Classical and Probabilistic Information Retrieval Techniques: An Audit

Khawar Iqbal Malik[1], Qaiser Abbas[2], Hira Arooj[3], Sadia Niazi[4], Yousaf Saeed[5], Ijaz Asghar[6], Muhammad Ilyas[7]

[1]Department of Computer Science & IT, The University of Lahore, Sargodha, Pakistan

[2,7]Department of Computer Science & IT, University of Sargodha, Sargodha, 40100, Pakistan

[3]Department of Mathematics & Statistics, The University of Lahore, Sargodha, Pakistan

[4]Department of Psychology, University of Sargodha, Sargodha, 40100, Pakistan

[5]Department of Information Technology, University of Haripur, Haripur, Pakistan

[6]Department of English, University of Sargodha, Sargodha, 40100, Pakistan

Email: qaiser.abbas@uos.edu.pk

(Received: 24 July 2021 ; Accepted: 10 Sep 2021 ; Issue Published: 12 Sep 2021)

ABSTRACT

Information retrieval is acquiring particular information from large resources and presenting it according to the user's need. The incredible increase in information resources on the Internet formulates the information retrieval procedure, a monotonous and complicated task for users. Due to over access of information, better methodology is required to retrieve the most appropriate information from different sources. The most important information retrieval methods include the probabilistic, fuzzy set, vector space, and boolean models. Each of these models usually are used for evaluating the connection between the question and the retrievable documents. These methods are based on the keyword and use lists of keywords to evaluate the information material. In this paper, we present a survey of these models so that their working methodology and limitations are discussed. This is an important understanding because it makes possible to select an information retrieval technique based on the basic requirements. The survey results showed that the existing model for knowledge recovery is somewhere short of what was planned. We have also discussed different areas of IR application where these models could be used.

KEYWORDS: Information Retrieval, Vector Space Model, Boolean Model, Probabilistic Models, Indexing Searching, Inference Network Model.

1. INTRODUCTION

The collection of information is usually considered an informatics branch that manages the representation, access and store the information. Information processing involves the management and compilation of data from large repositories of data [1]. Information Recovery (IR) is a technique for describing, processing and gathering information in response to a user request (quest) for the exploration of knowledge [2]. The cycle starts with the user question in several steps and concludes with sufficient user details. Filtering, scanning, matches of document and rating procedures are also remaining phases. The primary aim of the information retrieval is to bring out facts or documents that meet the users' data needs. IR systems (IRS) typically implement the following processes to achieve this objective [3]:

- Indexing process: The presentation of documents is in summary form.
- Filtering process: Removal of stop words and common words, known as a linguistic module.
- Searching process: Main route of IRS contains several methods for retrieval of records that are suited with the query.

There are two basic steps to determine the efficiency of retrieval of information [4].

- Precision: That is the proportion of documents obtained that are applicable to the question.
- Recall: It might be a number of records relating to the application that has already been retrieved.

$$Precision = \frac{No.of\ Relevant\ Document\ Retrieved}{No.of\ Document\ Retrieved}$$

$$Recall = \frac{No.of\ Relevant\ Document\ Retrieved}{No.of\ Relevant\ Document}$$

The method of retrieval of information is listed below:

- The knowledge recovery process begins when a user uses any graphical or interface to construct some question in the framework
- Such user-defined requests provide, for example, descriptions of the information needed by search engine users.
- Doesn't IR single-question suit to appropriate data-object instead of the various sets of data-objects that consider the most important item for further evaluation?

- The appropriate documents are ranked to determine the most important document for the particular query. This is the main difference between the search and the retrieval of information.
- Submit it to the core of the device after the question. This section has access to the contents management module, directly linked to the backend i.e., the large data object collections.
- Once the results of the core system are created, some GUI returns them to the user.

The cycle continues and outcomes are changed before the user fills up with what he really wants. The following figure sketches the processing of textual queries performed by an IR system.

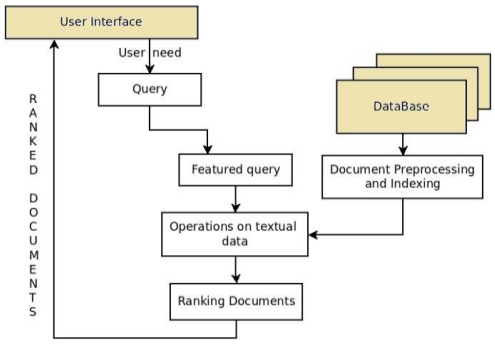

Figure 1: Information Retrieval System Process

This paper is organized as follows. Section 2 and 3 offer a brief introduction to IR models, including Boolean, VSM and Probabilistic. In Section 4 and 5, indexing and searching techniques for these models has be discussed. Section 6 covers the area where these models are used as commercial applications.

