Information Retrieval in Business

Revati Vilas Wable

Department of Information Technology, National Institute of Technology Raipur

G.E. Road, Raipur, Chhattisgarh-492010, India

reva.wable@gmail.com

Abstract— For several years people have realized the importance of archiving and finding information. With the need of computers, finding useful information from such collections has become a necessity. Information retrieval has become an important research area in the field of computer science and gained importance in several fields like business, healthcare, agriculture, medicine, law and many other fields. This research paper focuses on the need, models and the processes involved in information retrieval. A case study on INSYDER system has been proposed to gain holistic knowledge of information retrieval in the field of business.

Keywords— Information Retrieval, IR models, Processes, IR tools, INSYDER system.

I. INTRODUCTION

Information retrieval is finding material that could be in the form of a document consisting of unstructured nature that provides the required information [1]. It collects this information from large collections stored on computers. It is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. An information retrieval system [2] is a software system that provides access to books, journals and other documents; stores and manages those documents. It is often said that information is not knowledge without information retrieval systems. The objective of an information retrieval [3] system is to minimize the time it takes for a user to locate the information they need or in other words to provide the information needed to satisfy the user's question. Satisfaction does not mean finding all the required information on a particular issue. Thus, an Information Retrieval system collects and organizes the information in more than one subject area to equip users with all the relevant information as soon as requested. It helps in identifying the potential candidates for business on the basis of the idea they want. Hence, an information retrieval system does not inform the user on the subject of his inquiry but the existence or non-existence and whereabouts of documents related to his request

Figure 1 explains the working of information retrieval process. We can classify information retrieval as free recall, cued recall, and recognition. Therefore data technology is useful in managing vital production knowledge and supports the information it helps the assembly, management, and house owners of the corporation to rise, run their business [4] and earn most profits. As per Olson (2003), the term 'Information Retrieval' was coined in 1952 and gained popularity in the research community from 1961. Information Retrieval was seen as the organizing function to obtain major advances in libraries that were no longer storehouses of books [5] but a place where information was indexed and catalogued. The concept of Information Retrieval helped some documents or records containing information that have been organized in an order suited for easy retrieval and that is why it was designed to retrieve the documents or information required by the user community. It should be such that the right information is available to the target users. Some of the tools generally used for information retrieval [6] are bibliography, index and abstract; shelf lists and library card catalogue. Evaluation in information retrieval is considered as the process of systematically determining a subject's worth, merit, and significance by using necessary criteria governed by a set of standards. The primary issues of the information retrieval systems [7] are: Query Evaluation, Document and Query Indexing, and System Evaluation.

An example of information retrieval problem would be: consider a fat book owned by many people like Shakespeare's Collected Works. Now if you want to determine which plays contain the words 'Brutus' and 'Caesar' and not 'Calpurnia' one way would be to start at the beginning and read through all the text [8]. However, the simplest form of document retrieval for a computer would be a sort of linear scan through these documents. This process is generally referred as grepping through text. It can be a very effective process when the speed of modern computers allows useful possibilities for wildcard pattern matching through the use.

In our modern technology driven society data, facts, and knowledge have a higher priority than they were a few years ago. With greater use of the internet, information has become more and more accessible. When we want to access information, it has to retrieve from online sources like the most famous Google search engine (that is why search engines are common). The answer to all the queries is "Information Retrieval" [9] that gathers the information more precisely and recovers to ensure a discipline in computer and information science. Above all the importance of search engines and accuracy is due to the complex information retrieval systems that the use. They recognize the intentions or needs behind that specific search terms and thus provide relevant data on search queries.

Fig. 1: Information Retrieval Process

Fig. 2: Model of Information Retrieval

Figure 2 explains the various concepts involved in the model of a information retrieval system. Each concept has been discussed in the latter sections of this paper. For simple querying in modern computers collections of the size of Shakespeare's Collected Works [10] is a bit under one million words of text in total hence you don't need anything more. The primary aim of these systems is to arrange all knowledge collected from each level, summarize it, and present it in a manner that facilitates and improves the standard of the choices being created to extend the company's profit and productivity. Information system is incredibly essential for running and managing a business these days.

II. SIGNIFICANCE

Worldwide organizations rely majorly on modern technology that includes information systems to develop new ways to generate revenue, engage customers and streamline time-consuming tasks. With an adequate information system [11], businesses can save time, money and take smarter decisions. This technology can be automated depending on the requirement. Here are some of the uses of information retrieval systems in business:

- 1) To represent the contents of analyzed sources in a way that matches users' queries and analyze them. Further, it can represent them in a form that will be appropriate for matching the database. This can be achieved through the design of sophisticated search interfaces.
- 2) To identify the information related to the areas of interest of the user and act as a bridge between the world of creators of information and the users that use this information. Thus, even a small business should have a well-organized information storage and retrieval system to improve their performance and be able to compete with large scale businesses. It also helps them to learn, adapt and comply these techniques to increase its growth rate.
- 3) It provides organizations with immediate value along with ways to capture tacit knowledge. It also focuses on information that already exists in electronic formats and vendors install bases which put them in a better position than small startups that helps them to grow.

- 4) To make adequate adjustments in the system based on the feedback from the users and retrieve the information that is relevant to the business. Additionally, it can be made easier to provide and monitor all the internal controls designed to cross check frauds, waste and abuse and ensure the business is complying with information privacy requirements along with the required electronic system.
- 5) These systems are designed to capture, process, store and retrieve information to hold a business together. Thus, it identifies the resources relevant to the areas of interest of the target users' community for analyzing the contents of the sources and further represent them in a manner such that it will be suitable for matching all the queries of the users.
- 6) To analyze users' queries and to represent them in a form that will be suitable for matching with the database. A proper information retrieval system includes an effective indexing system (that not only decreases the chances of information will be misfiled but also enhances the retrieval process of information). The result is time-saving benefit which increases office efficiency and productivity while decreasing other issues like stress and anxiety.
- 7) To match the search statement as per user's requirement from the stored database. This information is retrieved via a variety of tools and techniques used to determine the relevance of information and their ranking. It also follows compliance regulations and tax record-keeping guidelines significantly for businesses to increases their confidence by checking that the business is fully complying.

Thus continuous changes in all aspects of the system [12] with rapid developments in information and communication technologies related to changing patterns of society, users and their required information. It uses complex algorithms hence less human error. Furthermore, employees can focus on the core aspects of a business rather than spending hours collecting data, filling out paperwork and doing manual analysis, simplified by several data analytics tools.

III. MODELS

Fig. 3: Models of information retrieval systems

A model of information retrieval should select and rank the relevant documents as per user's query. Figure 3 represents the prominent models [13, 14, 15, 16, 17] of an information retrieval system. In this section, we have discussed these models in detail. Here, the texts of the documents and the queries are represented in the same way, so that selection and ranking of document can be formalized by a matching function. This function will return a retrieval status value (RSV) for each document in the collection.

A. Boolean Model

This model is evolved from set theory based on the principle of “exact match”. Here, we can pose any query [14] in the form of a boolean expression of terms i.e., one in which terms are combined with the operators and, or, not but with a disadvantage that it is not able to rank the returned list of documents. This model of information retrieval is known as a classical information retrieval model as it was the first and most adopted one. It is virtually used by all commercial information retrieval systems. Each document either matches or is unable to match the query. The results retrieved in the exact match are a set of documents (without ranking to match the query. As a result, a ranked list of documents is obtained. The Boolean model discussed is the most common exact match model. The retrieval function used in this model decides whether the specified document is relevant or not.

B. Inference Network Model

Document retrieval is modeled as an inference process in an inference network in this model. Most techniques used by information retrieval systems can be implemented under this model. In this model, a document searches a term with certain strength on the basis of requirement, and the credit from multiple terms is converted into a query to evaluate the equivalent of a numeric score for the document [15]. From an operational perspective, the strength of instantiation of a term for a document can be considered as the weight of the term in the document and ranked. Ranking of vector space models and the probabilistic models are done in a similar manner. The strength of instantiation of the specified term for the required document is not specified by the inference network model, and any formulation can be used.

C. Probabilistic Model

The most significant feature of the probabilistic model is its attempt to rank the identified documents by their probability of relevance of the query. Stephen Robertson [16] formulated the probability ranking principle. In this model the identified documents and queries are represented by binary vectors such as $\sim d$ and $\sim q$, each vector element representing a document attribute or term that occurs in the document or query.

