Iqbal Open University Islamabad, Pakistan, abid_as44@yahoo.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

Butt, Kiran and Hussain, Abid, "Evaluation of Scholarly Information Retrieval Using Precision and Recall" (2021). *Library Philosophy and Practice (e-journal)*. 6502.

<https://digitalcommons.unl.edu/libphilprac/6502>

Evaluation of Scholarly Information Retrieval Using Precision and Recall

Kiran Butt¹, Abid Hussain²

Abstract:

Introduction: An information retrieval system is widely used to locate the information for academic, research and reference purposes by the internet users in a library or Information services department.

Aims of Study: This study aims to analyze the search engine's behavior on a given query and results for scholarly information, followed by examining precision and recall.

Design/Method: The present study was conducted in a controlled environment where three search engines were selected to examine its recall and precision against 50 queries with basic and advance searching queries. The queries were executed in all search engines. This quantitative study was used on the General, and Subject focus search engines.

Finding(s): The finding of study shows that the precision and recall of FindLaw, was highly followed by Lycos with medium and Yandex the low. Further study can explore the precision and recall of other searching engines with different subject.

Research Limitation(s): The study highlights the retrieval effectiveness of three search engines in a specific subject.

Originality/value: The research work is authentic and does not contain any plagiarized work.

Keywords: Precision & Recall, Advance Search Queries, Search Engines-behavior, Information Retrieval.

Introduction

Information retrieval is the process of obtaining information resources that are relevant to an information need from a collection of those resources (Deo, Gangrade & Gangrade, 2018). Searches can be based on full text and content indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also

¹Deputy Manager, ISD, Riphah International University, Islamabad, Pakistan little scholar189@gmail.com

² PhD Students, DLIS, Allama Iqbal Open University, Islamabad, Pakistan abid_as44@yahoo.com

searching for describes data, from database of texts, images, sounds or videos (Agboola & Shaibu, 2019).

Jones & Kelly, (2018) reported that automated information retrieval systems are used to reduce what has been called information overload. The information retrieval system is a system that provides access to research journals, documents, books and other literature available in a database; stores and manages those documents. The currently used search engines (i.e., Google, Yahoo, Bing, OneSearch etc) are the most visible information retrieval application. There are many search engines on the web helping the community in finding their required information. The search engines save the time of users and bring most relevant results in a variety of formats including documents, articles, books, videos, audio, images, and text (Uyar, & Aliyu, 2015).

The search engines are classified into three main types called Crawler based search engines, Human-powered directories, and Hybrid search engines. Furthermore, advanced technologies introduced subtypes of the searching tools such as general-purpose searching engines like Google, Bing, Yahoo, etc. Meta Search engines i.e., Skyscanner, Kayak.com, and subject focus searching engines, RefSeek, and WolframAlpha (Hussain, 2015). The internet user is not getting desired and appropriate results quickly against his submitted query. There is a flood of information available on web, few is relevant and few irrelevant. Millions of information is available online, some is relevant while, some may be irrelevant in the response of a user query (Jan, et al, 2018; Hussain, et al, 2020).

To find desired pieces of information among the mass results, it is difficult for ordinary users as well as information technology experts. The performance of search engines is improving day by day. Some of the searching engines used semantic and artificial intelligence web technologies either partially or as much as possible (Usmani, Pant, & Bhatt, 2012). The internet is an educational tool of the day. The searching engines are a significant tool to find most relevant information in a short time. Many libraries are using computer software for cataloging the library material that is specially designed for searching and retrieving relevant information.

After a comprehensive review of the literature, no such study was found that analyzed the precision and recall of searching engines using the sample, and complex query in the field of law. The study results will be helpful for law graduates, practicing advocates, and librarians working in law colleges, departments, and courts at country level.

