Crux and Crushed of the Information Retrieval

Sandeep Kaushik[1]
A.N.Jha[2]
Assistant Professor, Department of I.T
L.N. Mishra College of Business Management, Muzaffarpur[1]

Ex-Faculty, Department of Maths (L.N.T.College Muzaffarpur)

B.R.A.Bihar University, Muzaffarpur[2]

Abstract:

In modern days the world has become highly dependent on world wide web. Now it has captured every walk of human life. Creation of global web succeeded in allowing the people to share their information and data with each other globally. Such uses and presence of data and information created trillions of databases. Now in this complex scenario searching a particular information or data with accuracy is the need of hour. We need specialist tools. Search engines are one of the answers but retrieving meaningful information is difficult. To overcome this problem with search engines many modern technologies have been implemented which may retrieve meaningful information intelligently. Semantic web technologies are playing a major role. In this paper we are presenting the modern techniques on the search engine in intelligent web search technologies.

Keywords:

Semantic Web, Information Retrieval

Introduction:

An information retrieval seems to be very easy task with a general psyche of the common user of the local database and they keep this psychology with them even they venture into the globally distributed databases. But in such huge depository searching the correct and accurate data or information may be a case of mirage of a dessert of sahara. When a user searches an information or data they may have a very special understanding of the idea about the particular search which may not be understood by the search engines. Traditional search engines search for keywords to present the result for the queries from users. The search engines usually search web pages by matching the key words in the available web pages for the required information. Some search engines filter the pages from searching unnecessary pages by using advanced approaches. Such search engines presents topic wise results with advance filtering. The current globally present information data on web is highly unexpected in its size, verity and standards. The reason is that it came into existence with almost-complete lack of coordination in its creation. This happened due to the diversity of backgrounds and motives of its participants. Thus it made highly difficult to search content on the huge depository.

Problem definition

Conventional web search engines are the most widely used systems for searching and retrieving of information. But it is highly deplorable state that despite the overload of information, we regularly fail to locate relevant information. They are based on keyword searching. They search key words in web pages to answer the query of users. In this conventional information searching the advantage is that it is

- Simple
- Quick
- Easy

But beyond these advantages it has several disadvantages and they are

- Lack of knowledge expression and
- Lack of query understanding and processing, which leads to unsatisfactory search quality.

A user may require some information with his own understanding of web services associated with retrieved content. But the generic search engines do not provide the web services associated with the users understanding and the request automatically.

Some of the examples can be seen here as the proof of the misunderstanding between the user and the web services.

Example 1.

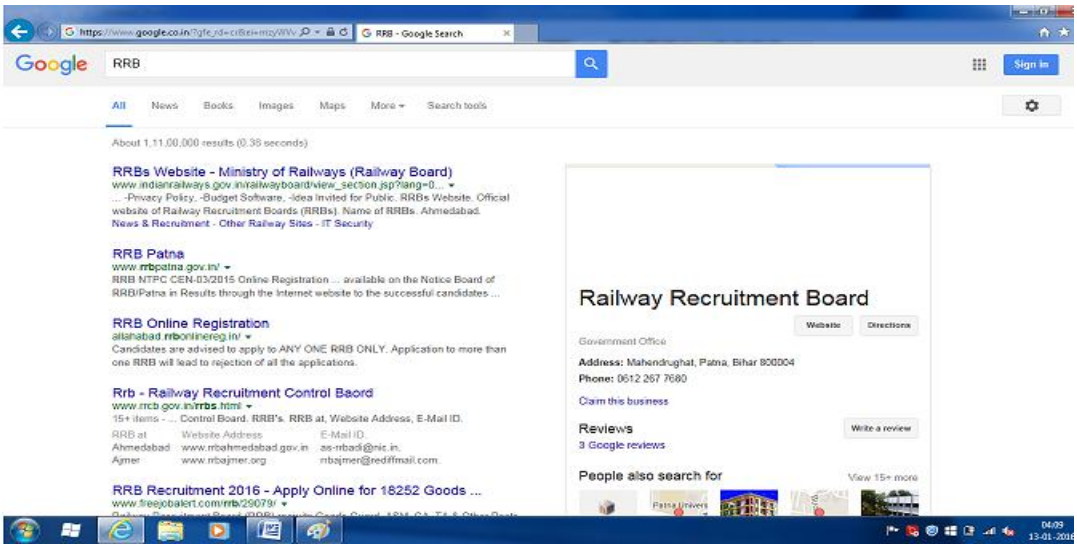

Fig 1 :- User queried for "RRB" with the intention to acquire some information regarding **Regional Rural Bank** but it presented the information regarding **Railway Recruitment Board** due to lack of understanding of users idea about the query. It gives result completely irrelevant to the user's query.

