



## Web search and page ranking optimizing for images on cloud host base web semantic using ontology

<sup>1</sup>Shaifali, <sup>2</sup>Dr. Mohit Gangwar

<sup>1</sup> M-Tech Research Scholar, BERI, Bhopal, Madhya Pradesh, India

<sup>2</sup> Principal, BERI Bhopal, Madhya Pradesh, India

### Abstract

We study in this paper about the web search and page rank for topographies and high-level semantic concepts. The main purpose of the Semantic web and ontology is to integrate heterogeneous data and enable interoperability among disparate systems. 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 defines a set of representational primitives with which a domain of knowledge is modeled. Cloud-based computing is an emerging practice that offers significantly more infrastructure and financial flexibility than traditional computing models. Ontologies are playing very important part in many areas such as intelligent information retrieval, knowledge management and organization, electronic commerce. As the ontology is a standard, explicit and formalized description for shared conceptual model, the educational enterprises can be integrated by applying semantic express, shared knowledge described by ontology and automatic inference mechanism. In our proposed algorithm Web Search Optimizing on Cloud Host base Semantic Web using Ontology (WSO-CHSWO), where Cloud-based computing is an developing repetition that offers significantly more organization and financial flexibility than outdated computing models. The Semantic Web machine interpretable information to make a machine process able form for stating information. In this paper, we present the infrastructure of different websites for resulting the page rank.

**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

### 1. Introduction

The Artificial-Intelligence literature contains many definitions of an ontology; many of these contradict one another. For the purposes of this guide an ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins. Classes are the focus of most ontologies. Classes describe concepts in the domain. For example, a class of wines represents all wines. Specific wines are instances of this class. The Bordeaux wine in the glass in front of you while you read this document is an instance of the class of Bordeaux wines. A class can have subclasses that represent concepts that are more specific than the superclass. For example, we can divide the class of all wines into red, white, and rosé wines. Alternatively, we can divide a class of all wines into sparkling and no sparkling wines. In practical terms, developing an ontology includes: defining classes in the ontology, arranging the classes in a taxonomic (subclass–superclass) hierarchy, defining slots and describing allowed values for these slots, filling in the values for slots for instances <sup>[13]</sup>.

We can then create a knowledge base by defining individual

instances of these classes filling in specific slot value information and additional slot restrictions. An ontology typically consists of a hierarchical description of important concepts in a domain, along with descriptions of the properties of (the instances of) each concept. Ontologies will play a pivotal role in the Semantic Web by providing a source of shared and precisely defined terms that can be used in meta-data. The degree of formality employed in capturing these descriptions can be quite variable, can be extended by statements in the language <sup>[14]</sup>. As has been ranging from natural language to logical formalisms, but increased formality and regularity clearly facilitates machine understanding. The recognition of the key role that ontologies are likely to play in the future of the web has led to the extension of web markup languages in order to facilitate content description and the development of web based ontologies, e.g., XML Schema <sup>[1]</sup>.

### Web Ontology Language (OWL)

Web Ontology Language (OWL) defines more classes that let RDF authors define more of the meaning of their predicates within RDF <sup>[15]</sup>. Four classes of predicates defined by OWL include: owl: Symmetric Property, owl: Transitive Property, owl: Functional Property, and owl: Inverse Functional Property. (The OWL namespace is <http://www.w3.org/2002/07/owl#>.) Each of these classes is rdf: sub Class Of rdf: Property.

Applications can use these classes, by convention, to make inferences about data. You would use these classes in an ontology like this:

Defining amazon: price

amazon: price rdf: type owl: Functional Property

Because these classes are defined in the OWL ontology as being sub-classes of rdf: Property, applications can infer the following:

Defining amazon: price

amazon: price rdf: type rdf: Property

That's the same statement as earlier. So, when you use a sub-class in place of the 'parent' class, you're being strictly more informative. Anything the application knew before it still knows (if it has inference capabilities and knows the OWL ontology), and it knows more because the sub-class is more specific<sup>[15]</sup>.