2. IR MODELS

There are two explanations why information retrieval models are available. The first is that models provide direct analysis and the basis for scholarly debate. Secondly, models may be a guide for implementing an actual retrieval system. The specifications of representing the document, representing the query and recovery features are given by an IR model [5].

The key to IR models might be categorized as boolean, vector space and probabilistic models [5] [6]. The remainder section is the discussion of these models briefly.

2.1. Boolean Model

The Boolean model is solidly grounded in science due to its natural utilization of archive sets, which gives an incredible method for data recovery. The key drawback of the Boolean model is the incapacity to rate documents. For most recovery applications, positioning is critical and positioning augmentations have been proposed in the Boolean model [7]. These augmentations depend on those models that accept the requirement aimed at positioning through their beginning stage. The Boolean model takes into consideration the operations of Boolean algebra including AND (&&), OR (||) and NOT (!), for inquiry. In the Boolean model, an archive is related with a lot of watchwords. Inquiries are likewise articulations of catchphrases isolated by AND (&&), OR (||), or NOT (!). The recovery work in stated model treats a record as important otherwise insignificant. Figure 2 demonstrates some of the problems of the Boolean recovery methods by the shaded regions.

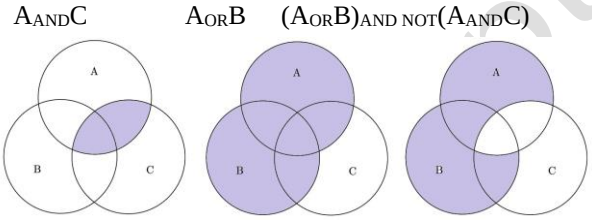

Figure 2: Venn Diagram: Boolean Mixtures of Sets

2.2. Vector Space Model

Gerard Salton besides his partners, proposed a model that depends on Luhn's likeness that takes a more grounded hypothetical inspiration [7]. We considered a list and the investigation as implanting vectors in a Euclidean space of high dimensions where each term is calculated differently. The vector space model is finely defined by its attempts to identify records by comparing questions with each database [6]. In the vector space model (VSM), records in addition to question are expressed by way of a vector and the point among the two vectors by utilizing the closeness of the cosine method. The similarity of Cosine is defined as.

$$Sim(d_j, q) = \frac{d_j \cdot q}{\|d_j\| \cdot \|q\|} = \frac{\sum_{i=1}^{N} w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^{N} w_{i}^{2},_{j}} \cdot \sqrt{\sum_{i=1}^{N} w_{i}^{2},_{q}}} \qquad (1)$$

Where vectors are records and inquiries.

$$dj=(w_{1j},w_{2j}, \ldots\ldots, w_{tj})\ q = (w_{1q},w_{2q}, \ldots\ldots, w_{tq})$$

The vector space model presents the term weight plan studied through idf weighting. These loads need a term recurrence (tf) factor, which estimates the recurrence of an event of archive or inquiry. A backwards record recurrence (idf) factor estimates the converse of the quantity of reports that encompass a question or else archive-term.

2.3. Probabilistic Model

Though Maron and Kuhns presented positioning via the likelihood of significance, it was Stephen Robertson who transformed the thought hooked on a guideline. He figured the likelihood positioning rule, which he credited towards William Cooper, equally pursued by Robertson in 1977 [8]. For the probabilistic model, the most important standard is its attempt to identify reports by their probability of pertinence. Records and questions are answered by vectors ~d and ~q, each part of the vector indicating whether a record property or word exists in the archive or inquiry. Rather than probabilities, the probabilistic model uses chances O(R),

Where O(R)=P(R)/1-P(R), R implies "document is relevant" and $\bar{R}$ implies "document isn't relevant".

For the determination about the significance of documents with respect to a query, both these models use statistical information in the form of term frequencies while they vary in the way they use the word frequencies. A brief comparison of VSM and probabilistic models is presented here in Table I along with their motivation, goals, and issues.