Example 2.

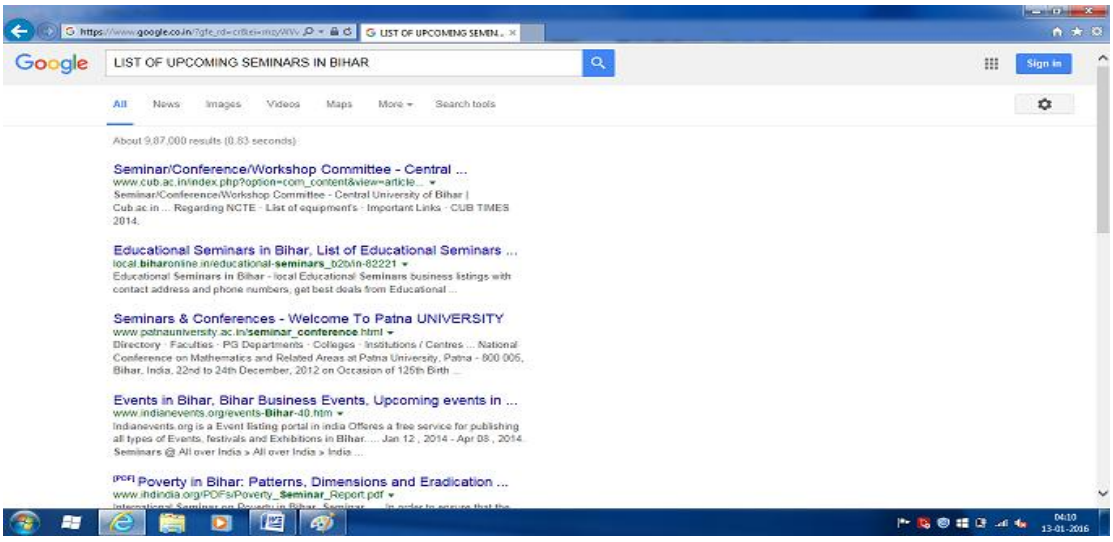

Fig 2:- If the query is a location dependent query the result relevant to the location may not be retrieved appropriately. This can be easily seen in this fig, where user wants to see the list of upcoming seminars in Bihar, whereas the results retrieved are not accurate.

Traditional conventions

Ever since the invention of hypertext, documented by Vannevar Bush in the 1940's significant progress has made the formation of the world wide web, in the 1990's. Web usage has shown tremendous growth to the point where it now claims a good fraction of humanity as participants in 2015, by relying on a simple, open client-server design. The server communicates with the client via a protocol and fulfills their purpose. Client asks for services and server responds as per their understanding. Each of these seemingly innocuous features has contributed enormously to the growth of the web.

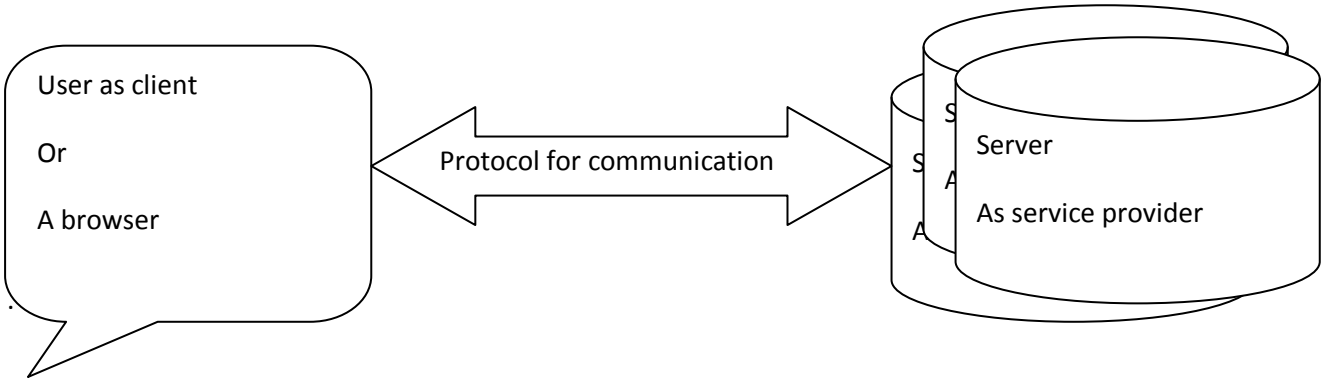

Fig 3: User asks for some information on web. The browser sends a request to the web application, and fetches the information from back-end databases then selects a web page and returns it to the browser.

The first generation of web search engines used some classical search techniques such as matching the texts. Some earliest web search engines had to contend with indexes containing tens of millions of documents, which was a few orders of magnitude larger than any prior information retrieval system in the public domain. Indexing, query serving and ranking required the engaging of

tens of machines to create highly reliable systems. The first generation of web search engines was largely successful at solving these challenges while continually indexing a significant fraction of the web. However, the quality and relevance of web search results is not considered for much of importance. As conclusion drawbacks of the current web can be listed as

- The web content lacks a proper structure regarding the representation of information.
- Ambiguity of information resulting from poor interconnection of information.
- Usability to deal with enormous number of users and content ensuring trust at all levels.
- Incapability of machine to understand the provided information due to lack of universal format.

Intelligent Search Engines:

The scale and scenario of present era in web contents and their use has triggered the researches in the field of intelligent search engines which may answer meaning based. Meaning based search is an interesting research area. The reason is that meaningful search is a key aspect for human conversations. In fact, the fast development of the meaning based searches has led researchers to focus on the development of techniques based on synonym recognition to improve the discovery of resources on the web. The internet data resources are increasing day by day. The traditional search engines such as Yahoo, Google use traditional search techniques, which do not satisfy user requirements to catch high grade information from web. Hence they need to search related reliable and latest information more precisely and efficiently. A couple of Intelligent search engines are designed and implemented for different working environments, and the mechanisms that realize these search engine are distinct.

To overcome the limitations of the current search engine, we need to use modern techniques in information retrieval. One of the best modern techniques is **semantic web and ontology.**

Semantic web is an extension of current web that allows the meaning of information to be precisely described in terms of well defined vocabularies that are understood by machines. This term was coined by Tim Berners-Lee.Due to its semantic searching which overcome the bottleneck of keyword searching, it is often called as next generation's web and web of linked data that can be processed by machines. It is one layer above what any search engine retrieves just by analyzing the keywords. In semantic web query is analyzed syntactically and semantically both. Thus the result retrieved is more accurate and relevant to user.

The semantic web vision is based on structuring the knowledge that is present in current web so that they can be understood and processed by machines autonomously. According to W3C, Semantic Web research goal is to develop a series of computer-readable expressions and processing of semantic information in language and technology to support extensive and effective automatic reasoning in network environment. To achieve this goal, Semantic web requires a multi-technology framework known as Semantic Web Stack.

Semantic web stack is a set of network information storage, organization, and all aspects of security authentication complete system, involving **XML** (Extensive Markup Language), **RDF**(Resource Description Framework), **Ontology**, digital signature techniques and methods, contributing to the network information retrieval system based on semantic level. The key ingredients of Semantic Web Stack are

- ***XML***: - XML is "corner stone" of Semantic Web. As WWW is based on HTML, Semantic Web is based on XML as its framework language. It creates application independent document and

data. It has a standard syntax for metadata and standard structure for both document and data.

- ***RDF: -*** RDF is world wide web consortium (W3C) standards designed as a metadata data model. A description is a set of statements about the resource. The RDF model is often called a "triple" because it has three parts: subject, predicate, object.
  - Subject: - This is the resource that is being described by the ensuing predicate and object.
  - Predicate: - This is a function from individuals to truth-values with an parity based on the number of arguments it has.
  - Object: -This is either a resource referred to by the predicate or a literal value.
  - Statement: -This is the combination of the three elements, subject, predicate, and object.

Ontology is a term of philosophy which can help greatly in intelligent search with semantic web design. The ontology philosophy emphasises on theory of existence. It is an explicit specification of conceptualization. For semantic web searches it can describe some domain with its common sense. It can be intensely used with semantic web searches techniques for adding some extra meaning to the searches for a common sense. In our problem definition we have described how the traditional search engines have presented the wrong results with their traditional search methods. They searched railway recruitment boards instead of regional rural banks. In their search keyword rail came immediately and before the word regional. The common sense was absent from the point of view that it was searched by some special intension and user. Ontology is considered as a backbone of the semantic web searches.

The ontology is the foundation of vocabularies and effective communication on the Semantic Web. It allows different systems or applications to cooperate that were not formerly designed to interoperate. It not only supports interoperability and common understanding between the different parties, but is a key component in solving the problem of semantic heterogeneity.

Ontology is comprised of four main components: concepts, instances, relations and axioms. The present research adopts the following definitions of these ontological components:

Concept is an abstract group, set or collection of objects. It is the fundamental element of the domain and usually represents a group or class whose members share common properties. This component is represented in hierarchical graphs, such that it looks similar to object oriented systems. The concept is represented by a super-class, representing the higher class or so called parent class, and a subclass which represents the subordinate or so-called child class. For instance, a university could be represented as a class with many subclasses, such as faculties, libraries and employees.

Instance is the ground-level component of an ontology which represents a specific object or element of a concept or class. For example, lnmcollege could be an instance of the class BRA Bihar University.

Relation is used to express relationships between two concepts in a given domain. More specifically, it describes the relationship between the first concept, represented in the domain, and the second, represented in the range. For example, study could be represented as a relationship between the concept person and university or college.

Axiom is used to impose constraints on the values of classes or instances, so axioms are generally expressed using logic-based languages such as first-order logic; they are used to verify the consistency of the ontology.

Limitations:-

An extra care is required for quality and should be taken into account when developing languages for encoding ontologisms and semantic web. Syntax is one of the most important features of any language, so it should be well-defined. The semantics of knowledge should be well defined, because it represents the meaning of that knowledge. Formal semantics should be established in the domain of mathematical logic in a clearly defined way that will lead to unambiguous meaning, since well defined semantics will lead to correct reasoning. Semantics can be considered a prerequisite to support reasoning. On the other hand, reasoning will help to check and discover consistent ontology, to verify unintended relationships between classes and to classify individuals into classes.

Conclusion:-

The aim of ontology is to achieve a common and shared knowledge that can be used by people and different systems. Ontology play an important role in achieving interoperability across organizations and on the Semantic Web, because they capture domain knowledge and their role is to create semantics explicitly in a meaningful way and providing the basis for meaningful understanding between parties. Ontologism has become a popular research topic in many communities. We have described a reliable and an efficient system, which suggests the user all the effective details to know about an educational domain. It is reliable because though it can be inputted with synonymous words and misspell, it retrieves the similar result and does not provide an irrelevant results. it saves the user's inconvenience to move on to more pages to search for the more result. The system can be further refined of with more words in the search interface which can yield more filtration of the query result. The system can be better used with more performance indicators which can better model user requirements.

References:-

- T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web", Scientific Am, May 2001
- D. Fensel, F.V. Harmelen, I. Horrocks, D.L. McGuinness, and P.F. Patel-Schneider, "OIL: An Ontology Infrastructure for the Semantic Web", Presented at IEEE Intelligent Systems, 2001
- R. Neches, R. Fikes, T.W. Finin, T.R. Gruber, R.S. Patil, T.E. Senator, and W.R. Swartout, "Enabling Technology for Knowledge Sharing", Presented at AI Magazine, 1991
- N.F. Noy, "Semantic Integration: A Survey Of Ontology-Based Approaches", Presented at SIGMOD Record, 2004
- Y. Qu, W. Hu, and G. Cheng, "Constructing virtual documents for ontology matching," inProc. 15th Int. World Wide Web Conf., 2006
- http://www.semanticweb.org/
- http://www.w3.org/
- http:// Wikipedia.org