“Role of Search Engines in Intelligent Information Retrieval on Web.”

Ramprakash¹, Sanjay Kumar Malik², Nupur Prakash³, SAM Rizvi⁴, Sudeep Marwaha⁵

¹M.Tech.(IT) IV Sem, University School of Information Technology, GGS Indraprastha University, Delhi
²University School of Information Technology, GGS Indraprastha University, Delhi
³University School of Information Technology, GGS Indraprastha University, Delhi
⁴Dept. of Computer Science, Jamia Millia Islamia, Delhi
⁵Scientist (S.S), Indian Agricultural Statistics Research Institute, Pusa, Delhi

ramprakash.usit@gmail.com¹, sdmalik@hotmail.com², nupurprakash@rediffmail.com³, samsam_rizvi@yahoo.com⁴, sudeep@iasri.res.in⁵

ABSTRACT
With the growth of the web, information explosion has taken place in the form of “Big Bang”. A well-support semantic based search engine needs to display the few specific pages from the billions available in which users have interest. Therefore, search engines have become one of the most important and helpful tools for obtaining information from the Internet. But most of the search engines are suffering from drawbacks of human and time consumption. Instead of caring about the exact meaning of semantics of information, the machines on the current web are caring about the location and display of information only. Because of this shortcoming of the current web, the search results produced by even most popular search engines are not satisfactory. Search engine activities in current web are not particularly well supported by software tools except for keyword-based search and suffer from lack of semantics and as well as semantic based search engine also lack of knowledge of semantics like RDF, Ontology, Query language like SQL. This paper presents an analysis of the role of search engines in intelligent web and to present a theoretical model to explore some method and possibilities for intelligent and meaningful retrieval of information on web for the ordinary users who are not the aware of semantic knowledge.

KEYWORDS
Search Engine, Intelligent Retrieval, RDF, IIR, Semantic Database

1.0 INTRODUCTION
With the growth of the web [1], information explosion on the Internet has taken place in the form of “Big Bang”. Search engines are the most commonly tools for finding relevant information from the Internet. But today’s search engines are not perfect in intelligent information retrieval. It is common that search engines display many web page links which are not the user’s need. However, users give the right keywords and search engines display pages involving these keywords and, yet the results are useless and irrelevant. It is extremely difficult for experienced searchers to choose the exact terms to retrieve the intelligently meaningful information [2]. Current web which is biggest collection of database or information lacks the existence of semantic and as a result the information available in the web is only human understandable, not machine. So that, a search engine activity in current web are not particularly well supported by software tools and suffers from lack of knowledge of semantics. However, there are lots of search engines available but the way to retrieve meaningful information is difficult. To overcome this problem in search engines to retrieve relevant and meaningful information intelligently, semantic web technology deals with a great role. Only this technology gives the desire results by search engines to the user.

In this paper, we present the role of Search Engines in intelligent information retrieval on web and describe the theoretical model of intelligent information retrieval (IIR) search engine.

2.0 INTELLIGENT INFORMATION RETRIEVAL (IIR) SEARCH ENGINE: INTRODUCTION
Information retrieval in the World Wide Web is become an active area. Information retrieval for searching information on the web is not a fresh idea but has different challenges when it is compared to general information retrieval. Different search engines return different search results due to the variation in indexing and search process. None of the search engines come close to indexing the entire web, much less the entire Internet. Content not indexed. The distributed nature of information has two kinds of research problems in search engine that must be displayed in order to effectively answer to queries

• Information saved in semantically different and mostly without any semantic annotation at all. Here, a question arises that is, how can a search engine map a query to documents where information is available but does not retrieve in intelligent and meaningful information?
• The results of a query produced by search engines are distributed across various documents that may or may not be strongly connected with a hyperlink. How can a search engine recognize such a huge distributed result?
The concept of Semantic Web [1, 4] can solve the first problem in the World Wide Web with semantically annotations to produce intelligent and meaningful information by using query interface mechanism and ontologies. The second problem can be solved by graph-based query models [5, 6] but it does not clear how these methods can be applied when explicit links are not available. To overcome the problems imposed by traditional search engine, there are various number of semantic search engine have been developed [7, 8]. These semantic search tools only enhance the performance of traditional search engine but not suitable for ordinary end users who are not familiar with semantic database, ontologies, or SQL-like query languages. In this regard, we propose theoretical model of intelligent information retrieval (IIR) search engine. The intelligent information retrieval (IIR) search engine we present theoretical model here, gives several means to address these issues.

- IIR Search Engine can tackle the problems of knowledge by providing initial phase which is strongly supported by software tools.
- IIR Search Engine can deal with the complex queries and convert into formal queries by using Perfect Match Algorithm.
- IIR Search Engine should produce meaningful information which is really user’s need.

So that, IIR search engine would make it possible for simple user to extract intelligently meaningful information without having specific knowledge of semantic web and need not to learn query language like SQL.

3.0 THE UNDERLYING TECHNOLOGIES IN SEMANTIC SEARCH ENGINE

In this section, we represent how present semantic search technologies handle user’s need? There are mainly four different strategies of semantic search engines according to users demand.

- Form based Search engine
- RDF based Search engine
- Semantic based keyword Search engine
- Question –Answering Search engine

The SHOE search engine [9] is an example of form based semantic search engine. This type of search engine supplies web forms so that users can easily specify queries. Further, they may have problems to express the queries on the information retrieval. The TAP search engine [7] is an example of semantic based keyword search engine. There are two steps search process i.e. (i) to find a perfect and suitable for the user entered keyword and (ii) retrieve information which are highly related to the match of the users entered keyword. Such types of search engine frequently deliver comprehensive support the clustering of search results. Both these strategies are able to increase the search performance by using semantic data knowledge. There is also one problem of knowledge overhead which is essential for users to be having proper knowledge of ontologies, RDF so that users can easily retrieve meaningful information easily.