Review of Relevant Literature

The review of literature is a significant part of the research process and identifies the area that needs to be addressed by the researcher. A review of literature provides a pitch to the current study about what has been already done by the scholars (Liang, 2020). It has also pointed the method, population, and strategies employed by the investigators for the same nature of studies. To collect the relevant literature on this topic of precision and recall of search engines a dedicated search of different databases, research journals, and digital libraries was conducted. The most relevant and latest studies were included in this section.

Kanwal Ameen and Midrar Ullah claimed that the rapid emergence of information and communication technology is a need for library and information services and is important for users too. The development of internet learning resources compels its users for information literacy skills programs. Effective information literacy programs can make the users of online resources independent in optimizing the searching and retrieving of information by using searching strategies for accurate and precise results (Ameen, & Ullah, 2016).

The topic of information retrieval is predefined, and the evaluation was carried out by using the standard system for retrievals metrics of precision and recall. The documents are regarded as information items that consist of text, in any natural language. They have a known length, and their content serves as the source of indexing features (Järvelin, 2007).

A study conducted by Ali, Sabha, and Gul, Sumeer, highlighted the retrieval effectiveness of search engines taking into consideration both precision and relative recall. Google.com is the most visited search engine, and Yahoo.com is the fourth most visited site globally. A total of 15 queries were selected randomly from Library and Information Science (LIS) in 2014 and classified manually into different queries. The analysis revealed that the precision and relative recall of Google and Yahoo was on top. The queries using concepts in the field of LIS were tested in this analytical study. Results showed that the precision of Google was high with (mean =1.10) followed by Yahoo with (mean = 0.88). This study analyzed only two searching engines and recommended that researchers can analyze more search engines in further studies (Ali, & Gul, 2016).

Arora, Monika, Uma Kanjilal, and Dinesh Varshney have evaluated the precision and recall of information retrieval systems by targeting Yahoo and Google. The authors stated that the significance tests are used to evaluate functional, performance (precession and recall), collection, and interface evaluation. The response time and the relevancy of the results of Google and Yahoo were significant. Google is the best alternative for getting web-based scholarly

documents, and Yahoo offers a good combination of recall and precision. The findings of this study established the case that precision is inversely proportional to recall (Arora, Kanjilal, & Varshney 2016).

In the evaluation of an information retrieval system, the system uses a precise answer to a well-formulated query from a structured database (Zoghi et al., 2014). The Information retrieval system also assists the users in satisfying the information needs by interpreting his information needs and providing the information items that are relevant to them. There is a great diversity in the users need ranging from answers to precise questions, requiring specific information to broad navigation of information items.

The search requests or topics and the documents are unstructured. The natural language text internally takes care to represent information needs (Radlinski & Craswell, 2013). They are associated with independent indexing features that derive from their content through natural language processing. These indexing features are usually wording whose semantics help to describe the document's main themes and summarize its content. Precision is the relationship between the number of retrieved relevant documents concerning to a query statement and the number of documents that have been retrieved based on it (Belew, 2001).

Objectives

The following objectives are set to study precision and recall of three search engines:

1. To pinpoint the search engines for relative information in the field of Law
2. To analysis the Precision of selected three search engines
3. To find out Recall of selected three search engines

Methods and Materials

The methodology is based on following step which enable the authors to complete all prerequisites for the study and then practically executed the formula and exam the defined environment.

1. Study of Related literature which was available on the same topic.
2. Three search engines were selected (Two General,One of Specific Subject)which is enlisted below:
 - a) <https://www.lycos.com/>(General Search engine)
 - b) <https://www.yandex.com/>(General Search engine)
 - c) <http://public.findlaw.com> (Law Subject Specific)

3. Selection of field (Subject area) from which the relevant key words chosen for construction of queries, (Filed selected = LAW).
4. Searching queries were developed from an online glossary of Law. A total of 50 (25 Single words & 25 phrase) were selected for the study in hand attached as “Annexure 1”.
5. To check how search engines hits and gives the results. The simple and advance search methods was used by including Boolean operators(AND, OR, NOT), Wildcards & Truncation (*, ?), Double-quote (“ ”) simple word and phrase without any addition of mention above operators.
6. Each query was submitted to the selected engines which retrieved a large number of results but only the first five results were evaluated to limit the study in view of the fact that most of the users usually look up under the first ten hits of a query which is displayed by default for user from search engines.