For these conditional probabilities that a document d is relevant $P(R|(q,d))$ or irrelevant $P(I|(q,d))$ to the query q is calculated. It uses odds $O(R)$, where $O(R) = P(R)/1 - P(R)$ and R refers to “document is relevant” and $\sim R$ refers to “irrelevant document”. The documents with probabilities of relevance are ranked in decreasing order of their relevance and those exceeding a cut-off threshold c are the retrieved document set, defined as $R(q) = \{d | P(R(q,d)) \geq (d | P(I(q,d)), P(R|(q,d)) > c)\}$, with $P(R|(d,d))=1$ and $P(I|(d,d))=0$. Hence the ranking of the probabilities provides an effective way to obtain information.

D. Vector Space Model

This model was invented by Salton and his working group. In the Vector Space Model, documents and query are represented as a vector and the angle between the two

vectors is evaluated using the similarity cosine function [17]. It introduces a term weight scheme known as tf-idf weighting having a term frequency (tf) factor that measures the frequency of occurrence of the terms in the document or query texts and an inverse document frequency (idf) factor that measures the inverse of the number of documents that contains a query or document term.

The idea behind this is to represent text and query by weighted term vectors $D = (d_1, d_2, \dots, d_n)$ respectively. These two representations are then used to measure a degree of similarity between the query (or a sample document) and a document $\text{sim}(D, Q) = \sum diq_i$ resulting in a ranked list. For the similarity measurement different methods have been proposed like the cosine measure:

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

IV. PROCESSES

Web pages mostly contain semi-structured and dynamic information along with links that may not easily be accessible. Hence searching through the World Wide Web is significantly different from searching data in databases, which are static and centralized. A number of query languages, based on semi-structured data models and mostly represented as labeled graphs have been developed. The main problem is how to convert knowledge into information for data mining algorithms to function properly. Although most web documents are text-oriented, a considerable amount of information is not easily accessible through common search methods so documents can't be retrieved without accessing each one individually. In this section each and every process involved has described in detail with examples to provide a better understanding to the readers. Algorithms and best strategies have also been discussed to help find out adequate information precisely. The ultimate aim of information retrieval system is to find the relevant information or document that satisfies user information need and to achieve this goal; it usually implements the following processes:

A. Indexing Process

Indexing is used to explain the absoluteness of documents, text and media. Index terms may be derived from the document itself or from a document-impartial supply. Extracting the phrases from the document itself (creator absoluteness) continues the author's unique aim and expresses his expertise. The index can also use controlled or uncontrolled vocabulary and can be constructed up manually or robotically, managed vocabulary is derived from an authorized term listing, a glossary, to overcome problems like homonyms and a lack of understanding of the unique terms, using the glossary himself a consumer can locate the appropriate listed documents more without difficulty. The managed phrases are particularly used in online databases, e.g. supplied by hosts like STN [18], dialog and many others. It is where the documents required by the users are transformed into searchable data structures and so it can be referred to as the process of extraction. It creates a core functionality of the information retrieval process as it is the first step and assists in efficient retrieval of information. In the process, first, the document surrogates are created to represent each document. Secondly, it requires analysis of original documents that include simple (identifying meta-information e.g., author, title, subject and others) and complex (linguistic analysis of content) data.

Many of those systems allow the consumer to list the controlled terms. Indexes are the data structures that are used to make the search faster. However a disadvantage is the fact that the controlled phrases are frequently in the back of new developments. The guide relation of index terms lacks the consistency of indexing, it's miles a subjective indexing. Having some human beings to index the identical file leads typically to using many one of a kind index terms, relying on the background of those people. Nonetheless in a professional environment one attempts to reduce the hassle of indexer inconsistency and inter-indexer consistency through finding individuals with area knowledge, the usage of controlled vocabulary similarly to out of control and other indexing policies. Automatic indexing [19] guarantees the consistency of index phrases, as the algorithms behind are continually the identical, producing the equal effects. The idea in the back of computerized indexing is to find characteristic phrases representing a file excellent. Consequently as a minimum requirements for the illustration exist: identity of suitable content material unit (consider) and resolution of term weights to distinguish vital phrases from much less essential ones (precision) for the content material illustration. An outline of an automatic indexing is supplied by the subsequent:

1. Calculate the frequency of each term k inside each document i ($FREQ_{ik}$).
2. Sum-up the frequency of each time period k inside the entire series Σ ($FREQ_{ik}$).
3. Sort the terms in line with their lowering frequency.
4. Define a higher and lower threshold for the frequency.
5. Take away all phrases above or underneath that threshold.
6. The last phrases are then indexing terms of the record.