Table 1. Parametric Comparison of IR Models

Parameters	Information Retrieval Classical Models		
	Boolean Model	**VSM**	**Probability Model**
Orientation	Query oriented, and work on extract match	Partial match	Partial match
Elementary method	Conventional	Non conventional	Non conventional
Optimum	Performance is not prime	Performance is optimal	Performance is provided to be optimal
Set Theory	Weights binary, Extended boolean case based, Fuzzy set,	Vector space, Neural networks Generalized vector space, Latent semantic indexing,	Probabilistic, Inference Network, Belief network

3. TYPES OF PROBABILITY MODELS

3.1. Principle of Probability Rank Model (PRP)

The probabilistic retrieval model is built on the probability classification theorem, which says that the document retrieval system based on its likelihood of relevance to the query, will evaluate the documents in the light of all the available data [9]. The theory considers that there is confusion regarding the need for information and documentation. The probabilistic methods of retrieval can use a range of data sources and the most common method is the statistically relevant and non-relevant distribution of words.

By assuming the independence between the query terms the model will be:

$$\log\big(p(Rel|d,q)\big) = \sum_{t\in q} \log \frac{p(t|Rel).p(\bar{t})\big(\overline{Rel}\big)}{p(t|\overline{Rel}.p(\bar{t}|Rel)}$$

In this probability *p(Rel|d,q)*, *Rel* symbolizes the event of a document *d* existence relevant to a query *q*.

3.2. The Binary Independence Retrieval Model

The vital principle of the classic approach to probabilistic information retrieval is a relatively modest model i.e., the so-called BIR. In all the following sections, precise assumptions underlying this model are discussed.

In the BIR model, we need to measure the likelihood for a given document as in most other possibly IR models. The d_m unique query is tested with respect to a specific question q_k. The estimation of this probability is $p(R|d_m, q_k)$. The underlying principle is that words are differently distributed in related and non-related documents. This theory is called the hypothesis of clusters. Let $T = \{t_1, \dots, t_n\}$ distinguish the set words. Then we can represent the set of terms d_m^T occurring in document d_m as a binary vector $x=x_1 \dots x_n$ with $x_t = 1$, if $t_i \in d_m^T$ and $x_t = 0$ otherwise. We used two transformations, also used for deriving probabilistic IR models, to establish a formula for this probability [10]:

- Bayes' theorem test (in the form $P(a|b) = P(\frac{b|a)P(a)}{P(b)})$
- Use probability rather than odds where $O(y) = \frac{P(y)}{P(\bar{y})} = P(y)/[1-P(y)]$

This helps us to measure the chances of a binary vector x data that is important for the question query q_k as

$$O(R|q_k, x) = \frac{P(R|q_k, x)}{P(\bar{R}|q_k, x)} = \frac{P(R|q_k)}{P(\bar{R}|q_k)} . \frac{P(x|Rq_k)}{P(x|\bar{R}q_k)}$$

3.3. The 2-Poisson Model

There is a specific model relating to the representational essentials. It is affecting pattern of 2-Poisson. Another methodology, like the index model, attempts to determine whether or not an index word is intended for a document [11]. For a specific word, there are two types of documents. The number of incidences tf_{tm} of the term t_t inside the document d_m is observed with the assumption of distribution in the two document groups of this function as different. For a specific document class k_{ij}, μ_{ij} is the expectation of t_t with the probability that a document contains l occurrences as follows.

$$P\big(tf_{tm} = l \big| d_m \epsilon k_{ij}\big) = \frac{\mu_{ij}^t}{l!} e^{-\mu_{ij}}$$

Two document classes $k_{t1}$and $k_{t2}$for each term are there, so $\pi_{t1} + \pi_{t1} = 1$ in the 2-Poisson model. The probabilistic index term weighs $P(d_m \in k_{ij}|tf_{tm} = 1)$, which can be derived. The parameters π_{ij} and μ_{ij} can be estimated without feedback information from the document collection [12].

3.4. Bayesian Inference Netwrok

If IR is considered an unsafe inference, the structure of document-to-quest comparisons is more complicated than the relevant model.

Historical recovery is seen in this model as a guessing process in an induction [13]. The majority can be conducted according to this model using IR frameworks. In simpler implementation, a document installs a term of a certain standard and credit is granted to find out what can be compared to a statistical value for the archive in specific words. Since an operational viewpoint, the instantiation

quality can be considered as the heaviness of the word in the study, and the positioning of the archive (within a less demanding model) ends up as positioning in the vector space model and probabilistic models. The quality of a recording term instantiation does not depend on the layout, and any method can be used.