Calculation of Precision: The precision is the fraction of a search output that is relevant for a particular query. Its calculation, hence, requires knowledge of the relevant and non-relevant hits in the evaluated set of documents (Clarke & Willet, 1997).

Thus, it is possible to calculate absolute precision of search engines which provides an indication of the relevance of the system. In the context of the present study precision is defined as:

$$\text{Precision (Pr)} = \frac{\text{Sum of the scores of information retrieved by a search engine (X)}}{\text{Total number of results evaluated (Y)}}$$

Estimate Pr = f $\frac{X}{Y}$

To determine the relevance of each page, the following scale was devolved and used which enabled authors to calculate precision.

Table 1: Scale for Calculating of Precision

S.No	Results in 1st five hits by Search Engines	Score	Rank
1	Books, Journals, Proceeding, Case Law and Court Decision	3	High
2	Reference Sources i.e., Glossary etc.	2	Medium
3	Online Sources	1	Low
4	Others then above	0	Zero

As shown on Table 1 when a query displays a list of hits which include Books or Journaletc. will be given high score (3), a hit which include any reference sources in the response as a hit will be score medium (2) followed by online hit (1) and others will be score as zero (0).

The above-mentioned scale will help the researchers to calculate the scores of searching engines for each single and phrase terms.

Calculation of Recall: The recall on the other hand is the ability of a retrieval system to obtain all or most of the relevant documents in the collection. Thus, it requires knowledge not just of the relevant and retrieved results but also of those not retrieved. There is no proper method of calculating absolute recall of search engines as it is impossible to know the total number of relevant in huge databases.

However, Clark and Willett, (1997), have adapted the traditional recall measurement for use in the web environment by giving it a relative flavor. This study also followed the method used by Clark and Willett by pooling the relevant results (corresponding here to scholarly documents) of individual searches to form the denominator of the calculations.

The Relative Recall Value is defined as below:

$$\text{Recall (Re)} = \frac{\text{Total number of hits retrieved by a search engine (X)}}{\text{Sum of retrieved by all Three search engines (Y)}}$$

$$\text{Estimate Re} = \mathbf{f} \frac{X}{Y}$$

Analysis of the Results

The data was retrieved from two general, and one subject domain searching engines. The collected data was analysed by using Microsoft Excel Sheets and the data was put to formula which was formulated for digging of results for finding the relative recall and precision in the field of Law.

Table 2: Precision of Three Search Engines

S.No	Search Engine(s)	$\sum Xi$	$\sum Yi$	Percentage	Position
1	Lycos	377	750	50.27	2nd
2	Yandex	349	750	46.53	3rd
3	FindLaw	403	750	53.73	1st

The analysis of the table 2 shows that the precision of search engines with its positions. The *Findlaw* is on the top with (53.73%) and got 1st position, followed by ‘Lycos’ (50.26%) 2nd,

and ‘Yandex’ got 3rd position with (46.53%). It is noteworthy that Find Law—A subject search engines is providing precise results for the structure and unstructured queries.

Result of Single Words and Phrases

The figure 1 (below) shows the separate result for single words and phrases, *FindLaw* was given the good results and marked highest in single word query as well as in phrases precision and *Yandex* has been marked “lowest” for the less numbers of results in the field of law.