Nonetheless with this approach no time period weighting has been made to assign weights for the distinguishing of important and much less important phrases. An affordable measure of importance is acquired with the aid of the tf-idf equation, favoring terms with an excessive frequency especially documents (tf) but with a low frequency basic in the series (idf). As a result from a contrast of automatic indexing methods and manual keyword indexing using the abstracts of the Cranfield series, argue that automated indexing is not always as good as manual-indexing techniques. In their conclusion of the evaluation they suggest that "weighted terms need to be used, derived from file excerpts whose duration is at the least equivalent to that of an abstract".

A. Filtering Process

Filtering is a name used to describe a spread of strategies regarding the shipping of statistics to folks who need it. A more specific definition describing the information Filtering problem is given by thinking about some dynamic records objects. The concept of information filtering machine fits the characterizations [20] of the statistics gadgets towards the consumer profiles, descriptions of the customers' information desires, to obtain a relevance estimate of the facts items with respect to the facts wishes." rise inside the identity of their article the

query, whether or not information retrieval and information filtering are "sides of the equal coin", insinuating the connection of the 2 disciplines. They finish that records Retrieval and information filtering are certainly two aspects of the identical coin. The "coin" is that both disciplines address the identical goal, gratifying the information desires of humans, using similar techniques. These systems deal with the ranking of semi-structured or unstructured data (usually textual) in order of relevance. Nonetheless, contemplating selective dissemination of information (SDI), as commonplace for on-line database search as the genuine software area of information retrieval, filtering has been around inside the information retrieval context for a long term. despite the fact that environmental problems, e.g. person modeling, consumer tracking to construct profiles, social and privateers elements have in no way been the focus of information retrieval research, but are very an awful lot inside the middle of studies.

Fig. 4: Information Filtering System

Figure 4 shows the basic working or mechanism of the information filtering process/system. It refers to the selection of relevant information or rejection of the irrelevant one from a stream of incoming data. The filtering agent uses filters to filter out all the irrelevant incoming documents. Thus, it presents user only those documents which match the user's interest. During this period the filtering system [21] becomes more effective by learning the user's preferences and hence develops a great accuracy in performing the filtering tasks. It performs tasks like interfacing along with the source document subsystem thereby managing the user-profile and calculating the relevance of a document vector as per user-profiles and communicating with the user. These systems are generally applied to attain information for user's long term interests. Different criteria may be used to filter documents or articles. One such filtering based on the same concept is collaborative filtering.

- *Collaborative Filtering:*

It is a form of social filtering based on the subjective evaluations of other readers attached as annotations to the shared document. Schemes using collaborative filtering [22] use human judgments so that it does not suffer from the problems which automatic techniques have with natural language like polysemy, synonymy, and homonymy while other language constructs at a pragmatic level such as sarcasm, humor and irony may be recognized.

B. Search Strategies

Based on the user's requirement the required information is retrieved. For this various searching algorithms such as linear search, binary search, brute force search and many others are applied to the World Wide Web to get the most preferred information. It includes the use of one's knowledge about online searching systems, indexing vocabularies and conventions practiced in the generally used text database construction. A good understanding of the same and how it is implemented in the system searched makes searching easier. Some of the most suited search strategies are as follows:

1) *Brief search*: This search consists of a single query statement like Information Retrieval in Business or Information Retrieval and Business Intelligence [23]. It might act as a good starting point for further in depth querying using other search strategies to provide appropriate information. It identifies the sources of information relevant to the areas of the target user community and accordingly analyzes the contents of the sources. It also represents contents of analyzed sources that will match queries and further analyzes user queries that will match with database. Brief search is one of the most commonly used searching techniques on the World Wide Web.