OWL symmetric properties tell applications that the following inference is valid. If the application sees the statement S P O, and if P is typed as a symmetric property, then O P S is also true. For instance, we think of the has-friend relation are being symmetric. If you're my friend (ME HAS\_FRIEND YOU), I'm your friend (YOU HAS\_FRIEND ME).

OWL transitive properties work like this. If the application sees the statements X P Y and Y P Z, and if P is typed as a transitive property, then X P Z is also true. rdfs: sub Class Of is a transitive relation. If Mammal is a sub-class of Animal and Animal is a sub-class of Organism, then Mammal is a sub-class of Organism.

OWL functional and inverse-functional properties indicate how many times a property can be used for a given subject or object. A functional property is one that has at most one value for any particular subject. An example is the has Birthday relation between a person and his or her birthday. Everyone has just one birthday, so for any given subject (person), there can be just one object (birthday). But, the owns relation between an owner and owner is not functional. People can own more than one thing.

## XML

The Extensible Markup Language (XML) has become a standard language for data representation and exchange. XML<sup>[2]</sup> is a Standard, flexible syntax for data exchanging Regular, structured data. Database content of all kinds: Inventory, billing, orders etc. It has small typed values and irregular, unstructured text. It can consists of documents of all kinds: Transcripts, books, legal briefs etc.<sup>[16]</sup>. With the continuous growth in XML data sources, the ability to manage collections of XML documents and discover knowledge from them for decision support becomes increasingly important. Mining of XML documents significantly differs from structured data mining and text mining. XML allows the representation of semi-structured and hierarchal data containing not only the values of individual items but also the relationships between data items. Element tags and their nesting therein dictate the structure of an XML document. XML was designed to transport and store data. XML tags are not predefined. You must define your own tags XML documents<sup>[3]</sup> form a tree structure that starts at "the root" and branches to "the leaves".

## User Roles and Perspectives

The Administrator A database administrator, or DBA, would be the first user to interact with the system. The DBA is the entity responsible for maintaining a data set, and thus, for creating policies that regulate access to the data set. It is possible that the DBA does not have access to the queries that a user will make, and it is also possible that the DBA only has access to a data set's metadata<sup>[17]</sup>.

In order to create a policy, a DBA must have a list of the fields in a database, and in particular, the data types of those fields as URIs. It is up to the DBA to determine what kinds of policies they wish to implement. The DBA may be bound by local and national laws, by department practices, or by any number of other factors in creating policies. In all likelihood, the policies that the DBA needs to implement will be expressible in terms of the primitives.

## The User

The second major user of this system is someone who wishes to access the database. There are two possible modes of operation here. If the DBA or a system administrator has configured the policy assurance system, it is possible that the user will see no change, other than having some queries rejected for lack of compliance.

## The Auditor

The third user of the policy assurance system is the auditor. This is a person or entity charged with the responsibility of assuring that the policies that the DBA wrote are correct, and that the system achieves compliance. The auditor would be able to access the query history, the policies, and the reasoned outputs, and manually verify that things are working correctly. 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.

## 2. Literature Survey

The core technique of Semantic web mining is ontology [4]. In computer science, ontology represents a set of precisely defined terms about a specific domain and accepted by this domain’s community. Ontology is an explicit specification of a conceptualization. an ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology [5] together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins. Classes are the focus of most ontology [6]. Classes describe concepts in the domain. For example, a class of wines represents all wines. Specific wines are instances of this class. The Bordeaux wine in the glass in front of you while you read this document is an instance of the class of Bordeaux wines. A class can have subclasses that represent concepts that are more specific than the superclass. For example, we can divide the class of all wines into red, white, and rosé wines. Alternatively, we can divide a class of all wines into sparkling and non-sparkling wines. In practical terms, developing an ontology includes: defining classes in the ontology, arranging the classes in a taxonomic (subclass–superclass) hierarchy, fining slots and describing allowed values for these slots, filling in the values for slots for instances.