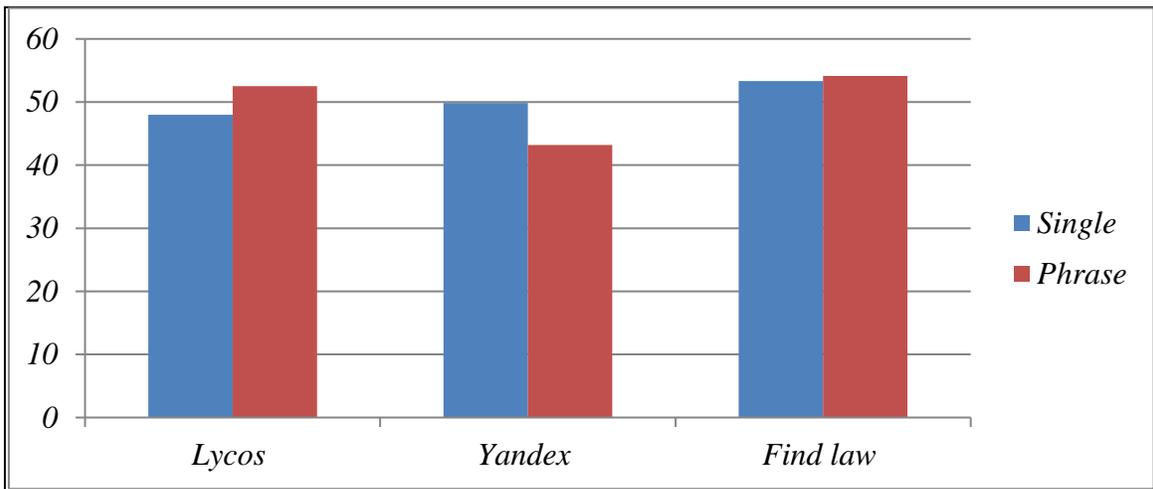


Figure 1 Precision of Searching Engines for Single and Phrase Query

Table 3: Recall of Selected Searching Engines

S. No	Search Engines	Recall (%age)	Cumulative Percentage
1	Lycos	33.85	33.85
2	Yandex	30.77	64.62
3	FindLaw	35.38	100

The analysis of table 3; driven the recall for three selected searching engines. The *FindLaw* (A subject focus search engines) is achieved 35.38% is the best recall as per study conducted in controlled environment, followed by *Lycos* 33.85%. The low 30.77 % recall was recorded for *Yandex*. It’s amazing to note that the percentage of results near with each other’s with a little gap from 33.77 % to 35.38% consequently.

