IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 21-24

A Study of Information Retrieval Systems

Mrs. Rashmi G.Dukhi
G.H.Raisoni Institute of Information Technology, Nagpur, India

Abstract-*Information has become the most significant source of our day-to-day life. Information available on internet may create some confusion among its users because of its diversity. In order to get proper and exact information from internet, users need to know the effective techniques and strategies. This paper focuses on the functional process of retrieval which help users to get the required information and also to save their valuable time. This paper also contains the examples in which information retrieval techniques are used.*
Keywords: *Information retrieval; Search process; Search strategies; Retrieval techniques*

I. Introduction

Information retrieval (IR) is the activity of obtaining information system resources relevant to an information need from a collection. Searches can be based onfull-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for metadata that describe data, and for databases of texts, images or sounds.

Automated information retrieval systems are used to reduce what has been called information overload. An IR system is a software that provide access to books, journals and other documents, stores them and manages the document. Web search engines are the most visible IR applications. An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy.

An object is an entity that is represented by information in a content collection or database. User queries are matched against the database information. However, as opposed to classical SQL queries of a database, in information retrieval the results returned may or may not match the query, so results are typically ranked. This ranking of results is a key difference of information retrieval searching compared to database searching[1].

Depending on the application the data objects may be, example, text documents, images[2], audio,[3]mind maps [4] or videos. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by document surrogates or metadata.

Most IR systems compute a numeric score on how well each object in the database matches the query, and rank the objects according to this value. The top ranking objects are then shown to the user. The process may then be iterated if the user wishes to refine the query[5].

H.P.Luhn first applied computers in storage and retrieval of information. Different types of information retrieval systems have been developed since 1950's to meet in different kinds of information needs of different users. Information retrieval system offers different search approaches those deals with three basic aspects.

These aspects are as follows.

- Information storage and organization.
- Information representation.
- Information access.

II. Basic Model Of IR Systems

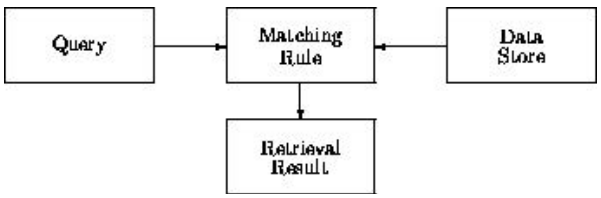

Figure 1. General Model of Information Retrieval Systems

III. Functional Processes Of IR

Four Major Functional Processes are:-

A. Item Normalization
B. Selective Dissemination of Information
C. Document Database Search
D. An Index Database Search along with the Automatic File Build process that supports index files.

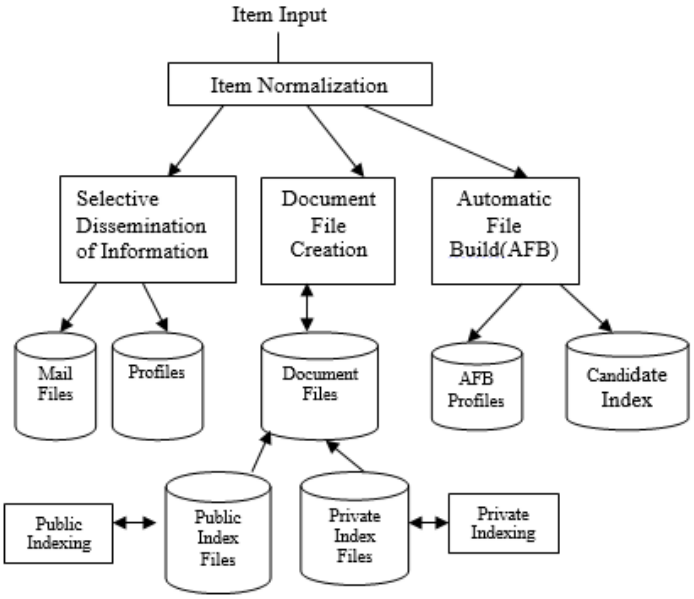

Figure 2. Functional Processes of Information Retrieval Systems

A] Item Normalization

In any integrated system is to normalize the incoming items to a standard format. It provides logical restructing of the item. Operations during item normalization is identification of processing tokens, characterize tokens, stemming of token.