The Semantic Web [7] is designed to let users make explicit statements about any resource, and maintain that data themselves in an open and distributed manner. Several standards such as the Resource Description Framework (RDF) [8] and Web Ontology Language (OWL) [9] have been developed to realize the layer cake of the Semantic Web. World Wide Web has resulted in a revolution in the way information is transferred between computer applications. A common underpinning is especially important for the Semantic Web as it is envisioned to contain several languages, as in Tim Berners-Lee’s “layer cake” diagram first presented at XML 2000. The diagram depicts a Semantic Web

Architecture in which languages of increasing power are layered one on top of the other. Unfortunately, the relationships between adjacent layers are not specified, either with respect to syntax or semantics.

The Semantic Web is the name used to encapsulate the 2001 vision presented by Berners-Lee *et al.* [1] of a new Web as an information space usable by machines rather than humans. Web researchers realized that the rapid adoption of the Web and the associated information overload would necessitate alternative solutions and technologies where autonomous programs or machines assist humans to manage the information available on the Internet. Several definitions and descriptions of the Semantic Web were published since its inception as practitioners and researchers adopted the notion of a Web semantically enriched [10]. Some of these definitions are provided below in order to highlight the diversity thereof. The list is by no means exhaustive. a Web enriched with semantic meta-data that enables agents to execute complex information management tasks on behalf of its users a mechanism that contributes towards data, information and knowledge exchange and integration across communities and applications; a comprehensive architecture of meta-data language functionality that can be instantiated with different technology standards and specifications. The biggest challenge in the next several decades is how to effectively and efficiently dig out a machine-understandable and queryable information and knowledge layer, called Semantic Web, from unorganized, human-readable Web data.

In this paper [11], 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 research paper author Xiaohui Tao, Yuefeng Li *et al* [4]. 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 [7], 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, they 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. 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.

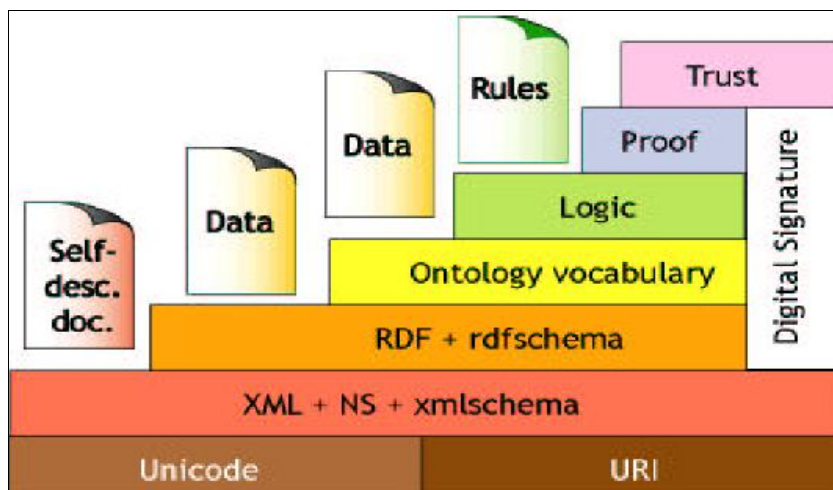
In this research paper author Xujuan Zhou, Sheng-Tang *et al.* [12], it is well known that taking the Web user profiles into account can improve the effectiveness of Web mining systems. However, due to the lively and complex nature of Web users, automatically acquiring worthwhile user profiles was found to be very stimulating. Ontology based user profile can possess more accurate user information. Their investigation emphasizes on acquiring search intentions information. They presents an 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.

**Proposed Technique**

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  $p_i$  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  $TP_i$  ).



**Fig 1:** Semantic web mining architecture

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.

**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 demonstrates 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.

**Experiment and Analysis Result**

In this experiment used Matlab to create Ontologies from web based images tag by using Matlab programming language for implementation. Used data from web based host database from 15 different type of websites.