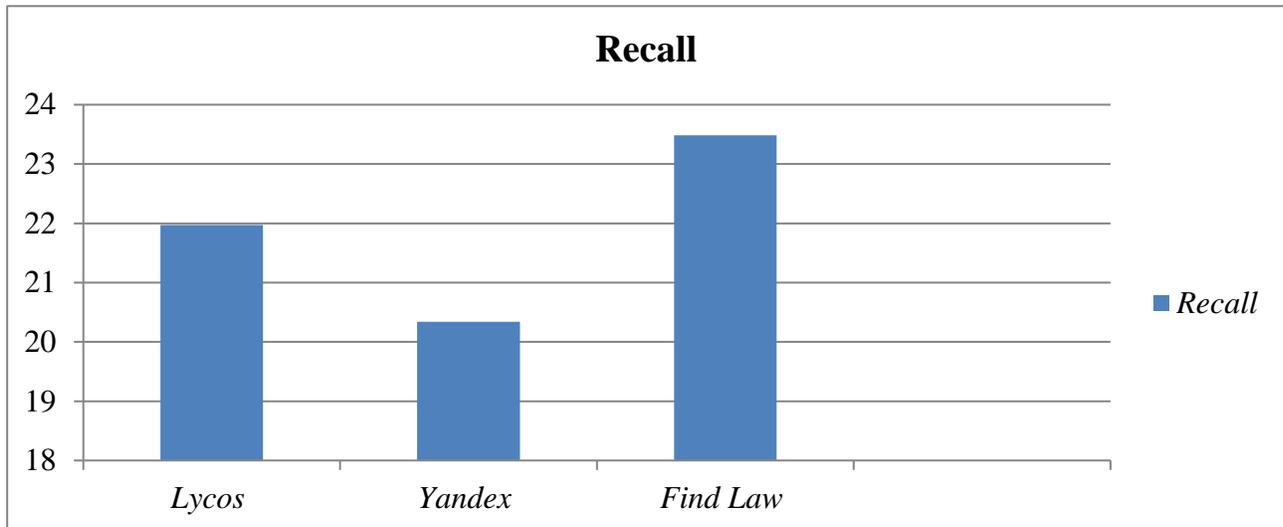


Figure 2, Recall of Three Search Engines

The relative recall for the terms attached as annexure 1 was searched using the selected three searching engines for the calculating the recall has presented on Figure 2. The results revealed that 'Find Law' was higher among all followed by Lycos in the response of queries for recall.

Discussion

A record is considered relevant or irrelevant when calculating recall and precision of scholarly information. There may be slightly relevant or somewhat irrelevant, maybe very relevant, and completely irrelevant records while searching for information needs by a user of the library. A recall is difficult to measure because it is often unknown that how many relevant records are available in the database. The recall is often estimated by identifying a pool of relevant records and then determining how many searches have been performed for its retrieval.

The study was conducted for the evaluation of retrieval using three different search engines. That includes Find Law, Lycos and Yandex. This study is used to find out the occurrence of the relevant records in the first attempts of the search engines. Relevance of the top five hits was determined as pre-define criteria on the results of selected searching engines. Practical searches were performed in a controlled environment on all searching engines for calculating recall and precision. Basic and advance searching queries were used by the authors for recording the results. The results show that 'Yandex' was low in term of relevant recall of information need and top position was chilled by 'Find Law'.

The recall and precision are invaluable to any experienced searcher. Knowing the goal of the search; to find everything on a topic, just a few relevant documents, or something in-between; to determines that what strategies the searcher will use. There is a variety of searching techniques may be discussed for future work, which may be used to affect the level of recall and precision. A good searcher must be adopting the searching technique and properly using them.

Conclusion

The current study was aimed to analyze the search engine's behavior on a given quarry and results for scholarly information, followed by examining precision and recall. The study was conducted in a controlled environment where five search engines were selected to find of its recall and precision. A set of 50 queries with basic and advance searching queries were executed on search engines. The major finding of the study revealed that the precision and recall of Find Law is high followed by Lycos with medium and Yandex is low. Based on the analysis the authors suggested that furtherstudy can explore the precision and recall of other searching engines with a different subject area.

References

- Agboola, B., &Shaibu, R. (2019). Impact of ICT on Information Retrieval System in Academic Libraries: The Experience of Federal University Gashua Library, Yobe State, Nigeria. *Library Philosophy and Practice (e-journal)*, Paper no 2350, 1-15. <https://digitalcommons.unl.edu/libphilprac/2350/>
- Ali, S., & Gul, S. (2016). Search engine effectiveness using query classification: a study. *Online Information Review*, 40(4), 515-528.
- Ameen, K., & Ullah, M. (2016, October). Information literacy instruction: An overview of research and professional development in Pakistan. In *European Conference on Information Literacy* (pp. 555-562). Springer, Cham.
- Arora, M., Kanjilal, U., & Varshney, D. (2016). Evaluation of information retrieval: precision and recall. *International Journal of Indian Culture and Business Management*, 12(2), 224-236.
- Belew, R. K., &Belew, R. K. (2000). *Finding out about: a cognitive perspective on search engine technology and the WWW*. Cambridge University Press, New York, USA.
- Clarke, S. J., & Willett, P. (1997, July). Estimating the recall performance of Web search engines. In *Aslib proceedings*. MCB UP Ltd.
- Deo, A., Gangrade, J., &Gangrade, S. (2018). A Survey Paper on Information Retrieval System. *International Journal of Advanced Research in Computer Science*, 9(1), 778-781.