Document node, query or definition may accept the true or false value of each node. We presume contrary to the models that there are two types of concepts, namely t_t and query concepts r_t. The network's centered arcs reflect probabilistic node dependency. A node's likelihood depends only on its parentage values. This relation must be established in the node as a function., The document node is set to' true' to estimate the probability of $P(d \rightarrow q)$and then the probabilities of the nodes are determined before P(q= true) is obtained. According to the node's combination feature, the inference network can involve multiple nodes, such as boolean links and probabilistic correlationsas in different models [14]. Suppose the idea of representation is a clear example r_1 is 'IR' which is defined as an OR-combination of the two document concepts (terms) t_1 = 'information retrieval' an t_2 = 'document retrieval'. Then the probability of r_1 being true can be computed by the function $P = (r_1 = true) = 1 - (1 - P = (t_1 = true))(1 - P(t_2 = true)$

This method provides certain benefits as compared to the model of interest, as the maximum probabilistic IR models can be transformed into a network of inferences. Unlike these, the network method does not allow a closed probabilistic formula to be extracted, and hence uses more dynamic interdependencies. The network methodology allows multiple pieces of evidence to be integrated. For example, data on text similarities and information from external sources such as a thesaurus may be considered.

The Boolean model is based on set theories and is represented as Boolean form. In contrast to Boolean, the vector space queries and documents are represented as vectors in a certain dimension spacehere dimensions are the most similar terms, and the report is found in space [15]. The probabilistic approach assumes that each question receives an ideal response and is guided by partial matching using the non-conventional methods and tests.

Table 2. Comparison of Different IR Models over specific Criterion

Criterion	Boolean Models	Vector Space Model	Probabilistic Models
Query Matching	Query-oriented, exact match	Partial match	Partial match
Basic Approach	Conventional	Non-conventional	Non-conventional
Recall	All or No document retrieved due to perfect match only	Due to 'partial matches' and term weight, higher recall rate than the Boolean model. Ever not possible to find out the weights of words in changing repository, the recall rate may be negative.	Better recall other two approaches
Precision	All or No document retrieved	Retrieved documents depends on terms weight w.r.t other documents & query	Retrieved document based on the likelihood of a phrase appearing in a single document versus the same term appearing across the whole corpus
Working	It processes queries as Boolean statements processed	Uses indexed weights and partial matching.	Used the optimum set of probabilistic index words.
Representation	Binary weights	Weighted Index terms	Binary weights. Text classified as either belonging to ideal set or being irrelevant.
Type of Information	Semantic information not observed	Semantic information's observed during retrieval	Semantic information's observed during retrieval
Word Contingency	Not provided the number of word occurrence	Have info regarding No. of word occurrence	Word occurrence info available in matrix
Output	Strict match	Best match	Best match
Pros and cons	It's simple but in some cases not ranked the documents	Easy to understand but more complex than binary Model	Not restricted to words only since it replaces 'keywords' by 'concepts' but most complex model

In above Table II, a comparison is presented for all three types of models on some specific criterions. According to these comparisons, we have seen that each of the 3 information retrieval models has its own pros and cons. The Boolean model works on accurate match, but it does not provide an answer when the user is unsure of what he wants to retrieve. Similarly, the VSM and the probabilistic model both have their own unique method of achieving partial matching in data acquisition [16].

It's an important element of search engines to offer a diversity of outcomes results to users because it is impossible to determine the user's specific intent all at once. The idea that the browser are not built in an orderly form and is currently not well-ordered is often stressed by researchers [17].

4. INDEXING TECHNIQUES

Some common techniques for indexing knowledge recovery, including reversed indices and signature files are available.

4.1. Signature Files

Each document gives a bit ("signature") of a signature file method by hacking its words and coding overlaid. The results of the paper signatures are stored sequentially in the separate section called the Signature Log [18].

4.2. Inversion Indexes

The list of keywords defining the document content for recovery purposes that be described in each document [19]. Quick recovery is possible if we reverse these keywords. Keywords can be saved, for example alphabetically.

We maintain a list of indicators on the eligible documents in each keyword's postings file and in the index file. Nearly all commercial systems follow this approach.

5. SEARCHING TECHNIQUES

Present are many searching algorithms, including linear search, brute force search, binary search, and several more:

5.1. Linear Search

A method is used to put in a linear search algorithm that regulates all items of the list, by one sequence, and locates list or sequence-unique feature or keyword. Linear search is the most simple search algorithm. The slow search speed in the ordered list is one of most important drawbacks. This search is often known as sequential search.

5.2. Brute force

Brute force exploration is an overall problem-solving technology consisting of listing with all potential solution candidates and checking whether any candidate meets the requirement of the problem. Gross force algorithm is simple to use and always seeks a solution if it is necessary.

5.3. Binary Search

Binary search algorithm identifies the element's key match. A matching item is identified, which returns its index or location. Else if search key is smaller than the mid-element value, the algorithm replicates its function in the left-hand mid-element sub-array or similarly, in the right sub-array when the value of the search is greater. If the remaining list is empty, it cannot locate the key and a special "not found" sign is returned.

6. AREAS OF IR APPLICATIONS

As established, retrieving information used to be an activity involving only a few people: bibliographers, paralegals and similar skilled investigators. The world has changed, and hundreds of millions of people use the online search enginto find information each day [20]. The extraction of information is rapidly becoming the dominant method of accessing information. IR is also used in many NLP technologies for the extraction of textual data like tweets, written in different languages and other variety of jobs in health monitoring systems [21] [22] [23]. Specific knowledge recovery system implementations are as follows:

6.1. Digital Library

A library is digital when digitally compiled assortments are stored and accessible to computer. The digital content can be locally deposited otherwise accessed by computer networks remotely. A digital library is programed for accessing info [24].

6.2. Search Engines

Some of the furthermost common requests' techniques of information retrieval in large text sets are search engine. The finest known cases are online search engines, but there are also other searches, such as: desktop search, social search, federated search, mobile search, and enterprise search [25].

6.3. Media Search

A file recovery program is a computer system that uses a large digital archive to access, scan and retrieve images [26].

6.4. Information Filtering System

A filtering system is a system which removes redundant or unnecessary information through an (semi) automated or computerized method from an information stream before presentation to a person. Its key goal is to control information overload andmprove the semantic signal to noise ratio [27]. In order to do so, the user profile is compared with other comparison features. The information element (content-based approach) or

the user's social environment (collaborative approach to filtering) can give these characteristics a source.

6.5. Geographic Information Retrieval

Recovery of geographical information (GIR) improves the retrieval of geographical information. GIR seeks to answer textual questions, such as "Which war was waged in Greece?" and "restaurants in Beirut." The text indexing and interpretation are usually distinguished from geographical indexing in GIR. GIR [28] is a key component in the seminary similarity and disambiguation of the word-sense system.

6.6. Legal Information Retrieval

Recovery of legal knowledge is the science of retrieval relates to law, including law, case law, and scholarly research. It is necessary to provide laymen and legal professionals with appropriate legal knowledge to gain access to the law. Due to the huge and increasingly growing number of legal documents accessible by electronic means, has increased its value [29].

7. CONCLUSION

It is summarized that the collection of information about search and retrieval methods from record collection. This survey addresses the basics of the retrieval of information. We describe the information collection method with its simple measurements. We explored conventional information retrieval modeling along with various search techniques, indexing procedures and the IR applications

REFERENCES

[1] P. Mahalakshmi and N. S. Fatima, "An Art of Review on Conceptual based Information Retrieval," *Webology*, vol. 18, no. 1, pp. 51-61, 2021.

[2] M. Sharma and R. Patel, "A survey on information retrieval models, techniques and applications," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 11, pp. 542-545, 2013.

[3] A. Hammache and M. Boughanem, "Term position-based language model for information retrieval," *Journal of the Association for Information Science and Technology*, vol. 72, no. 5, pp. 627-642, 2021.

[4] D. W. Oard and D. e. a. Soergel, "Building an information retrieval test collection for spontaneous conversational speech," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 2004, pp. 41-48.

[5] D. Hiemstra, "*Information Retrieval Models*",Vol 10, issue 4, pp- 125-145, 2009.

[6] F. Crestani, M. Lalmas, C. J. Van Rijsbergen and I. Campbell, "Is this document relevant?… probably" a survey of probabilistic models in information retrieval," *ACM Computing Surveys (CSUR)*, vol. 30, no. 4, pp. 528-552, 1998.

[7] G. Salton, E. A. Fox and H. Wu, "Extended Boolean information retrieval," *Communications of the ACM*, vol. 19, no. 11, pp. 1022-1036, 1983.

[8] S. E. Robertson, "The probability ranking principle in IR," *Journal of documentation*, ,Vol 4 , issue 3, pp. 187- 191,1977.