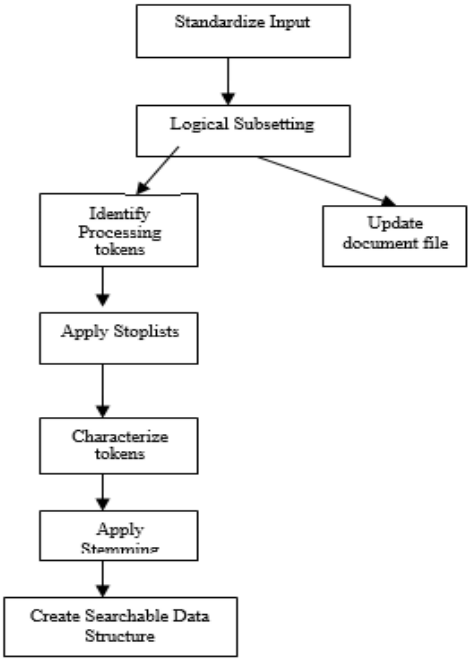

Figure 3. Steps of Item Normalization

B] Selective Dissemination of Information

It provides the capability to dynamically compare newly received items in the information system against standing statements of interest of users & deliver the item to those users whose statement of interest matches the contents of the item.

C] Document Database (DDB) Search

DDB Search process provides the capability for a query to search against all items received by the system. It is composed of search process, user entered queries & the document db which contains all items that have been received, processed & stored by the system.DDB can be very large, hundreds of millions of items or more. Items in DDB do not change once received. The value of much information quickly decreases over time.

D] Index Database Search

In this process, the user can logically store an item in a file along with additional index terms & descriptive text the user wants to associate with the item. It is also possible to have index records that do not reference an item, but contain all the substantive information in the index itself. There are 2 classes of index files public & private. Every user can have one or more private index files leading to a very number of files.

IV. Examples Of IR Systems

A] Boolean Library Catalog

The illustrative model in Figure 2 was, itself, based on the typical form of current first and second generation online library catalogs. The Representations (catalog records) are derived, via Representation Making Rules, from the Source Objects (books, periodicals, microforms, etc. in the collection). Some of this is the direct copying of fragments (e.g. titles, call numbers); some is a more complex intellectual representation derived from External Knowledge Sources (e.g. assignment of subject headings and classification numbers). In practice the representations are largely indirect copies, being derived directly from external sources, especially in the form of other cataloger's previously created catalog records (copy cataloging) rather that directly from the source object. The Searchable Index is limited in practice to a small number of the fields of the catalog records (Representations).

User Queries are accepted into the system from users either as well-formed and normalized Formal Queries, in the case where the searcher is experienced and uses the ``command line interface", or in some less well-formed format that must be go through a Query Development process before searching.

The Query Development process commonly has an option (or requirement) that it be a two stage process. Two of the primary examples of such transformations in library catalogs are index browsing and transformations which rely on External Knowledge Sources. In the former, the User Query is used to select a subset of the terms from the Searchable Index which the user *scans*, and from which, the user then selects one or more terms which are used to retrieve the Representations associated with these terms. Such selection by ``browsing" is, in effect, a two-stage retrieval process.

Queries may also be transformed with various sorts of External Knowledge, such as thesauri, dictionaries, controlled vocabulary lists, subject headings, etc. In online catalogs, these sources nearly always have some sort of syndetic structure (``broader term", ``narrower term", ``use for", ``see also", etc.) which can be used, either algorithmically or by hand, to harmonize the User Query with the system vocabulary (Searchable Index) for better results. For example, a syndetic structure may be in place so that one form of a query term will be represented as another, e.g. a search for Mark Twain will retrieve Samuel Clemens and vice-versa.

Retrieved sets are normally re-ordered by main entry before being output as a display or as a stream of records. This re-ordering of retrieved sets is, in effect, an automatic, ``hardwired" partitioning instruction. Future online catalogs will probably allow the user to choose what re-ordering or aggregating (by date, by availability, etc.) to invoke.