- Hussain, A. (2015). Search Engines as an Effective Tool for Library Professionals. *DESIDOC Journal of Library & Information Technology*, 35(6), 389-397.
- Hussain, A., Ibrahim, M., Saeed, S., & Salaam, A. (2020). Examining Usability of the Internet by the Teaching Staff of the Peshawar Medical College, *Library Philosophy and Practice (e-journal) Paper No 5007, 1-10*.<https://digitalcommons.unl.edu/libphilprac/5007/>
- Jan, S., Hussain, A., Ibrahim, M., & Saeed, S. (2018). Use of internet by the teaching faculty of Peshawar Medical College, Peshawar, Khyber Pakhtunkhwa, Pakistan. *Journal of the Pakistan Medical Association*, 68(3), 459-462.
- Järvelin, K. (2007). An analysis of two approaches in information retrieval: From frameworks to study designs. *Journal of the American Society for Information Science and Technology*, 58(7), 971-986.
- Jones, S. L., & Kelly, R. (2018). Dealing with information overload in multifaceted personal informatics systems. *Human-Computer Interaction*, 33(1), 1-48.
- Liang, X. (2020). Internet of Things and its applications in libraries: a literature review. *Library Hi Tech*, 38(1), 67-77.
- New York State Court System- NYSCS (2021). Legal Glossary, available online at: <http://www.nycourts.gov/lawlibraries/glossary.shtml> accessed on August 25, 2021.
- Radlinski, F., & Craswell, N. (2013). Optimized interleaving for online retrieval evaluation. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 245-254). DOI: <https://doi.org/10.1145/2433396.2433429>
- Usmani, T. A., Pant, D., & Bhatt, A. K. (2012). A comparative study of Google and Bing search engines in context of precision and relative recall parameter. *International Journal on Computer Science and Engineering*, 4(1), 21-34.
- Uyar, A., & Aliyu, F. M. (2015). Evaluating search features of Google Knowledge Graph and Bing Satori. *Online Information Review*, 39(2), 197-213.
- Zoghi, M., Whiteson, S., de Rijke, M., & Munos, R. (2014). Using confidence bounds for efficient on-line ranker evaluation. *WSDM*, 14, 24.

Annexure 1

Search Queries for Precision and Recall

1. One-word queries (25 Words)

- 1.1. Law
- 1.2. Legal
- 1.3. Acknowledge*
- 1.4. Affidavit
- 1.5. Amend?
- 1.6. Calendar
- 1.7. Chattel
- 1.8. "Citation"
- 1.9. Complaint
- 1.10. Decision?
- 1.11. Decree
- 1.12. Defendant
- 1.13. Dismissal
- 1.14. "Docket"
- 1.15. Easement
- 1.16. Evidence
- 1.17. Execution
- 1.18. Exhibit?
- 1.19. Forum
- 1.20. Garnish
- 1.21. Heari*
- 1.22. Impleader
- 1.23. Injunction
- 1.24. Interpreter
- 1.25. *Jurisdiction

2. Multi-word queries (25 Phrases)

- 2.1. "Bifurcated-trial"
- 2.2. Bill of costs
- 2.3. Case-file NOT files
- 2.4. Certified copy
- 2.5. "Charge to jury"
- 2.6. Common law
- 2.7. Court AND reporter
- 2.8. Cross AND claim
- 2.9. Directed OR verdict
- 2.10. Dissolution of marriage
- 2.11. Equitable OR distribution
- 2.12. Fair preponderance?
- 2.13. First paper NOT paper
- 2.14. Habeas corpus
- 2.15. Index-number NOT Index
- 2.16. Endorsed AND complaint
- 2.17. Inter AND alia
- 2.18. Joint-trial NOT Single
- 2.19. Judgment OR roll
- 2.20. Jury-instructions NOT Jury
- 2.21. Legal age?
- 2.22. Natural person*
- 2.23. Opening statement
- 2.24. Peremptory challenge
- 2.25. Show caus*