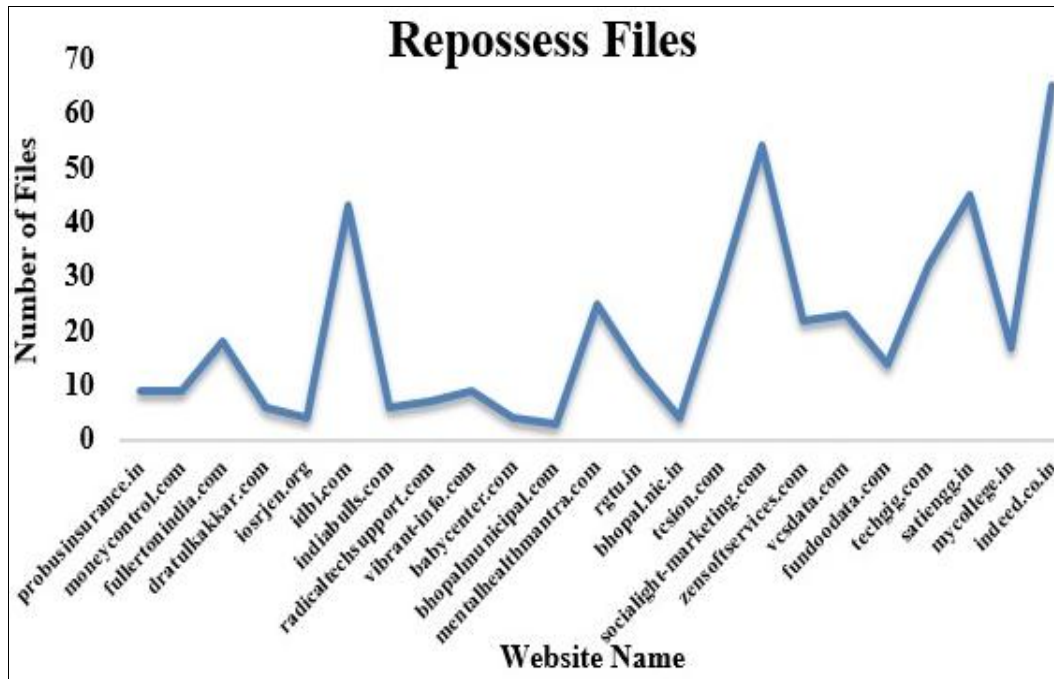


Fig 2: Repossess file from web domain

Table 1: Show Website domain wise ontology base experimental result

Web Domain	Current Page Rank	Time	Precision	Recall
probusinsurance.in	1	12	0.47	0.50
moneycontrol.com	6	11	0.73	0.91
fullertonindia.com	4	34	0.77	0.91
dratulkakkar.com	0	13	0.68	0.92
iosrjen.org	3	12	0.57	0.67
idbi.com	4	87	0.66	0.86
indiabulls.com	4	12	0.71	0.92
radicaltechsupport.com	0	11	0.61	0.64
vibrant-info.com	1	19	0.59	0.63
babycenter.com	6	12	0.52	0.83
bhopalmunicipal.com	4	11	0.52	0.73
mentalhealthmantra.com	0	56	0.64	0.93
rgtu.in	0	26	0.60	0.85
bhopal.nic.in	4	11	0.50	0.73
tcsion.com	5	34	0.39	0.62
socialight-marketing.com	0	67	0.68	0.91
zensoftservices.com	2	42	0.63	0.93
vcsdata.com	3	34	0.62	0.94
fundoodata.com	3	23	0.51	0.96
techgig.com	5	41	0.47	0.73
satiengg.in	6	53	0.60	0.83
mycollege.in	3	21	0.57	0.67
indeed.co.in	6	72	0.55	0.75

In experiment, built a computer ontology and input into the model. To test the developed method, Created ontology for

different web site from web host domain.