B] Full-Text Retrieval

In a simple case of retrieval from full-text, electronic texts would be stored (copied into the system) to become the Representation, and all of the texts would be searchable for the occurrence of specified phrases, words or word fragments. In such a case, if all of the text can be searched, the Searchable Index is actually (or logically) co-extensive with the Representation. The Retrieved Set could consist of either partial or complete copies of the those texts which satisfy the Matching Rule. The syndetic structure component of the Searchable Index is absent.

One degree more complex would be to represent the relative location of pairs of terms or to impose some vocabulary control in the form of stop words so that the significance of a term could be represented more

reliably. More sophisticated still would be information storage systems which use or include algorithmically generated representations (e.g. weighted vectors of terms) of the texts.

C] Message Filtering

Systems for filtering electronic messages (or other objects) constitute an example in which objects are represented, filtered (searched) and then discarded or relegated to other storage. In this case the User Query, once developed, remains indefinitely in place as a stored instruction (Matching Rule) which is used to select messages (Representations) as soon as they have been copied into the system. A stored, ``standing" query resembles the default alphabetizing of retrieved sets in online catalogs: Both are, in effect, latent matching instructions, instrumental in partitioning whatever data may come their way. In this sense, filters with stored queries and transient data objects are symmetrical with retrieval systems with trier transient queries and stored data objects.

Primitive retrieval systems based of a serial scan of searchable records, such as the mid-twentieth century "rapid selector" machines for scanning long spools of microfilm, can be seen as an intermediate design between typical modern filters and typical modern retrieval systems.

V. Conclusions

The information is need of current age human, therefore a number of information extraction and retrieval applications are developed recently that supports the current age information needs. During the implementation of different kinds of data search and retrieval techniques a number of methods for structured and similarly unstructured information processing is developed recently. Among most of the work is devoted to the structured data processing, but their limited efforts are made for retrieving information from the unstructured data sources. The traditional unstructured data processing techniques either not much efficient, or not accurate for adopting and using in real world application therefore a new technique is required to investigate and develop by which the user query relevancy and performance are both improved.

References

[1]. Jansen, B. J. and Rieh, S. (2010) The Seventeen Theoretical Constructs of Information Searching and Information Retrieval. Journal of the American Society for Information Sciences and Technology. 61(8), 1517-1534.
[2]. Goodrum, Abby A. (2000). "Image Information Retrieval: An Overview of Current Research". Informing Science. 3 (2).
[3]. Foote, Jonathan (1999). "An overview of audio information retrieval". Multimedia Systems. 7: 2–10. CiteSeerX 10.1.1.39.6339. doi:10.1007/s005300050106.
[4]. Beel, Jöran; Gipp, Bela; Stiller, Jan-Olaf (2009). Information Retrieval On Mind Maps - What Could It Be Good For?. Proceedings of the 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'09). Washington, DC: IEEE.
[5]. Frakes, William B.; Baeza-Yates, Ricardo (1992). Information Retrieval Data Structures & Algorithms. Prentice-Hall, Inc. ISBN 978-0-13-463837-9. Archived from the original on 2013-09-28.
[6]. C.S. Naga Manjula Rani, "Importance of Information Retrieval", Oriental Journal of Computer Science& Technology, vol. 4, no. 2, pp. 459-462, 2011.
[7]. M. Kc, M. Hagenbuchner, A. C. Tsoi, "Quality Information Retrieval for the World Wide Web", International Conference on Web Intelligence and Intelligent Agent Technology, IEEE, pp. 655-661, 2008.
[8]. H. C. Yang, C. H. Lee, "Mining Unstructured Web Pages to Enhance Web Information Retrieval", International Conference on Innovative Computing, Information and Control, IEEE, Volume 1, 2006.
[9]. J. D. Rose, J. Komala, M. Krithiga, "Efficient Webpage Retrieval Using WEGA", Procedia Computer Science, Volume 87, pp.281 – 287, 2016.
[10]. L. P. Jing, H. K. Huan, H. B. Shi, "Improved Feature Selection Approach TFIDF in Text Mining", Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, IEEE, pp.944-946,2002.