2) *Block building approach*: The idea behind this is to divide the required information into several concepts, to search for these concepts separately and combine the results in a bottom-up manner. Each of these concepts gives a result within the database, combining these three results using the set operation i.e. AND that retrieves the desired documents. By using related words or acronyms the set operation i.e. OR might lead to an expansion of the final result set. Figure 5 explains the working of block building [24] approach:

Fig. 5: Example of block building approach

The approach is advantageous when complex information is to be retrieved such that it consists of more than three entities/concepts such that the user can keep a track of how well each entity is represented in the given database. The only problem is that the user has to be aware of the Boolean set operations used otherwise this search will fail to execute. Only some search engines provide the number of how many times a keyword in the query has been found.

3) *Successive Fractions Approach*: This approach starts with a broad set of documents, successively filtering it by using the set operation AND to narrow the result set till a feasible size is reached to retrieve the desired information.

In comparison to the block building approach discussed above this search strategy is impractical for searching on the World Wide Web [25] as most of the time no status information about the search is given. Consider a situation where an online database search system keeps track of the different search sets and combines. It narrows the search by applying limiting techniques like Boolean operators AND, OR or NOT. This process is carried out step-by-step until the search is reduced to a manageable number of hits.

4) *Facets strategies*: This strategy can be described as variations of the block building approach. The most suited concept first strategy recites that the user selects as the first facet which is believed to be the most specific in the information needed. This type of search allows navigation along with several independent dimensions. It is significant to communicate the user's current location and navigation options and this can be done in three ways (to communicate navigational state): Breadboxes, Multi-Selectable Facets and Inline Breadcrumbs concept [26]. The lowest postings first strategy dictates he chooses the concept he believes to be the rarest in the database as the first concept.

Fig. 6: Most specific facet strategy

Figure 6 represents the most specific facet strategy where the most specific and next specific concepts are analyzed and added to the result set. Apart from the search strategies, searching the World Wide Web is different from searching online databases. Describing the documents contained using a formal and content based approach. Other than internet search engines, the user of an online database gains an overview of all his search sets and can combine these easily. As the HTTP protocol is stateless this is difficult to achieve for search-engines.

5) *Citation Pearl Growing*: This searching technique is based on the idea to find a relevant document and similar ones by using vocabulary from this document, descriptors or classification codes. The user stops the search when he has enough documents to satisfy his information need. This technique is in accordance with the relevance feedback option that some retrieval systems offer and is also useful for searching the internet [27], as the user can select appropriate keywords from relevant documents. The only problem with this technique is that the system support from the search-engines is mostly not given / specified.

V. INFORMATION RETRIEVAL TOOLS

Several information retrieval tools are available on the Internet. User can choose the tool of their choice to retrieve the desired material. Because of different people accessing information for different reasons user needs to use the right tools to locate their material. Its primary goal is to supply the

right information to the right user at a right time. Different techniques [28], materials and methods are used for retrieving the desired information. It provides organizations or businesses with immediate value i.e. important information / data to try to figure out ways to capture tacit knowledge. Some examples of information retrieval tools would be classification schemes, catalogues, indexes and other information retrieval tools in the library include the following: almanacs, handbooks, periodicals, atlases, encyclopedia, directories, dictionaries, and concordances among others. Internet search engine, subject directory, online database, online public access catalogues (OPAC) and digital library. Information Retrieval thus with the help of these tools provides a means to get at information that already exists in electronic formats. These tools differ in structure as their function and use different methods and techniques for storing and retrieving the information. This reflects the target audience and indeed their intended use. Some of these tools are:

- 1) OPAC i.e. online public access catalogues are generally used by students to find books from online library instead borrowing it from the library. It is a computerized catalogue [29] containing bibliographic records of items in a library. However, in digital age students rely heavily on the Internet and they usually use internet search engine to find sources of information. The information found in search engine may be a web page, images and any other type of file. It uses web crawler to retrieve the information from millions of web pages on the web which is further stored as the search engine index thereby making the search engine the most comprehensive coverage of the web.
- 2) Even subject directories like Yahoo and DMOZ are used in order to locate the required information. These directories [30] are created by assigning the submitted sites manually to a suitable subject category by the directory developers itself. Despite this some students assume subject directories as Internet search engine whereas the real internet search engines are AOL Search, AltaVista, and Google.
- 3) Online databases also provide access to the remote databases through the so called database vendor or service provider. Examples of such databases [31] are IEEE, Elsevier, and ACM. It is also defined as digital library as it is an organized collection of information with associated services wherein the information is stored in digital formats and is accessible over a network. This ensures a high quality resource.