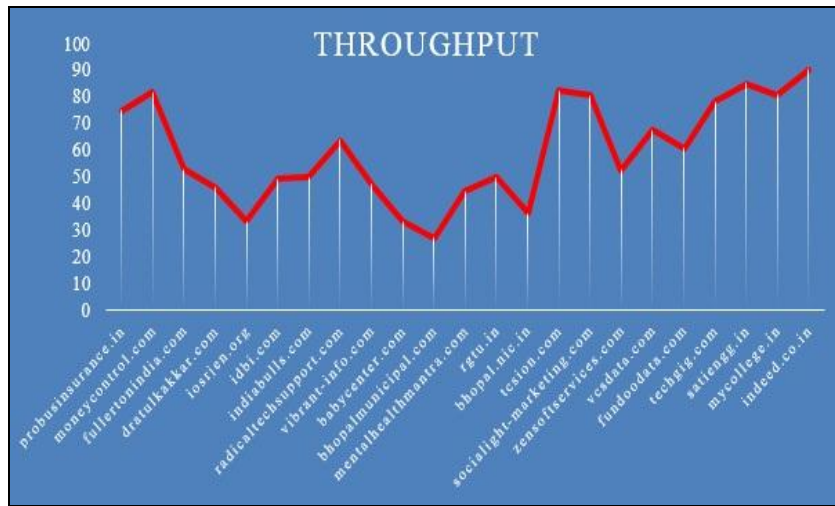


Fig 3: Throughput of Repossess file from web domain

Some concept pairs were chosen from each ontology and the relationships between them. The results are shown in Table 1.

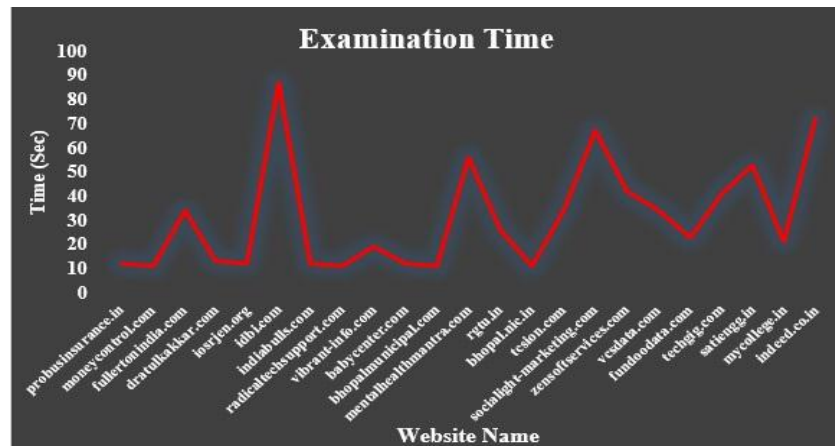


Fig 4: Examination Time for website

The process of this experiment using component under specified conditions, observing and recording the results, and making an evaluation of some aspect of the ontology and web semantic base result.

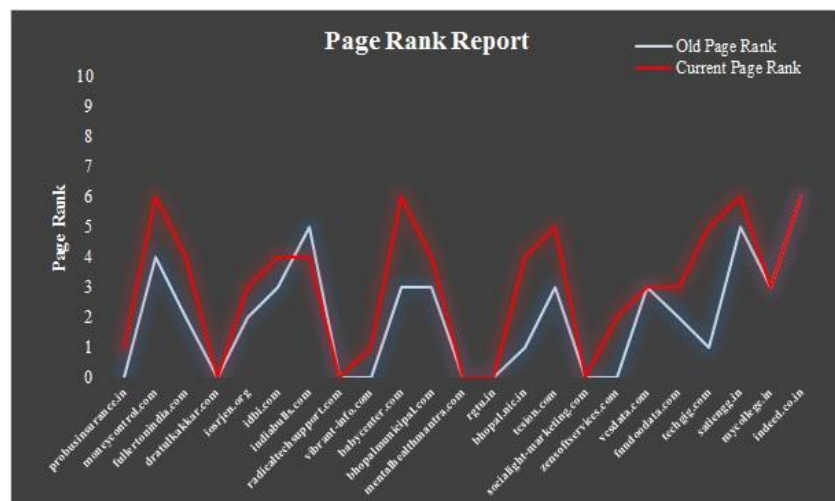


Fig 5: Compare results before and after repossess of website

In order to evaluate the throughput of the Ontology, we let one web domain have 100 files to three different tags. For our algorithm, search 10 file for 1 tags. Then we calculate throughput of searched file per second.