[9] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin," *Communications of the ACM*, vol. 35, no. 12, pp. 29-38, 1992.

[10] N. Fuhr, "A probability ranking principle for interactive information retrieval," *Information Retrieval*, vol. 11, no. 3, pp. 251-265, 2008.

[11] S. Robertson, "Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval," in *Conference Proceedings 17th Annual International ACM SIG1R Conference on Research and Development in Information Retrieval. Google Scholar Google Scholar Digital Library Digital Library*, 1994.

[12] H. Arooj and K. I. Malik, "A Control Chart Based on Moving Average Model Functioned for Poisson Distribution," *International Journal of Current Science Research and Review*, vol. 3, no. 10, pp. 104-112, 2020.

[13] S. Acid, L. M. De Campos, J. M. Fernndez-Luna and J. F. Huete, "An information retrieval model based on simple Bayesian networks," *International Journal of Intelligent Systems*, vol. 18, no. 2, pp. 251-265, 2003.

[14] M. Neil, N. Fenton and L. Nielson, "Building large-scale Bayesian networks," *The Knowledge Engineering Review* , vol. 15, no. 3, pp. 257-284, 2003.

[15] M. e. a. Hecker, "Gene regulatory network inference: data integration in dynamic models—a review," *Biosystems*, vol. 96, no. 1, pp. 86-103, 2009.

[16] S. Raman, V. K. Chaurasiya and S. Venkatesan, "Performance comparison of various information retrieval models used in search engines," in *2012 International Conference on Communication, Information and Computing Technology (ICCICT)*, IEEE, 2012, pp. 1-4.

[17] C. C. Marshall and F. M. Shipman, "Which semantic web?," in *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, 2003, pp. 57-66.

[18] W.-c. Lee and D. L. Lee., "Signature file methods for indexing object-oriented database systems," in *Proc. ICIC, Vol. 92.*, 1992.

[19] G. Navarro, D. Moura and N. E. S., "Adding compression to block addressing inverted indexes," *Information retrieval*, vol. 3, no. 1, pp. 49-77, 2000.

[20] J. Allan, J. Aslam, N. Belkin, C. Buckley and J. Callan, "Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, University of Massachusetts Amherst," in *ACM SIGIR Forum*, New York, NY, USA, ACM , 2003, pp. 31-47.

[21] N. Tabassum, T. Alyas, M. Hamid, M. Saleem, S. Malik, Z. Ali and U. Farooq, "Semantic Analysis of Urdu English Tweets Empowered by Machine Learning," *Intelligent Automation and Soft Computing*, vol. 30, no. 1, pp. 175-186, 2021.

[22] N. Tabassum, A. Rehman, M. Hamid, M. Saleem, S. Malik and T. Alyas, "Intelligent Nutrition Diet Recommender System for Diabetic's Patients," *Intelligent Automation and Soft Computing*, vol. 30, no. 1, pp. 319-335, 2021.

[23] I. Qutab, K. I. Malik and H. Arooj, "Sentiment Analysis for Roman Urdu Text over Social Media, a Comparative Study," *International Journal of Computer Science and Network*, vol. 9, no. 5, pp. 217-224, 2020.

[24] Wan, Gary Gang, and Zao Liu, "Content-based information retrieval and digital libraries," *1foldr Import 2019-10-08 Batch 11*, vol 13, issue 3, pp. 12-18, 2008.

[25] S. Buttcher, C. L. Clarke and G. V. Cormack, Information retrieval: Implementing and evaluating search engines, Mit Press, 2016.

[26] D. Zhou, S. Lawless and V. Wade, "Improving search via personalized query expansion using social media," *Information retrieval*, vol. 15, no. 3, pp. 218-248, 2012.

[27] U. Hanani, B. Shapira and P. Shoval, "Information filtering: Overview of issues, research and systems," *User modeling and user-adapted interaction*, vol. 11, no. 3, pp. 203-259, 2001.

[28] R. Hariharan, B. a. L. C. Hore and S. Mehrotra, "Processing spatial-keyword (SK) queries in geographic information retrieval (GIR) systems," in *19th International Conference on Scientific and Statistical Database Management (SSDBM 2007*, IEEE, 2007, pp. 16-16.

[29] M. Saravanan, B. Ravindran and S. Raman, "Improving legal information retrieval using an ontological framework," *Artificial Intelligence and Law*, vol. 17, no. 2, pp. 101-124, 2009.