VI. CASE STUDY

Here, we would review the real time application of information retrieval by considering the INSYDER system that helps in seeking business information from the World Wide Web. The advantage of using external information for business intelligence system helps enterprises to know more about its customers, suppliers, competitors, government agencies, and other external factors. Valuable information about external business factors is readily available on the internet but only a few are used as reliable data sources. Figure 7 shows the architecture [32] of the INSYDER

system having components mainly developed in Java. Only the component involved in semantic analysis has been developed in C++.

Fig. 7: Architecture of the INSYDER System

Further, the user-interface and visualizations have also been developed in Java using Java Foundation Classes or Swing. The scheduler's is responsible for the monitoring process of user's query on the internet. The watch function used for the same checks user-defined Web pages [33] for regular changes. These sources are defined in various XML documents, thereby enabling an easy maintenance and extension of the sources in an organized manner. Web-API supports easy access to the documents thereby acting as a set of functions and methods. For every document that matches user's requirement, calls the semantic analysis via a COM wrapper to get a relevance value. It further uses a semantic net that models the real world by a controlled vocabulary and can be individually adapted to various application domains.

INSYDER uses a dynamic search approach [34] for the online search to discover relevant information by following links. The sources are representation of starting points of a search. It determines the relevance ranking by using semantic analysis of documents. A significant aspect is the fact that ideas and components from different fields are combined. Nowadays, systems do a dynamic search with a metadata generation and the visualizations of the same leads to document inherent data i.e. new. The visualization of the query is performed in the following manner:

- 1) First step is the semantic analysis that includes various relationships such as narrower term, part-of broader term although they are not represented in the graph visualization but tells that there is a relationship.
- 2) To keep the overview (for user's assistance) the system was designed with a detailed and full view. This is done by taking the information from the tree view. For example, if the user clicks on a branch of the tree view then only that branch should be visualized in the graph. On clicking the root of the tree the result should be a graphical presentation of the entire tree in the graph.
- 3) Lastly, interaction with the graph representation includes all the terms represented in the graphical representation. It can be moved by keeping the

relation to their base node in a way that the elements are ordered automatically using an algorithm to ensure that elements are connected to a node such that most of them are viewable.

It uses two ranking algorithms: Natural Language which is the default ranking algorithm of the system and Concept Query [35]. Using the Visual Query the user can select terms to go on with semantic analysis. It offers an Interactive Relevance Feedback function to use the judgments made by the user about documents to derive new query terms from them. Firstly, we extract the feature concepts of a document, and then describe these concepts to its best. For example in a search on "information visualization" the user decides whether to have documents more like the first two suggested or nothing similar to the them. To calculate the ranking values the query [36] itself is put in a meta description for the comparison of the query and the document and the matching is done segment by segment. These values are then used for the visualizations, and the final ranking value is calculated by taking into account the mean value of all segments and the maximal value reached by one or more segments. To calculate the overall relevance of a document Boolean 'AND' with a 'NEAR' proximity operator is used and the 'OR' operation to compare systems using Boolean logic. The effectiveness of the system shall be measured by using the term frequency inverse document frequency (tf-idf) measurement as a baseline for the ranking algorithms.

VII. CONCLUSION

This research paper specifies the need of information retrieval in business. Further, it discusses the various models involved like the Boolean, Inference Network, Probabilistic and Vector Space Model in detail. The different processes involved (indexing, filtering and searching; brief search, block building approach, successive fractions approach facets strategies and citation pearl growing) in information retrieval along with the various tools used have also been discussed. Lastly, a case study based on INSYDER system has been proposed along with its architecture to gain a holistic understanding of how information retrieval is used and how it is helpful in business. It includes the visualizations of the search results with respect to Visual Query and the Concept Query Ranking Algorithm. With more advance techniques and proper research we can solve the problem to get a direct connection to the semantic net in dose co-operation. Moreover, the evaluations of the relevance feedback, ranking, and the usability of the Visual Query have to be made soon.

ACKNOWLEDGMENT

The author wishes to thank all the reviewers, faculty members who helped throughout and colleagues for their valuable suggestions and comments that help improve the contents of paper. This paper and the research behind it would not have been possible without their exceptional support and guidance.

REFERENCES

- [1] Soumya Shukla. 2019. "Information Retrieval- A Brief Overview." <https://medium.com/@soumya.vkshukla/information-retrieval-a-brief-overview-173bba8fe0e9#:~:text=Information%20retrieval%20is%20the%20activity,information%20need%20of%20the%20user.>
- [2] "Information Retrieval." 2014. ScienceDirect. <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/information-retrieval>.
- [3] Ceri S., Bozzon A., Brambilla M., Della Valle E., Fraternali P., Quarteroni S. (2013) The Information Retrieval Process. In: Web Information Retrieval. Data-Centric Systems and Applications. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39314-3_2
- [4] M. Sanderson and W. B. Croft, "The History of Information Retrieval Research," in Proceedings of the IEEE, vol. 100, no. Special Centennial Issue, pp. 1444-1451, 13 May 2012, doi: 10.1109/JPROC.2012.2189916.
- [5] N. Nikitinsky, D. Ustalov and S. Shashev, "An information retrieval system for technology analysis and forecasting," 2015 Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT), St. Petersburg, Russia, 2015, pp. 52-59, doi: 10.1109/AINL-ISMW-FRUCT.2015.7382969.
- [6] Stefano Ceri, Alessandro Bozzon, Marco Brambilla, and Emanuele Della Valle. 2013. "The Information Retrieval Process." Research Gate. https://www.researchgate.net/publication/312733684_The_Information_Retrieval_Process.
- [7] D. C. Ferreira Garcia, C. C. Gattaz and P. E. Cruvinel, "Information Retrieval: A Case Study on Contributions of Greimasian Semiotics to Semantic Computing in Agriculture for Knowledge Management," 2019 IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA, 2019, pp. 478-484, doi: 10.1109/ICOSC.2019.8665607.
- [8] Xiangzhu Gao, S. Murugesan and B. Lo, "A dynamic information retrieval system for the Web," Proceedings 27th Annual International Computer Software and Applications Conference. COMPAC 2003, Dallas, TX, USA, 2003, pp. 663-668, doi: 10.1109/CMPSAC.2003.1245413.
- [9] Amit Singhal. 2018. "Modern Information Retrieval:." IEEE. <http://160592857366.free.fr/joe/ebooks/ShareData/Modern%20Information%20Retrieval%20-%20A%20Brief%20Overview.pdf>.
- [10] G Mußler. 2002. *Information Retrieval techniques for retrieving Business Information from the WWW*.
- [11] Jon Espen Ingvaldsen. n.d. "Information Retrieval for Organizational Business Process Insight." <http://ceur-ws.org/Vol-263/paper7.pdf>.
- [12] Nodtvedt, Einar. 1980. *Information Retrieval in the Business Environment*.
- [13] Amati G. (2018) Information Retrieval Models. In: Liu L., Özsu M.T. (eds) Encyclopedia of Database Systems. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-8265-9_916
- [14] A. H. Lashkari, F. Mahdavi and V. Ghomi, "A Boolean Model in Information Retrieval for Search Engines,"

- 2009 International Conference on Information Management and Engineering, Kuala Lumpur, Malaysia, 2009, pp. 385-389, doi: 10.1109/ICIME.2009.101.
- [15] Akram Roshdi, and Akram Roohparvar. 2015. "International Journal of Computer Networks and Communications Security." *Review: Information Retrieval Techniques and Applications*, (September).
- [16] University of Twente. n.d. "Information Retrieval Models."
- [17] Balwinder Saini, Vikram Singh, and Satish Kumar. 2014. "Information Retrieval Models and Searching Methodologies: Survey." *International Journal of Advance Foundation and Research in Science & Engineering (IJAFRSE)*, (July).
- [18] Kiran Prakash Bachchhav. 2016. "Information Retrieval: search process, techniques and strategies." (February).
- [19] H. Kaur and V. Gupta, "Indexing process insight and evaluation," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7830087.
- [20] N. DeClariss, D. Harman, C. Faloutsos, S. Dumais and D. Oard, "Information filtering and retrieval: overview, issues and directions," Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Baltimore, MD, USA, 1994, pp. A42-A49 vol.1, doi: 10.1109/IEMBS.1994.412140.
- [21] T. Matsuda, K. Nakamura and N. Sakamoto, "An efficient Internet crawling and filtering system for the nationwide tendering information retrieval," Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), Halifax, NS, Canada, 2003, pp. 722-725, doi: 10.1109/WI.2003.1241304.
- [22] Cakmakov, D., Davcev, D. Information Retrieval and Filtering of Multimedia Mineral Data. *Multimed Tools Appl* 1, 367-382 (1995). <https://doi.org/10.1007/BF01215884>
- [23] J. Pokorny, "Web searching and information retrieval," in *Computing in Science & Engineering*, vol. 6, no. 4, pp. 43-48, July-Aug. 2004, doi: 10.1109/MCSE.2004.24.
- [24] Tetsuya Sakai, Douglas W. Oard, and Noriko Kando. 2021. *Evaluating Information Retrieval and Access Tasks*. N.p.: NTCIR's Legacy of Research Impact.
- [25] Balwinder Saini, Satish Kumar, and Vikram Singh. 2014. "Information Retrieval Models and Searching Methodologies: Survey." https://www.researchgate.net/publication/274837522_Information_Retrieval_Models_and_Searching_Methodologies_Survey.
- [26] S. Mishra, S. K. Satapathy and D. Mishra, "Improved search technique using wildcards or truncation," 2009 International Conference on Intelligent Agent & Multi-Agent Systems, Chennai, India, 2009, pp. 1-4, doi: 10.1109/IAMA.2009.5228080.
- [27] Y. Guo, "Methods and Techniques of Searching Terms Determination in Computer Information Retrieval," 2012 International Conference on Control Engineering and Communication Technology, Shenyang, China, 2012, pp. 948-951, doi: 10.1109/ICCECT.2012.79.
- [28] M. Ginsburg, "Annotate! a tool for collaborative information retrieval," Proceedings Seventh IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE '98) (Cat. No.98TB100253), Stanford, CA, USA, 1998, pp. 75-80, doi: 10.1109/ENABL.1998.725675.
- [29] N. L. Mohd Shuib, N. Abdullah and M. H. bin Ismail, "The use of information retrieval tools: A study of computer science postgraduate students," 2010 International Conference on Science and Social Research (CSSR 2010), Kuala Lumpur, Malaysia, 2010, pp. 379-384, doi: 10.1109/CSSR.2010.5773804.
- [30] J. Rabinow, "Presently available tools for information retrieval," in *Electrical Engineering*, vol. 77, no. 6, pp. 494-498, June 1958, doi: 10.1109/EE.1958.6445131.
- [31] J. T. Yao and Y. Y. Yao, "Web-based information retrieval support systems: building research tools for scientists in the new information age," Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), Halifax, NS, Canada, 2003, pp. 570-573, doi: 10.1109/WI.2003.1241270.
- [32] "INSYDER system." 2000. https://www.researchgate.net/figure/Architecture-of-the-INSYDER-system_fig1_30012845.
- [33] Back J, Oppenheim C (2001) A model of cognitive load for IR: implications for user relevance feedback interaction. *Inf Res* 6(2). <http://InformationR.net/ir/6-2/ws2.html>
- [34] Klein P, Müller F, Reiterer H, Limbach T (2003) Metadata visualization with VisMeB. In: Proceedings of the 7th international conference on information visualisation (IV 03). IEEE Press, New York, pp 600-605
- [35] Gabriela Mußler, Harald Reiterer, and Thomas M. Mann. 2000. "INSYDER : information retrieval aspects of a business intelligence system." Research Gate. https://www.researchgate.net/publication/30012845_INSYDER_information_retrieval_aspects_of_a_business_intelligence_system.
- [36] Reiterer, H., Tullius, G. & Mann, T. INSYDER: a content-based visual-information-seeking system for the Web. *Int J Digit Libr* 5, 25-41 (2005). <https://doi.org/10.1007/s00799-004-0111-y>