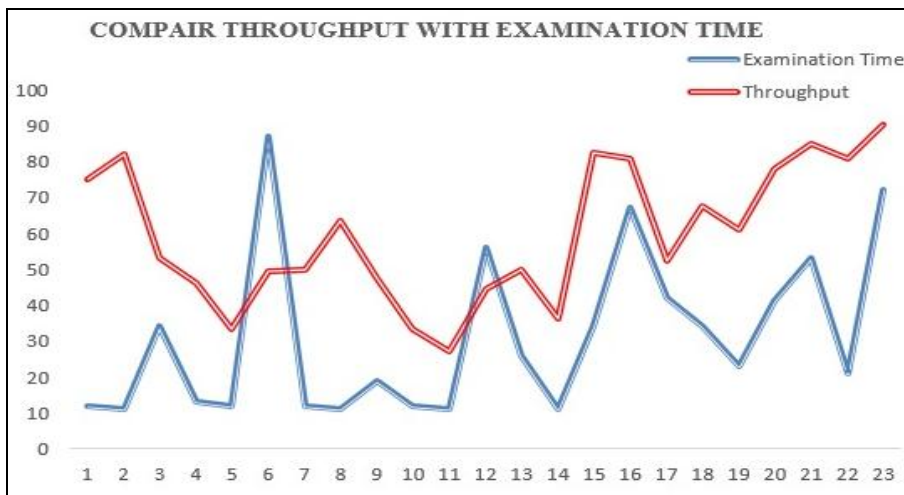


Fig 6: Compare throughput and examination time of website

The term testing can refer to many different but related activities, main purpose behind that is to evaluate some property of ontology based e-learning management system,

and eventually to increase our confidence in the precision of the product to our objectives.

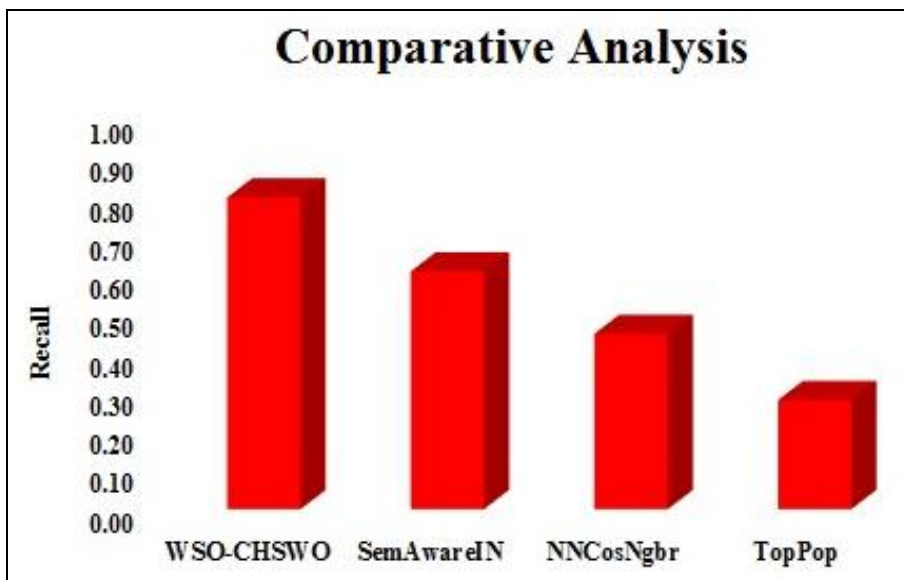


Fig 7: Comparative Analysis with other Algorithm

The test is conducted to assess certain aspects or property specified, and the aspect that sets the target for the test. As given in figure 7, recall of our algorithm is 0.80 compared to SemAwareIN 0.61, TopPop 0.28 and NNCosNgbr 0.45. Recall of SemAwareIN-Ex is compared with the recall of our proposed algorithm Web Search Optimizing on Cloud Host base Semantic Web using Ontology (WSO-CHSWO), is found to be 0.80, that is far better than that of other algorithm (SemAwareIN, TopPop and NNCosNgbr).

**Conclusion**

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. 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 an idealization.

