Comparative Study of Syntactic Search Engine and Semantic Search Engine: A Survey

A.C. Santha Sheela
Assistant Professor, CSE Dept
Sathyabama Institute of Science & Technology ,
Chennai, India

Dr.C.Jayakumar
Professor , CSE Dept
Sri Venkateswara College of Engineering
Chennai, India

***Abstract* - Search engine symbolizes an extremely powerful and valuable tool for fetching any sort of information from Internet. There has been numerous researches carried on search engines techniques, the major ones are syntactic and semantic. Referring to the Syntactic web, the results obtained are purely as per the keyword match. That is the query outputs numerous web pages against the keyword match that may not even be relevant or meaningful. Whereas, unlike the syntactic web, the semantic web is a revised or upgraded version of the web which produces quiet meaningful and specific output as it has the potential to comprehend the query effectively. Few examples of Semantic based search engines include Kosmix, Hakia, Cognition, Swoogle and Lexxe. Whereas syntactic based search engines are Google, Yahoo, Ask. The work performs a comparison amidst the performance of semantic and syntactic based search engine and evaluates them by employing certain queries.**

Keywords : search engine, syntactic search engine, semantic search engine, Information Retrieval (IR), Semantic Web.

I. INTRODUCTION

Search engines are basically utilized for searching and fetching any sort of information from the WWW or Internet. The user can get desired info based on his/her query using the Search engine. The information produced by the search engine can be a collection of text, images, video or audio. Information can be searched across the Internet by the means of 2 approaches: syntactic and semantic.

The Syntactic search engine fetches information from the Web based on user query but the resultant web pages are neither relevant enough nor in any specific order. That is, Syntactic based search engine are unable to comprehend the interpretation of the user query thus yielding in output with no specific order. Whereas in Semantic search engine the search is carried across the semantic web.

The content present in a semantic web is depicted in form of HTML form, RDF(Resources Description Framework) and OWL (web ontology Language). Moreover, here entire information is stored semantically, thus illustrating the output precisely by comprehending the keywords. Semantic web is a blend of techniques like ontology matching, word disambiguation and word clustering that forms a connection or link amidst syntax and semantics for building semantic web.

With the presence of voluminous information prevailing across the Internet, there exists numerable duplicate web pages on various websites, for instance in news and blog web page keep on increasing rapidly. Hence, the web mining research communities primarily concerns over semantic search in order to fetch desired specific documents from the WWW.

II. RELATED WORK

Cognitive Ontology Enrichment for Semantic Information Retrieval for semantic search system is obtained by first converting HTML to XML, next converting XML to OWL Mapping and performing search using SPRQL [1]. Semantic search engines likes woogle, falcon, SWSE have structured data that handles data in OWL or RDF format only. Moreover, XML takes in accord the Syntactic level but not the reasoning [2]. Employing SWSE(Semantic Web Search Engine) for Searching and Browsing Linked Data: [3].

Retrieval Comparison is carried out by relying upon Precision and Recall for information. It's depicted that semantic precision is 0.72 and syntactic precision is 0.42[4].Varying work survey is generated concerning this domain. Based on the precision and recall result, the evaluation of search engine's performance is performed.

III. PROPOSED SYSTEM

Here a special scheme is being proposed considering short sentence research in order to keep it distinctive. The system reveals similarity in the sentence on the basis of verification of repeated words via semantic analysis. It evaluates the sentence and thereafter divides it in 2 sections. One section will hold raw semantic vector and the other will hold order

978-1-7281-1599-3/19/$31.00 ©2019 IEEE

vector. The proposed system carries out such operations to search similarity in various sentences within the same document.

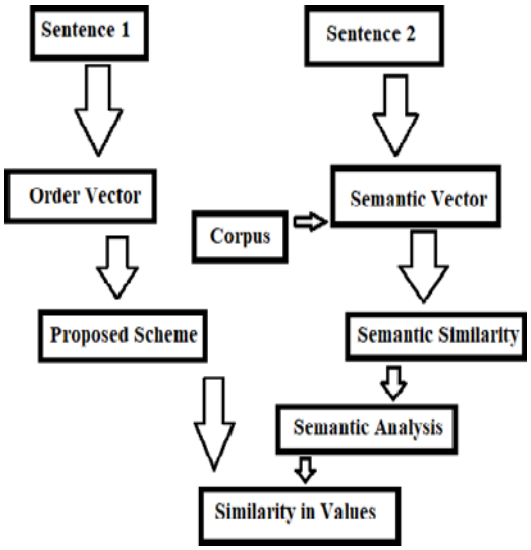

Fig. 1. *Semantic Analysis*

The system is evaluated by providing similar sentences along with similar sort of formation. By the means of semantic analysis, the separation is also carried out. Since there are various operations involved, time must be taken into consideration for performing similarity checking. Sophisticated model are designed for better products.

Though time consumption is not fulfilled as per satisfaction, but the similarity checking is achieved which holds extremely significant. Moreover there is an achievement of high similarity checking level. The system depicts few separation examples as depicted in Fig. 2.

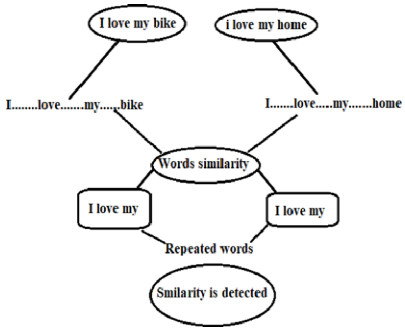

Fig. 2. *Level of knowledge*

Fig. 2 clearly depicts the system in the flow diagram.Initially, the user sentences are separated into various format, therafter they are evaluated by the semantic models by providing the values. Lastly, the sentences are transmitted to the user alon with a similarity report by the means of semantic analyzing. The proposed system involves various keywords to fulfill. Comparion of the existing system depicts that the simlarity checking level has been achieved and values are reported in-depth. In addition, sentence formation errors can be evalautyed by assessing the words on the basis of verbs and nouns. Using the proposed system there is an achievement of high similarity checking level. It yields in accuracy and similarity valueswhen used along with semantic analysis.

Table.1. *Sentence Comparison*

Sentences	Similarity Value	Detected
I love my bike I love my home	0.56	I love my
I love my job I like to go job	0	NIL

Table1 depicts the comparison of syntactic and semantic approaches based on different fundamental search facilities such as programming languages, functionality, nature etc…The proposed system strongly analysis the following: based on the comparison of no: of sentences, similarity by semantic analysis is nearly 0.561, in case there are two different sentences, similarity value is zero. The similarity check is achieved successfully by evaluating the sentences, rest of the comparison is carried out to obtain similarity level.

IV COMPARISONS BETWEEN SEMANTIC BASED SEARCH ENGINES AND SYNTACTIC BASED SEARCH ENGINES

978-1-7281-1599-3/19/$31.00 ©2019 IEEE

Segments For Comparison	Syntactic Search Engine	Semantic Search Engine
Definition	Structural characters with symbolic expression of language or phrases of the Language.	Complete Contextual expression of language and things or objects
Connections related	Syntactic characters or representation have some unique meanings.	Semantic expression includes syntactic characters and its assigned complete meanings of symbols, character, words and phrases. Meaning change of group of words.
Functionality	It does not use any methodology just matches exact keyword which is available in the website though ranking algorithm.	Ontology helps the search engine to identify the relation and understands it through keywords functionality of meaning through richer data integration.
Programming languages	HTML and XML are used to create meta data	RDF and OWL is used to create meta data
Nature	Forms of programming structure	Meaning of programming structure
Errors limitations	Syntax analyzers receive inputs, in form of tokens from lexical analyzers. Lexical analyzers are responsible for the token validity supplied by syntax analyzer. Syntax analyzers have following disadvantages such as 1.Token validity cannot be determined 2. Performing token validity cannot be determined 3.Declared token before usage cannot be determined 4. Initialized token before usage cannot be determined. These above tasks are easily achieved by semantic analyzer.	Semantic analyzer recognizes Typing mismatch, undeclared variable, reserved identifier misuse, multiple declarations of variables in scope, accessing an out of scope variable, formal and actual parameter

Fig. 3. *Comparisons between semantic based search engines and syntactic based search engines*

V ESTIMATION OF PRECISION AND RECALL

As with the rapid growth and popularity of WWW, many search engines have also spread its roots like Google, Yahoo, Hakia, Ask, Kosmix, Lexxe, cognition, Swoogle and so on. Amidst all prevailing search engines today, there are handful ones which offer relevant and specific information as desired by the user. The two factors namely precision and recall are used for measuring the accuracy or relevance of any information.

Precision (P) is the fraction of retrieved documents that are relevant.

Precision is the Number of relevant documents retrieved divided by Number of retrieved documents.

Recall (R) is the fraction of relevant documents that are retrieved.

Recall is the Number of relevant documents retrieved divided by Number of relevant documents.

The research work takes into account 6 different sort of search engine namely Google, Yahoo!, Wikipedia for syntactic based and Hakia, Bing and DuckDuckGo for semantic based in order to determine the performance of these search engine and the one resulting in the most desired information. For carrying out the task, queries are fetched sports fields including Cricket, Stump, Net, Marathon, and Indoor Stadium.

Table 2 depicts the query list. Input queries from Fig 4 are fed into both syntactic and semantic based search engine. On the execution of the query, 20 topmost we pages are manually classified as relevant to the query and irrelevant to the query. Output depicts that Google performed against Bing and Hakia but laid down second after DuckDuckGo. It's illustrated from the experiment that semantic based search engine yields better and effective output in contrast to syntactic based search engine.

Table 2: *Input query list from Sports Domain*

Query no	Query	Query no	Query
Q1	Cricket	Q6	Basket ball
Q2	Marathon	Q7	Individual
Q3	Stump	Q8	Swimming
Q4	Net	Q9	Skating Board
Q5	Indoor Stadium	Q10	Chess

Query/ Search Engine	Google	Yahoo!	Wikipedia	Hakia	Bing	Duck Duck Go
Q1	17	16	20	16	10	17
Q2	17	13	5	14	17	16
Q3	14	17	12	9	12	17
Q4	14	19	13	8	10	12
Q5	11	8	15	11	13	7
Q6	15	9	11	10	8	19
Q7	10	12	17	12	14	14
Q8	6	12	18	15	16	14
Q9	20	19	19	19	20	20
Q10	14	12	10	11	15	13
Total	138	137	140	125	135	139
Average	82.8	82.2	84	75	81	85.4

Fig.4. *Relevant Document Retrieved from syntactic and semantic search engine*

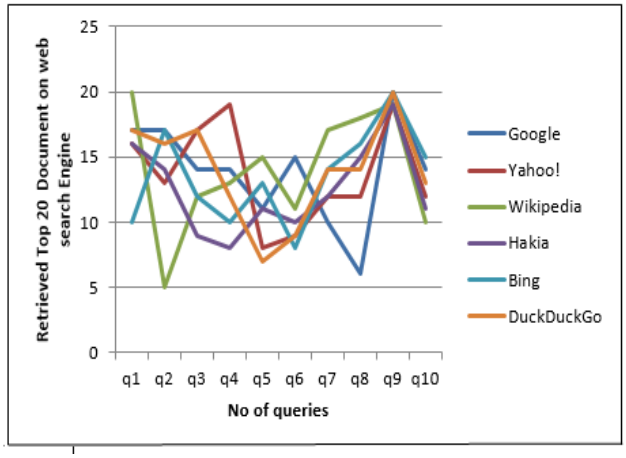

Fig 5: *Retrieved Relevant Document*

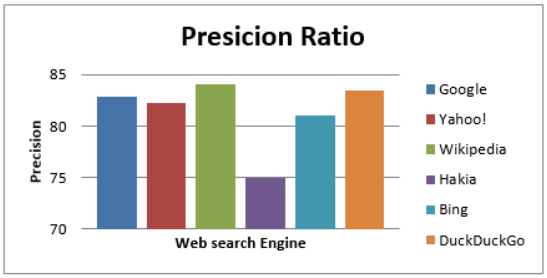

Fig 6: *Precision Ratio of different search engine*

V. CONCLUSION

The work illustrates regarding semantic analysis taking into account the repeated words present in the sentences. For every search, similarity value is assessed to ensure precise evaluation depending upon the sentences. Numerous syntactic and semantic search engine are bring compared and considered. It's elucidated from the experiments that the semantic search engine exhibits enhanced performance in contrast to syntactic search engine when relevant document needs to be fetched, measured by the factors of precision and recall. Fig 5 and Fig 6 depicts the output. The semantic search engine 'DuckDuckGo' exhibits better performance compared to Google.

REFERENCES

[1] G. Nagarajan , K.K. Thyagharajan, "cognitive Ontology Enrichment For Semantic Information Retrieval", National Journal Of System And Information Technology, Vol 5, No 2 , 2012,pp 168-179.

[2] Ranjna Jain, Neelam Duhan and A.K. Sharma, "Comparative Study on Semantic Search Engines", International Journal of Computer Applications (0975 – 8887), Volume 131 – No.14, December 2015.

[3] Aidan Hogan, Andreas Harth , J• urgen Umbrich , Sheila Kinsella , Axel Polleres , Stefan Decker, "Searching and Browsing Linked Data with SWSE: the Semantic Web Search Engine", Journal of Web Semantics, Vol 9 no.4 , 2011.

[4] Sanchika Gupta & Dr. Deepak Garg "Comparison of Semantic and Syntactic Information Retrieval System on the Basis of Precision and Recall", International Journal of Data Engineering (IJDE), Volume (2) : Issue (3), 2011.

[5] Carvalho, C., de Moura, E. S., Veloso, A., & Ziviani, N, " Website replica detection with distant supervision",Information Retrieval Journal, 21(4), 253-272, 2018.

[6] Luciv, D. V., Koznov, D. V., Chernishev, G. A., Terekhov, A. N., Romanovsky, K. Y., & Grigoriev, D. A., "Detecting near duplicates in software documentation" Programming and Computer Software, 44(5),2018, pp 335-343.

[7] Santha Sheela A.C and Jayakumar C."Duplicate Web Pages Detection With The Support Of 2D Table Approach", Journal of Theoretical and Applied Information Technology,Vol. 67, No.1, SEP 2014.

[8] Lou Junwei, Xiao, "Resaerch on Information Retrieval System based on Semantic Web and multi-agent",ICICCI,IEEE 2010.

[9] D. Tumer, M. A. Shah and Y. Bitirim"An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia," Fourth International Conference on Internet Monitoring and Protection, 2009.

[10] K.Varsha,P.Sandhya and K.Supreet, "Information Retrieval: Today and Tomorrow" IJCA, Vol.116, 2015.

[11] M. Andago, P.L Phoebe and A. M Thanoun, "Evaluation of a Semantic Search Engine against a Keyword Search Engine Using First 20 Precision," In Proceedings of the 29th Annual International Conference on Research andDevelopment in Information Retrieval, ACM Press, pp. 735–746, 2010.

[12] Albertoni R., Bertone A., & De Martino M., (2004), "Semantic Web and Information Visualization", Proceedings of the 1st Italian Workshop on Semantic Web Application and Perspective, DEIT, , Ancona, Italy, December10, 2004, pp.108-114

[13] Santha Sheela A.C and Jayakumar C., "Efficient Semantic Based Web Page Duplicate Detection Using SWP-RMA", Mitteilungen Klosterneuburg,Vol. 69(2), 2019.

[14] Hakia search engine availabeat :http://www.hakia.com/

[15] DuckDuckGo available at:htt://duckduckgo.com/