IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 6, Ver. II (Nov.- Dec. 2017), PP 01-06
www.iosrjournals.org

Web Search Optimization Using Semantic Web

[#1] Shaifali Nagar, Prof (Dr.) Mohit Gangwar[*2]
[#] *M-Tech Research Scholar, Department of CSE, Bhabha Engineering Research Institute Bhopal, (M.P.) India*
[*] *Principle (Department of CSE), Bhabha Engineering Research Institute Bhopal, (M.P.) India*
[#] *shaifali.nagar@yahoo.com,* [*] *mohitgangwar@gmail.com*

***Abstract**: An IT has developing in to the mature phase, this do not expect this state should continue. Something has to be done to certify that this can learn from other people by taking as much information or knowledgably data as this can and make it meaningful. Due to presence of vast amount of resources on the Web thus poses a serious problem of precise search. This is mainly because today's WORLD WIDE WEB (WWW) is a human-readable Web, where information data cannot be easily managed by machine. Highly cultured, efficient keyword based search engines that have grew today have not been able to bridge this breach. Ontology describes a set of representational primitives with which a field of knowledge is demonstrated. In our proposed algorithm Web Search Optimizing on Cloud Host base Semantic Web using Ontology (WSO-CHSWO), were Cloud-based computing is an developing repetition that offers significantly more organization and financial flexibility than outdated computing models. We comes up the idea of the Semantic Web machine interpretable information to make a machine process able form for stating information. Ontologies are singing very vital part in many areas such as intellectual information data, knowledge supervision and organization, electronic commerce base data. Today, search engine crawlers are regaining billions of sole URL's or web page. Based on the semantic Web technologies we proposed in this paper the design methodology of a semantic Web search engine with using ontology which provides precise search results for a domain specific search.*

***Keywords:** sole documents, intellectual information data, knowledge supervision, detecting replicate, organization, electronic commerce base data, replication, search engine, sole URL's, web page, Ontology, Semantic Web.*

Date of Submission: 27-10-2017 Date of acceptance: 18-11-2017

I. Introduction

Knowledge base data is so important in this current era for us in all aspects. While many sources have given decent numbers of info but there are still lacking in terms of knowledge reliability. An organization's ability to learn and handle knowledge processes or knowledge product is considered the new key accomplishment factor [1], investigation in information and knowledge recovery are actively conducted. They are also useful in avoiding researchers from digging every single document to information searching. But, regaining the real meaning of data is often fails to give the chosen result. The increasing size of data available on the Web as well as data's lack of creation, multidimensionality, large size and active evolution make information retrieval a tedious and hard task. There are 2 ways to progress the quality of web mining, one is to use improved mining technology for the present resource and the other is to make the web resource more machine logical for computer to process [2]. The last periods, the amount of web-based information data available has improved intensely optimized.

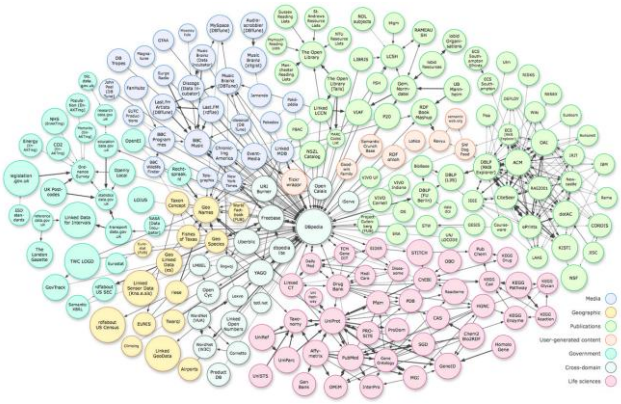

Figure 1: Linking Open Data cloud diagram

As presented earlier the term ”Semantic Web“ encompasses efforts to build a new WORLD WIDE WEB (WWW) architecture that enhances content with official semantics. This will allow computerized agents to reason about Web content, and carry out more bright tasks on behalf of the operator. The relation between “ontology”, “metadata” and “Web documents”. It depicts a small part of the CIA world fact book ontology. with semantic annotations given in an XML serialization of RDF-based metadata descriptions. For the country and the organization there are metadata definitions denoted by corresponding uniform resource identifiers The URIs are typed with the ideas COUNTRY and ORGANIZATION.

In adding, there is a association instance between the republic and association. This scenario is common to many people on the World Wide Web (WWW). A major tricky with searching on the Web today is that data obtainable on the Web has little semantic organization beyond simple operational arrangement of text, declared keywords, titles, and abstracts. As the Web expands exponentially in size, this lack of organization makes it very hard to efficiently glean knowledge from the Web, even with state-of-the-art natural language processing techniques, index mechanisms, or the assistance of an army of data-entry workers assembling hand-made Web catalogs. In short, there is no effective way use the World Wide Web (WWW) to answer a query like: The chief intent of HTML and HTTP is to assist operator level presentation and steering of the Internet; automated search or sophisticated knowledge-gathering has been a much lower priority.

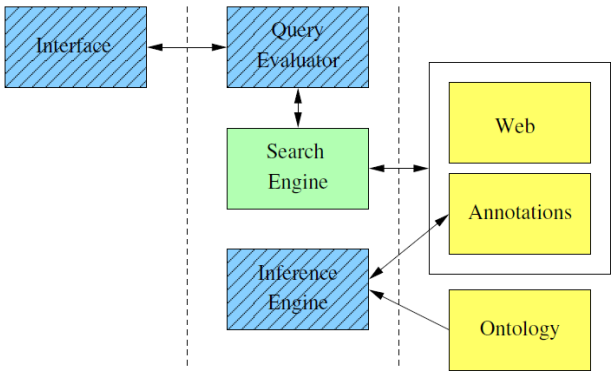

Figure 2: System Architecture

A major disadvantage of hand-built catalogs is the man hours required to construct them. Given the size of the World Wide Web (WWW), and the rate at which it is growing, cataloging even a modest percentage of web pages is task. Additionally, the criteria used in building any catalog may turn out to be orthogonal to those of interest to a user.

Ad-hoc robots that attempt to gather semantic information from the web typically gather only the limited semantic information inferable from existing HTML tags. The current state of natural language processing technology makes it difficult to infer much semantic meaning from the body text itself at a reasonable rate (if at all). In our experience, even limiting a web robot’s natural language understanding to a small topic like Computer Science Web pages still proves surprisingly difficult to implement, and like many ad-hoc methods, such algorithms are extremely brittle. Further, none of these approaches (except perhaps the last, for specific domains) allows for inferences about relationships *between* web pages, aside from simple facts about linkage.

Instead of trying to glean knowledge from existing HTML, another approach is to give authors the ability to embed knowledge directly into HTML pages, making it simple for user-agents and robots to retrieve and store this knowledge. The straight forward way to do this is to provide authors with a clean superset of HTML that adds a knowledge markup syntax; that is, to enable them to directly classify their web pages and detail their web pages’ relationships and attributes in machine-readable form using HTML.

Using such a language, a document could claim that it is the home page of a graduate student. A link from this page to a research group might declare that the graduate student works for this group as a research assistant. And the page could assert that “Laptop” is the graduate student’s last name. These claims are *not* simple keywords; rather they are semantic tags defined in some “official” set of attributes and relationships (an *ontology*). In this example the ontology would include attributes like “last Name”, classifications like “Person”, and relationships like “employee”. Systems that gather claims about these attributes and relationships could use the resulting gathered knowledge to provide answers to sophisticated knowledge-based queries. Moreover, user-agents or robots could use gathered semantic information to refine their web-crawling process. However, if the agent gathered semantic tags from Helena Laptop’s web page which indicated that Laptop was her last name, then the agent would know better than to search this web page and its links.

II. Literature Survey

In this paper [3], The vast availability of information, that added in a very fast pace, in the data repositories creates a challenge in extracting correct and accurate information. Which has increased the competition among developers in order to gain access to technology that seeks to understand the intent researcher and contextual meaning of terms. While the competition for developing an Arabic Semantic Search systems are still in their infancy, and the reason could be traced back to the complexity of Arabic Language. It has a complex morphological, grammatical and semantic aspects, as it is a highly inflectional and derivational language.

In this paper, we try to highlight and present an Ontological Search Engine called IBRI-CASONTO for Colleges of Applied Sciences, Oman. Our proposed engine supports both Arabic and English language. It is also employed two types of search which are a keyword-based search and a semantics-based search. IBRI-CASONTO is based on different technologies such as Resource Description Framework (RDF) data and Ontological graph. The experiments represent in two sections, first it shows a comparison among Entity-Search and the Classical-Search inside the IBRI-CASONTO itself, second it compares the Entity- Search of IBRI-CASONTO with currently used search engines, such as Kngine, Wolfram Alpha and the most popular engine nowadays Google, in order to measure their performance and efficiency.

The internet contains [4] vast amount of information that the search engines are able to provide search results that are based on page ranks. But the search results are not related to one particular user's environment. The main objective of this paper involves with search engine and search engine optimization methods. Here a new technique called as ontology search logs is introduced, which will be used for customized search logs according to the user define input. This application will be processed in any of the search engine. In this paper, a new system called as Semantic Search log Social Personalized Search is also proposed which would be able to provide results for search query that relates to a particular user's environment, his area of interests, his likes and dislikes, the data the he/she might have found to be useful for him while searching. Social networks are the domain in which it could obtain such user oriented information, which could be used for providing personalized search results. Supervised learning technique is used to learn about the user based upon his interactions inside the system. This process can be able to make applicable for each and every registered user in this application. User can give their basic information in their profile and get benefits from their each and every search. When the user getting register with the system it creates an ontological profile, according to the profile created by the user and when he/she getting login into the social network and interacts with it the system updates his/her ontological profile based upon the interaction. The search provision can be finding out in their home page after they get login. When the user searches a keyword using the search engine inside the social network, it refers to the ontological profile of the user and displays the Personalized Search results. The system should be able to intelligently identify whether a search result has been useful to him or not and save it for his future reference when he searches for the same or similar keyword next time.

In this paper [5] Data accuracy and reliability have been a serious issue in the vast emergence of information on the web. Advanced web searching has assisted in knowledge retrieving. However, most knowledge on the Web is presented in natural-language text that understandable by human but difficult for computers to interpret. Therefore, Semantic Web approach is widely used to give more reliable application. This paper presents a framework in enhancing knowledge retrieval processes using Semantic Web technologies. Instead of using ontology and categorization alone, we are injecting personalization idea from Relational Database (RDB) to ensure more reliable data are obtained. The proposed framework is discussed in details. A case study is presented to see the viability of the proposed framework in retrieving the meaningful information.

In this research paper author Nizar R. Mabroukeh [6] By the advent of social networks and tagging systems, The Internet has recently witnessed a big leap in the use of Web Recommendation Systems WRS. WRS are still limited by several problems, of which are sparsity, and the new user problem. They also fail to make full use and harness the power of domain knowledge and semantic web ontologies. In this article, we discuss how an ontology-based WRS can utilize relations and ideas in an ontology, along with user-provided tags, to provide top-n recommendations without the need for items clustering or user ratings. For this purpose, we also propose a dimensionality reduction method based on the domain ontology, to solve the sparsity problem.

A content-based web reference system is proposed based on a domain ontology. It relies on user-provided tags, that are mapped to ideas of this ontology. Similarity measures are used during mapping and a matrix of items ideas is built offline, which is used later for online top-n reference their system outperforms popular algorithms like Top-Pop and NNCosNgbr. There are two benefits in using an ontology over clustering of the tags. First, it saves the costly step of clustering, and second, a full ontology has a far better reasoning power than a topic taxonomy. In a full ontology there are numerous semantic relations that can be taken into thought to provide better understanding measures, and better interpretability.

In this research paper author Xiaohui Tao, Yuefeng Li et. al. [7] As a model for knowledge description and formalization, ontologies are widely used to signify user profiles in modified web information gathering. However, when on behalf of user profiles, many models have utilized only knowledge from either a global knowledge base or a user local material. they, a personalized ontology model is proposed for knowledge picture and reasoning over user profiles. This model learns ontological user profiles from both a world knowledge base and user local instance repositories. The ontology model is assessed by comparing it against standard models in web info gathering The results show that this ontology model is successful.

In this research paper author Xujuan Zhou, Sheng-Tang et. al. [8] It is well known that taking the Web user profiles into account can improve the effectiveness of Web mining systems. However, due to the lively and complex nature of Web users, automatically acquiring worthwhile user profiles was found to be very stimulating. Ontology based user profile can possess more accurate user information. their investigation emphasizes on acquiring search intentions information. they presents a approach of developing user profile for Web searching.

The model considers the user's search intentions by the process of PTM (Pattern-Taxonomy Model). Initial trials show that the user profile based on search meaning is more useful than the generic PTM user profile. Developing user profile that contains user search meanings is essential for real Web search and retrieval. Mixing ontology-based user profiles into the processing can be very beneficial for improving the competence of Web information search and retrieval.

III. Proposed Technique

Web mining is the use of data mining technologies to automatically interact and discover information from web documents, which can be in structured, unstructured or semi- structured form. XML has become very popular for representing semi structured data and a standard for data exchange over the web. The data based on XML is self described; it can be exchanged and handled without internal description.

There is tremendous amount of information and knowledge existing on the web and waiting to be discovered, shared and utilized. Ontology represents a set of precisely defined terms about a specific domain and accepted by this domain's community, ontology is an explicitly specification of a idealization.

All these have affected greatly the way web-based applications are designed and implemented with e-Learning systems could not include an exception. Besides, among all other "e" actions, e-Learning is one of the fastest growing and commonly accepted. For on line educational institute web site two important ontology's would need to be built one ontology describing all the educational services provided, with the relation between each other and the other ontology describing the web site. We have distinguished two stages in the whole process, one of offline tasks that includes data preparation, ontology creation and usage mining and one of online tasks that concerns the production of recommendations.

The proposed algorithms Web Search Optimizing on Cloud Host base Semantic Web using Ontology (WSO-CHSWO), starts with mapping the tags in a preprocessing step. Pages and their connected tags are stored in a database db. Compute the similarity between tags of each pages and each idea in the ontology, giving the similarity score Similarity(TP_i ; $IDEA_j$). To elaborate, each tag in pi is compared against each idea in the ontology, this is done by computing the similarity between the tag and the idea, both the tag and the idea are located as two words (say N1 demonstrating idea $IDEA_j$, and N2 demonstrating a tag TG from the set TP_i).

Tags Mapping: The web log is signified as tags comprising pairs of pages and their associated tags. Similarity is computed between each pages and each ontology leaf idea, and all similarity scores are stored in a matrix of pages ideas. Notice that the dimensionality of the matrix depends on the number of leaf ideas in the ontology.

Active User: As the active user arrives at a certain web page, the tags associated with this pages are re-claimed, and a vector is generated, that is similar to one row of the online *Web-Page database Db* generated in the previous step. The vector demonstrations the similarity between the active user tags and each idea. This vector is matched against each row in the matrix, and the top-n matching pages are used as the recommendation set.

Expanding data set: The recommendation set can be expanded by increasing n, and by expanding the active user vector using semantic relations in the ontology to include more ideas, not present in the matrix, from which reference can be drawn.

IV. Proposed Algorithm

Algorithm:

Web Search Optimizing on Cloud Host base Semantic Web using Ontology (WSO-CHSWO)
Input: Website host link, Tags, Pages, Domain Ontology

Output: Website base data, page rank data

Check each website authentication Ws-Val
Connect website host Ws-H
Analysis of website hosted database
Web-Page database Db,
Domain ontology DO,
Check website base tags
For each tags <Page P_i, its tags TP_i > in Db
For each idea $IDEA_j$ in DO
Calculate Similarity (TP_i ; $IDEA_j$), which is the average similarity between
Each tag in TP_i and $IDEA_j$
Check web-data downloadable
if <Page P_i, its tags TP_i > is valid then
if <Page P_i, its tags TP_i > is identical then
abounded data
Else
Download data in Db_r database,
Db_r [i; j] = Similarity (TPi ; IDEAj)
End if
End if
End for
End for
Return Db_r
Set top N recommended Page, P_S
Active User Vector, AUV_i
Extended set, S
Find the idea $IDEA_j$ to which AUV_i is highly similar
Rank all relations of $IDEA_j$ according to their relation hierarchy
For each relation connecting $IDEA_j$ with another idea C_j do
Instantiate idea C_j to generate Pages
Add pages to P_S, to get M
Store page rank in PgR_i
End for
Return M+
End

V. Results

The WORLD WIDE WEB (WWW) web search catches the information data from web sources (website host), may be from data warehouse and from own collective database. There is marvelous amount of information and knowledge are present on the WWW and to come to be discovered, communal and utilized.

VI. Conclusion

In our proposed algorithm Web Search Optimizing on Cloud Host base Semantic Web using Ontology (WSO-CHSWO), were Cloud-based computing is an developing repetition that offers significantly more organization and financial flexibility than outdated computing models. On which the idea of the Semantic Web machine interpretable information to make a machine process able form for stating information. Ontologies are singing very vital part in many areas such as intellectual information data, knowledge supervision and organization, electronic commerce base data. Today, search engine crawlers are regaining billions of sole URL's or web page. Based on the semantic Web technologies semantic Web search engine with using ontology which provides precise search results for a domain specific search.

The fundamental technique of semantic web searching is ontology. Ontology represents a set of precisely defined terms about a specific domain and accepted by this domain's community, ontology is an explicitly specification of a idealization.

Reference

[1] Maier, R., Hadrich, T., Peinl, R.: Enterprise Knowledge Infrastructures, 2nd edn. Springer, Berlin (2009)
[2] Edward H.Y. Lim, Hillman W.K. Tam, Sandy W.K. Wong, James N. K. Liu and Raymond S. T. Lee,Collaborative Content and User-based Web Ontology Learning System, IEEE, 2009, Pg 1050-1055

[3] Awny Sayed, Amal Al Muqrishi, "IBRI-CASONTO: Ontology-based semantic search engine",Egyptian Informatics Journal, Elsevier, 2017.
[4] P. Hema Priya, R. RangaRaj, "An Improved Search Engine by Semantic Web Using Ontology", International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, Volume 2 Issue 12, December 2013, World Wide Web (WWW).ijsr.net.
[5] Noryusliza Abdullah and Rosziati Ibrahim, "Semantic Web Search Engine Using Ontology, Clustering and Personalization Techniques" , B. Murgante et al. (Eds.): ICCSA 2012, Part IV, LNCS 7336, pp. 364–378, 2012, © Springer-Verlag Berlin Heidelberg 2012.
[6] Nizar R. Mabroukeh and C. I. Ezeife, "Ontology-based Web Recommendation from Tags", IEEE ICDE Workshops 2011
[7] Xiaohui Tao, Yuefeng Li, and Ning Zhong, "A Personalized Ontology Model for Web Information Gathering", Published by the IEEE Computer Society 2011
[8] Xujuan Zhou, Sheng-Tang Wu, Yuefeng Li, Yue Xu, *Raymond Y.K. Lau, Peter D. Bruza, "Utilizing Search Intent in Topic Ontology-based User Profile for Web Mining", IEEE/WIC/ACM International Conference, 2006

Input				Output	
S. No.	Web Host Link	Tags	Domain	Web Site Base Data	Page Rank Data
1	https://www.trivedieffect.com/the-science/publications/organic-compounds-publications/isotopic-abundance-analysis-of-biofield-treated-benzene-toluene-and-p-xylene-using-gas-chromatography-mass-spectrometry-gc-ms/	Gc-Ms Spectra of P-Xylene	TrivediEffect	Isotopic Abundance Analysis of Biofield Treated Benzene, Toluene and p-Xylene Using Gas Chromatography-Mass Spectrometry (GC-MS)	2
		Isotopic Ratio of (Pm+1/Pm) of Benzene			0
		Isotopic Ratio of (Pm+1/Pm) of P-Xylene			1
2	https://www.trivedieffect.com/the-science/tag/casein-enzyme-hydrolysate/	Casein Enzyme Hydrolysate	TrivediEffect	Evaluation of the Impact of Biofield Treatment on Physical and Thermal Properties of Casein Enzyme Hydrolysate and Casein Yeast Peptone	15
		Casein Yeast Peptone			11
		Physical and Thermal Properties of Casein Enzyme Hydrolysate			1
		Physical and Thermal Properties of Casein Yeast Peptone			1
3	https://www.trivedieffect.com/the-science/publications/organic-compounds-publications/characterization-of-physical-spectral-and-thermal-properties-of-biofield-treated-124-triazole/	1,2,4-Triazole	TrivediEffect	Characterization of Physical, Spectral and Thermal Properties of Biofield Treated 1,2,4-Triazole	11
		Thermo Gravimetric Analysis of 1,2,4-Triazole			1
		Dta Thermogram of Triazole			1
		Surface Area Analysis of Triazole			1
4	https://www.trivedieffect.com/the-science/publications/organic-compounds-publications/physical-thermal-and-spectroscopic-studies-on-biofield-treated-p-dichlorobenzene/	Para-Dichlorobenzene	Physical, Thermal and Spectroscopic Studies on Biofield Treated p-Dichlorobenzene	Physical, Thermal and Spectroscopic Studies on Biofield Treated p-Dichlorobenzene	102
		P-Dichlorobenzene			46
		P-DCB Chemical Reaction			14
		Mothballs and Naphthalene			0

Table: Page rank of website data

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Shaifali Nagar Web Search Optimization Using Semantic Web." IOSR Journal of Computer Engineering (IOSR-JCE) , vol. 19, no. 6, 2017, pp. 01-06