## Reference

1. Stefan decker and sergey melnik, frank van harmelen, The semantic web: roles of xml and rdf, an horrocks,iee internet computing. 2000; 15(3):63-74.
2. Debajyoti Mukhopadhyay, Chandrima Chakrabarti, Sounak Chakravorty, A New Semantic Web Approach for Constructing, Searching and Modifying Ontology Dynamically, arXiv:1101.5763 ARXIV, Computer Science Information Retrieval, 2011.
3. Masaya Ito, Fumiko Harada, Hiromitsu Shimakawa, Extracting Ontology from Tagging to Web Pages in Similar User Group, International Journal of Advanced Computer Science. 2011; 1(2):58-64.
4. Xiaohui Tao, Yuefeng Li, Ning Zhong. A Personalized Ontology Model for Web Information Gathering, Published by the IEEE Computer Society, 2011.
5. Ozgun Akcay, Orhan Altan. An ontology-based visualization for mobile geo information services, Scientific Research and Essays, 2011; 6(4):993-1000.
6. Paolo Ciccarese, Marco Ocana, Leyla Jael Garcia Castro, Sudeshna Das, Tim Clark. An open annotation ontology for science on web 3.0, Ciccarese *et al.* Journal of Biomedical Semantics, 2011.
7. Nizar R. Mabroukeh, CI Ezeife. Ontology-based Web Recommendation from Tags, IEEE ICDE Workshops 2011.
8. Cheng Wang, Ying Liu, Lihang Jain, Peng Zhang. A Utility-based Web Content Sensitivity Mining Approach, IEEE, International Conference on Web Intelligent Agent Technology, 2008.
9. Andre Vizine, Landro N. De Castro, Richard Gudwin, An Evolutionary Algorithm to Optimize Web Document Retrieval, IEEE, KISMAS, 2005.
10. Association Rules Mining. A Recent Overview, Sotiris Kotsiantis, Dimitris Kanellopoulos GESTS International Transactions on Computer Science and Engineering. 2006; 32(1):71-82.
11. Awny Sayed, Amal Al Muqrishi. IBRI-CASONTO: Ontology-based semantic search engine, Egyptian Informatics Journal, Elsevier, 2017.
12. Xujuan Zhou, Sheng-Tang Wu, Yuefeng Li, Yue Xu, Raymond YK, Lau Peter D Bruza, *et al.* Utilizing Search Intent in Topic Ontology-based User Profile for Web Mining, IEEE/WIC/ACM International Conference, 2006.
13. [http://liris.cnrs.fr/alain.mille/enseignements/Ecole\\_Centrale/What%20is%20an%20ontology%20and%20why%20we%20need%20it.htm](http://liris.cnrs.fr/alain.mille/enseignements/Ecole_Centrale/What%20is%20an%20ontology%20and%20why%20we%20need%20it.htm).
14. [https://books.google.co.in/books?hl=en&lr=&id=qR\\_3BwAAQBAJ&oi=fnd&pg=PA1&dq=We+can+then+create+a+knowledge+base+by+defining+individual+instances+of+these+classes+filling+in+specific+slot+value+information+and+additional+slot+restrictions.+An+ontology+typically+consists+of+a+hierarchical+description+of+important+concepts+in+a+domai&ots=9qBt-Vj1g9&sig=yHYpDqhkVSvAB8NAL\\_qANx6VTx0](https://books.google.co.in/books?hl=en&lr=&id=qR_3BwAAQBAJ&oi=fnd&pg=PA1&dq=We+can+then+create+a+knowledge+base+by+defining+individual+instances+of+these+classes+filling+in+specific+slot+value+information+and+additional+slot+restrictions.+An+ontology+typically+consists+of+a+hierarchical+description+of+important+concepts+in+a+domai&ots=9qBt-Vj1g9&sig=yHYpDqhkVSvAB8NAL_qANx6VTx0).
15. [https://link.springer.com/chapter/10.1007/978-3-540-92673-3\\_4](https://link.springer.com/chapter/10.1007/978-3-540-92673-3_4).
16. <https://www.sciencedirect.com/science/article/pii/S0920548908001633>.
17. [https://dspace.utamu.ac.ug/bitstream/123456789/85/1/%5BRamakrishnan\\_R.,\\_Gehrke\\_J.%5D\\_Database\\_Management\\_S\(BookFi.org\).pdf](https://dspace.utamu.ac.ug/bitstream/123456789/85/1/%5BRamakrishnan_R.,_Gehrke_J.%5D_Database_Management_S(BookFi.org).pdf).