4.0 NEED FOR INTELLIGENT INFORMATION RETRIEVAL

There is need of exploring more and more strategies for the intelligent retrieval of information on World Wide Web. Aim is to develop an intelligent information retrieval search engine. There are following key issues for the requirement of intelligent information retrieval from the search engines.

- It should hide the complexity of semantic search from end users.
- To support multiple keywords.
- To avoid the view – based search engine.
- It should give quick response of the given queries.
- Results displayed by search engines must satisfy the user queries and also easy to understand.

These requirements meet, if we take keyword based search technique instead of question answering technique in terms of searching. The problems faced by current keyword based semantic search engines can be overcome by providing Initial Phase where highly supports complex query having multiple keywords to extract meaningful information from the web.

5.0 PROCESS OF INTELLIGENT INFORMATION RETRIEVAL [IIR] SEARCH ENGINE

Figure 1 [10] shows the framework of IIR search engine. In this framework, several phases are used to separate the end users from the semantic database.

Figure 1. Framework of IIR search Engine
The function of each phases [11, 12, 13] are explained as:

- **Initial Phase**: This phase is used to permits the users to mention queries in terms of keywords. It extends the feature of the traditional search engine by query subject and mixing of multiple keywords.
- **First Phase**: The main task of this layer is to provide exact meaning of user entered keywords. It has two main subparts. The first part is entity index for semantic used to indexes documents and their semantic entities and other part is semantic entity search engine used to match the user’s keywords.
- **Second Phase**: This phase produces returned search result for user queries into formal queries.
- **Final Phase**: This final phase is used to provide a meaningful and specific formal query language that can be used to retrieve semantic relations from semantic data layer.
- **Semantic Database**: It contains semantic metadata which come from various data sources. This layer is maintained by using repository languages like SQL and PosGreSQL.

In this Framework, supported software tools are used to retrieve exact and meaningful information from the semantic database and can be used at each layer. This phase is equipped with supported software tools developed by like Java, RDF Schema, HTML and Java Servlets and Jena.

### 5.1 THE SEARCH PROCESS

There are following steps involved in IIR search process:

- **Make the sense of the users query**: It is the first step of the search processing. The main function of this process is to find out the best and exact meaning of the users query. For this purpose, blocks of the semantic entities are used to finding the semantic match. Although, process is not easy because more than one semantic entity matches the keyword. To avoid this situation, all matched semantic entities are first indexed in semantic database. It then, searches the indexed semantic database directly to retrieve the exact meaning of keywords. For this purpose, semantic entity index engine and semantic entity search engine are developed in the respective search engine to make the sense of the users query.

- **Translate the user queries into well-defined formal queries**: In this step, IIR search engine get the well defined formal queries as an output by taking semantic matches as input. There are two types of queries.
  1. **Simple query** where the types of semantic entity match combinations are already fixed.
  2. **Complex query** where the number of combination gets big when there are many keywords involved and there are lots of matches for each keyword.

   By using Perfect Match Algorithm for Query (explain in next section 5.2), the complex query problem can be solvable.

- **Query the back-end semantic database**: In this step, search engine query the back-end semantic database, which is produced by well defined formal queries.
- **Provide the rank**: The last step is to provide the best rank to the users query results intelligently according to user’s demand.

### 5.2 PERFECT MATCH ALGORITHM FOR QUERY

The Perfect Match algorithm constructs queries by walking through all the appropriate matches. It includes three main parts.

- **Initial block**: It deals with what user’s needs to be retrieved from the World Wide Web.
- **Retrieve block**: This block addresses how to retrieve the triples [14].
- **User’s Condition block**: In this block, conditions need to be satisfied according to user’s need are generated.

Figure 2 refers to the Perfect Match Algorithm for query.

```plaintext
1. Perfect Match Algorithm //Algorithm for query
2. \[//\] Instances of the subject which either have
3. relations with all the keywords or have relations
4. with some of the keywords
5. Initialize the query \[//\] It initialize the user
6. query when having multiple keyword for match
7. break the query into blocks
8. \{ for all blocks match the blocks with keyword
9. (loop 7 -22)
10. if keyword match
11. then check for class
12. if class match
13. then add query blocks for class-class
14. relations retrieval
15. move to next
16. else
17. if property match
18. then add query blocks for class-
19. property relations property
20. move to next
21. else
22. compose the queries by using the blocks
23. exit loop
```

Figure 2. Perfect Match Algorithm

### 6.0 CONCLUSION

Existing semantic search engines merely provide accurate answers to user queries. However, this type of search engines are only suitable for ordinary users who are not familiar with the semantic data, ontology and query languages and even do not know how to work on this type of search engine. So, our survey of underlying technologies of semantic search engine
reveals that current search engines provide little support for end users. In contrast with these efforts, our prototype and proposed model of IIR search engine furnish many purposes to address this issue.

- Initial phase, which can overcome the problems of knowledge overhead.
- Supporting relatively complex queries and breaks them into formal query.
- Providing relatively quick response to produce meaningful information retrieval from World Wide Web.

7.0 FUTURE SCOPE
Search Engines provide a foundation in web information retrieval. The proposed prototype is an attempt to overcome the problems of knowledge overhead and gives an idea to obtain a quick and meaningful response of complex queries. It may be beneficial in developing a case study on search engines and in intelligent information retrieval on web.

8.0 REFERENCES

[17] I. Rogers, “The google pagerank algorithm and how it works.”
[23] L. Page, S. Brin, R. Motawani and T. Winograd, “The pagerank citation ranking: Bringing order to the web: