iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** III **Month of publication:** March 2018

DOI: http://doi.org/10.22214/ijraset.2018.3732

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Web Revisitation and Personalized Content Retrieval using Modular Ontology

Nisha A K[1], Anu T P[2]
[1]*PG Scholar at Jawaharlal College of Engineering and Technology, Palakkad, India*
[2]*Assistant professor at Jawaharlal College of Engineering and Technology, Palakkad, India*

Abstract: ***Getting back to antecedently viewed sites may be a common nonetheless uneasy task for users as a result of the massive volume of in-person accessed info on the online. This technique leverages human's natural recall method exploitation episodic and long-term memory cues to facilitate recall and presents a private net revisitation technique through context and content keywords. Underlying techniques for context and content reminiscences acquisition, storage, decay, associate degreed utilization for page re-finding provides the user with tailored and personalized services includes a preponderating importance once developing an interactive system. personalized retrieval data is a very important drawback moon-faced. It usually depends on the user's preference and demand. However, anticipating user's interest remains another challenge to be overcome. this can be associated with one hand, to the shortage of specific process modeling for information discovery, and on another hand, to fitting properly with what the user needs. a replacement technique supported user profile ontology designed through ontology modularization is proposed. The discovering of such preferences is achieved by a brand new methodology of content-based ontology that aims to require into thought each the hierarchical and also the non-hierarchical (semantic) relations put together with connected properties. Three modules are presented. first is context acquisition module, the second is employed to keeping specific user's preferences delineated by these rules and therefore the third module permits for structuring the expanding rule***
Keywords: Web Revisitation, Semantic web, Ontology, Semantic web rule language, Spreading activation technique

I. INTRODUCTION

Everyone is familiar with browsing and seek the help of a browser whenever we have doubt on any topics. The information retrieval (IR) can be used to provide documents to the end users to satisfy their need for information. Significant data is separated from the web with the assistance of a web index. A web crawler is the centerpiece of an internet searcher; it is really a program used to peruse the World Wide Web(WWW) in a programmed way. This process of browsing WWW is called as web crawling [1]. The gigantic information on the web makes the perusing muddled. Showing immaterial information while perusing deludes the client. so an insightful framework is expected to channel the important information from the web. This is one of the major challenges faced by all the information retrieval systems. Traditional information retrieval systems and automated information retrieval systems are the major divisions of IR systems. Traditional information systems include Vector space models, Boolean model, and probabilistic models. Automated information retrieval systems include data mining, web crawling, indexing and machine learning. Data mining is used everywhere in this world. It is the process of discovering the relevant patterns from large datasets called as big data. The main concept of data mining relays upon getting the useful data. Web mining is one of the major parts of data mining. That is what this paper mainly concentrating upon. Web mining can be partitioned into three.Web structure mining, web content mining and web utilization mining. Web structure mining, for the most part, manages the association among the website pages. While web content mining is used to extract the relevant information, that is what are all familiar with. Web usage mining is used to recognize interesting web usage patterns in the web and to understand the uses of web related applications.

Lots of research are carrying out to build the best intelligent system to support and to enhance the performance of the information retrieval systems. Another main aspect of browsing is personalized retrieval. To make the retrieval personalized concept of recommender systems, semantic webs, relevance feedbacks are mainly used.

In this paper, customized information based recommender framework is accomplished by improving the question with new data construed from the unequivocal client's inclinations inside a space of utilization. This paper, concentrating on construing new data that prompts upgrading the data recovery undertaking, and furthermore to managing a few issues identified with the semantic model. From this stance, we are attempting to manage three primary confinements identified with the semantic search and recovery.

Notwithstanding, with the developing size of the ontology, dealing with the development of the multifaceted nature must be taken into thought. Toward this objective, we use in this paper, the ontology modularization for tending to the size many-sided quality.

The particular ontology is made out of sub-modules and each one portrays sub-ontology and indicates a sub-some portion of application information.

The overview of this paper is as follows. The following area presents a foundation for the personalization in the data recovery framework. The section3 presents the writing survey by talking about the approach of data recovery framework, the cosmology based data recovery framework, and the Ontology designing. Section4 portrays the general proposed structure. Section5 is committed to the exploratory outcomes. At last, the paper conc1udes and talks about future works.

II. RELATED WORK

Recommender framework [2] is thought to be a vital instrument in this computerized world. It can be utilized as a part of online web-based social networking to discover companions of comparative wavelength and investigating new associations, and in web-based business to purchase the item as indicated by the tastes of the clients. The primary point of recommender frameworks is to discover the client interests. Two methodologies are utilized as a part of recommender frameworks in discovering client's interests. Cooperative separating and substance based sifting. Collaborative filtering uses the recommendations from other users who have same preferences of the current user. Based on the likes and dislikes of the users and from the previous search behaviors suggestions are given to the users. Collaborative filtering techniques are classified as a model-based method, item-based method, and user-based method. User-based techniques [3] calculates the similarity between the users using the rating given to particular products by the users with similar characteristics. Item-based collaborative filtering [4] takes the items with similar characteristics and computes the similarity with the target item. The model-based collaborative filtering [4] methods check the user preferences by creating models. The user behavior can be understood from the way in which the user responds to other items. Content-based filtering techniques [5] can be connected to the information given by the client certainly or expressly. The data is created in view of this information given by the client. The item proposals are done based on the client profile. Be that as it may, content-based separating strategies face a few impediments, for example, recovery of deficient information and in recovering interdependencies and complex relations. The semantic web is the next approach used in intelligent information retrieval systems. By adding meaning to the web semantic web can be created. The idea of semantic web was introduced by Tim- Berners-Lee.[6]Semantic web helps in making the web understandable to both users and machines. The main components of the semantic web are Uniform Resource Locator(URL) [7], eXtensible Markup Language(XML), Resource Description Framework, ontology, and SPARQL. Ontology simply means the nature of being of any element. It helps in defining an element in details by describing the features and the relations of the element with other elements. Ontologies are mainly applicable in certain domains. How much data can be provided regarding the elements of the domain determines the accuracy of the results for the query given by the users. The figure1 shows various steps of an ontology learning process. There will be a domain ontology to which we apply all the logical operations. Intelligent systems use ontology concepts[12]. The knowledge identification and representation are the main steps in any intelligent information systems. Ontology gives several methods to connect web pages with some background knowledge. It can be used to reduce the semantic gap between the documents and the query[13].

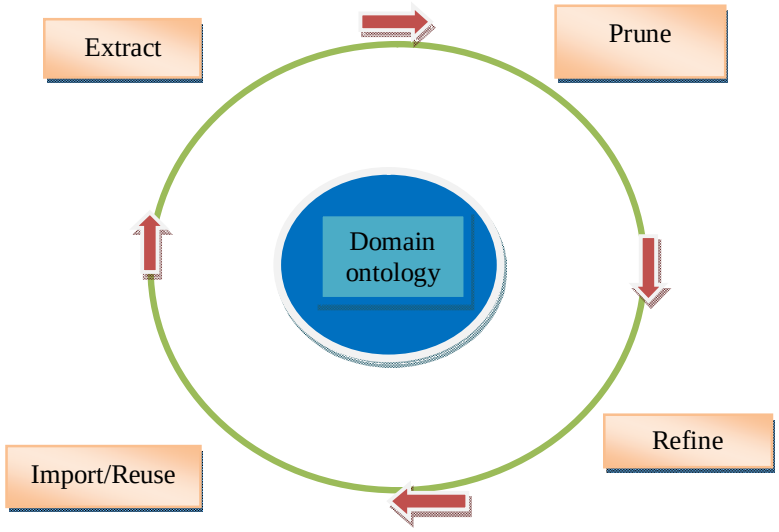


Figure 1: Ontology learning process

The ontology module as defined in [14] "is a reusable component of a larger or more complex ontology, which is self-contained but bears a definite association to the ontology modules, including the original ontology". Ideally, the ontology's modules aim to capture the meaning of the terms and concepts used. Videlicet, exactly the same inferring results should be obtained when answering queries according to these new modules. The idea is to define individual modules that can be assembled later into the enriched ontology with new concepts and relations within a domain of application. Utilizing the ontology measured quality permits thinking over the sub-modules and inducing new learning that is considered pertinent when executing the question procedure. In the following area, we show the second layer for thinking over the client profile, what's more, the space of use.

In philosophy, ontology can be defined as the study of existence. When coming to computer science ontology can be described as a mode to describe elements in concepts. Ontology converts the data to the machine-understandable form. Normally there use tools like protege2000[15],protégé 3.4 [16], SWOOP [17], Appolo[18] etc to construct ontologies. The domain ontology can be like medical, technical, electronics and so on. In order to build a domain ontology, a developer needs detailed and comprehensive knowledge regarding that particular domain. Protegue2000[3] stand alone ontology building open source tool with several plug-ins and extendable designs. Protégé 3.4[4] provides Graphical User Interface(GUI) and supports metamodeling.

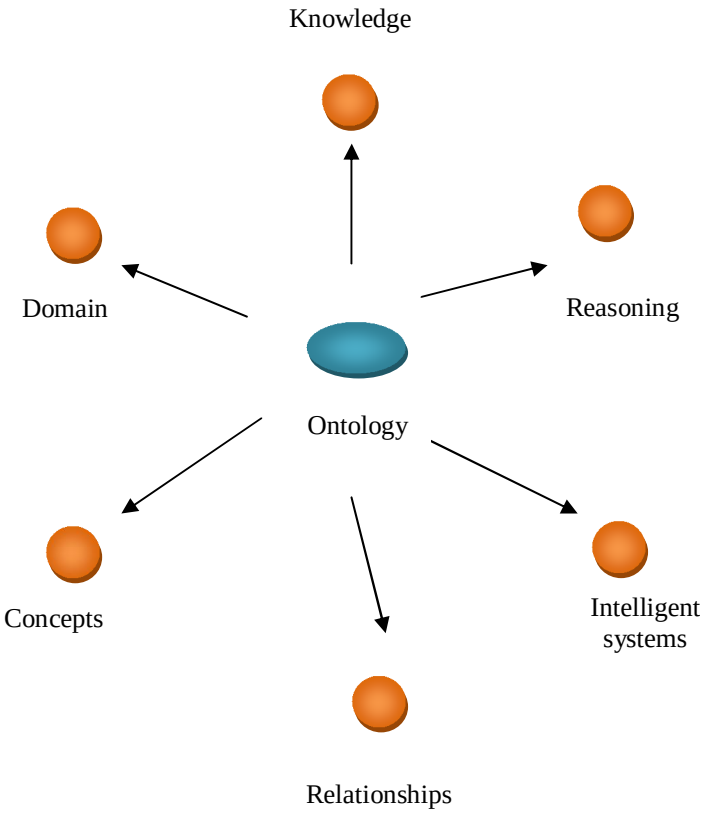


Figure 2: Ontology overview

SWOOP[5] is a Web Ontology Language editor tool. This provides a platform to compare, edited and merge several ontologies. Appolo[6] is a user-friendly tool which organizes ontologies in a hierarchical pattern.

In the paper web personalization[19] concept using ontology is presented. This paper is based on recommender system concepts which are under research. The web personalization uses the user's previous search details and it thus formulates an ontology based on the data obtained. For each user, there will be a personalized ontology. Which is reusable as per situations. When the user types a query it gets mapped to the user's ontology and the refinements are applied and the final result gets produced.

Ontology is one of the emerging concepts in the semantic web. It has several uses. The main advantage of ontology concepts is the way in which it describes each and every element and their relationships. Ontologies can be used in banking, e-commerce, the field of education, and also in retrieving personalized data in social media and in browsing, and in cloud computing[20] too.The user-based ontology creation helps the users to retrieve data as per their desire.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

III. METHODOLOGY

The foremost step is to get the data separated from the web. So that further extractions can be done. The user enters the query on the web and waits for the results. The application will retrieve the results as per the user's ontology. The user-based ontology helps to get personalized results. The hyperlinks will be extracted from the retrieved web pages. The process gets continued until all the hyperlinks get extracted. The further data mining and extraction is done on these hyperlinks. Final semantic matrix will be formulated based on these data extracted. And the final results based on ontology will be displayed to the user.

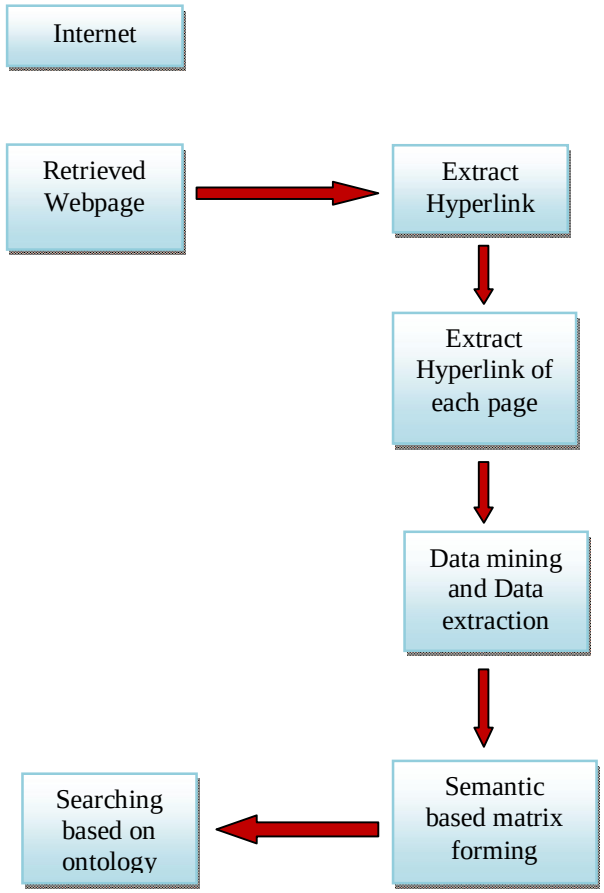


Figure 3: Extracting data

ontology module made out of the ideas (C) Instances (I) unequivocally proclaimed by the client. The motivation behind this module is to focus on the client's prerequisites while detailing the question procedure. ontology module made out of the ideas (C) Instances (I) unequivocally proclaimed by the client. The motivation behind this module is to focus on the client's prerequisites while detailing the question procedure.

Along with personalized retrieval, this system supports revisitation of the web pages[21] that the user previously visited. An individual web revisitation technique, that enables clients to return to their already engaged pages through access setting what's more, page content watchwords. Hidden procedures for setting and substance recollections' securing, stockpiling, and usage for website page review is examined. Dynamic tuning methodologies to tailor to person's retention quality and review propensities in light of significance criticism (e.g., weight inclination figuring, rot rate alteration, and so on.) are produced for execution change.

At the point when a client gets to a site page, which is of potential to be returned to later by the client (i.e., page get to time is over a limit), the setting Context acquisition module catches the present access setting (i.e., time, area, exercises surmised from the right now running PC programs) into a probabilistic setting tree. Then, the content extraction module plays out the unigram based extraction from the showed page fragments and acquires a rundown of probabilistic substance terms. The probabilities of obtained setting occasions and extricated content terms reflect how likely the client will allude to them as memory signs to return to the already engaged page.

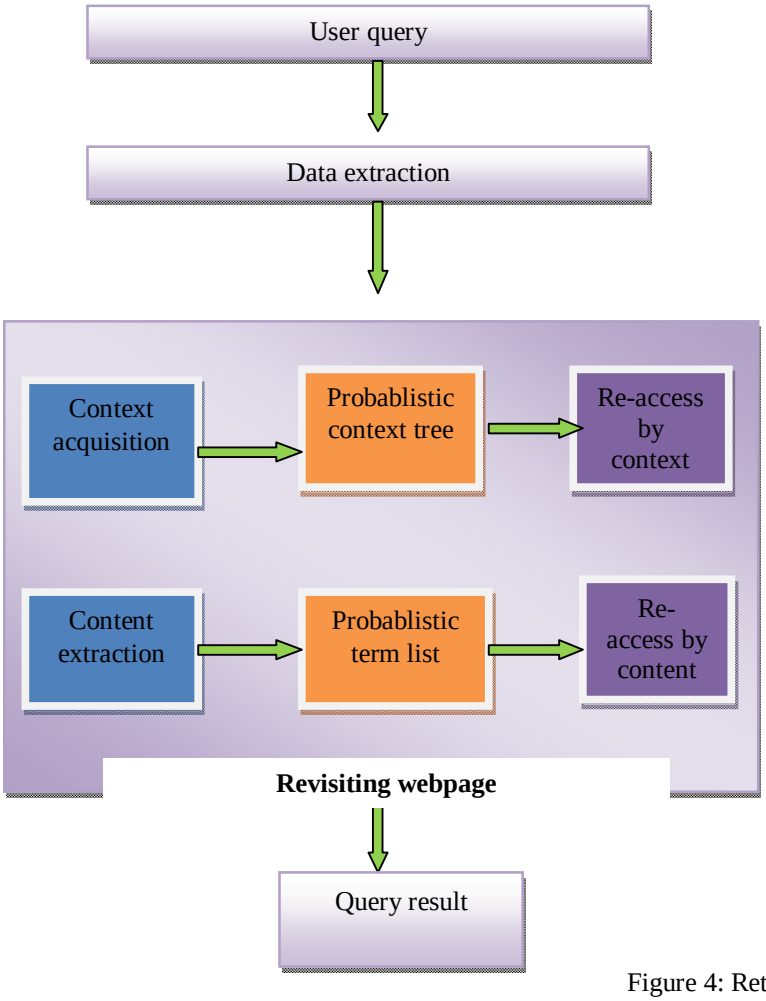


Figure 4: Retrieval process

Afterward, when a client solicitations to return to a formerly engaged page through setting as well as context keywords, the re-access by setting context extraction module and re-access by content extraction module look through the probabilistic tree and probabilistic term list, separately. The pertinence criticism component progressively tunes compelling parameters (counting recollections' rot rates, page perusing time edge, interleaved window measure edge, weight vectors in registering the affiliation and impression scores), which are basic to the development and administration of setting also, content recollections for individual web revisitation. Three sorts of client's access context are there, i.e., based on time, based on area, and simultaneous access can be caught. While get to time is determinate, get to the area can be gotten from the IP address of client's figuring gadget. Continuously monitor the change of user's focused program windows, which can be either a web page, a word file, or a chatting program window, etc., during user's interaction with the computer. Access context (i.e., time, location, and concurrent computer programming activities) is organized in a probabilistic context tree which adds generalized revisit queries of human user's understanding.

Rather than removing content terms from the full web page, here just considers the page portions appeared on the screen. There is numerous term weighting plans in the data recovery field. The most nonspecific one is to term frequency-inverse document frequency. For personalized web revisitation, merely counting the occurrence of a term in the presented page segment is not enough. Also, user's web page browsing behaviors can be obtained from users previous searches and the user's personalized ontology.

For a query Q, and retrieved web page W, Context keywords K_c, content keywords K_t, and the time t, matched webpage W_c, node v in tree V, mAs is the matching score then the algorithm will be as follows

ALGORITHM

Input: A query Q (W, K_c, K_t, t)

Output: Wm
For each hyperlink in the web page
Extract all the hyperlinks
Extract the content from the web page
Begin
Trees = getMatchContextTrees(W,Kc,t)
Lists = getMatchTermList(W,Kt,t)
Determine candidate matched page set Wc based on Trees and lists
For each w ϵ Wc Split w#tree into n smallest subtrees w#tree subi , (i=1,2,….n);
For i=1; i<=n ; i++
Determine matched nodes V subi of w#tree subi for each v ϵ, V subi doif v has a matched child nose in V subi then delete V from V subi else

$$mAS(Kc,v,t) = |Kc \cap v.title| / |v.title|$$

cRAnk(w#treesubi |(Kc,t) = $\prod v \epsilon subi \, . mAs(Kc, v, t)$
dRank(w#tree|Kc,t) =$\sum_{i=1}^{n} cRank(w\#tree\ subi\ |Qc, t)$

$$Rank(w|Q,t) = cRank\ (w\#tree|Kc,t)\ .\ dRank(w\#list|Kt,t)$$

Determine the matched page w with a highest ranking score
For each wϵWc do
if Rank (|w| , Q,t) < δ * Rank (wr |Q,t) then
Determine Wc by deleting w from Wc;
If W_c matched with user profile ontology display the results

IV. EXPERIMENTAL RESULTS

We at that point gave the accompanying directions to the clients before beginning the examination: 1) Each client is proposed to execute no less than one re-discovering undertaking every day. 2) For every re-discovering errand, four unique techniques are to be summoned. The execution succession is arbitrary. 3) When utilizing this system to do re-discovering, setting watchwords (from time, area, and action setting chains of command) and content catchphrases (from page's title and centered body content) could be input. With Memento, content watchwords (i.e., theme phrases removed from the page) and setting watchwords (i.e., point phrases extricated from the first furthermore, following pages) could be input. The clients could leave either setting or substance watchwords to purge if they would prefer not to. With History Searching, content keywords from page's title could be input. With Search Engine, content keywords from page's title, body text, and other descriptive information could be input. 4) At the end of a re-discovering assignment, the clients should record their search terms, the number of target pages they searched for, the positioning of the returned target page(s), and
the length of result page list.The Table1 below shows the comparison of finding rate between content and context terms in the web page revisitation and personalized retrieval system. This approach shows a similar result with the system proposed by Li Jin, Gangli Liu, et.al[21].

TABLE I
Comparing finding rate of content and context terms

Query Terms	Finding Rate
Content	0.8745
Context	0.9011
Content and Context	0.9388

At the point when a client does re-discovering, s/he, as a rule, has certain purposes as a top priority, such as setting up a task proposition, composing codes, and so forth. This framework endeavors to help clients to re-discover what they got to through past access time, area, and substance keywords. Past that, more client-driven setting factors (e.g., get to reason, ability, foundation, intrigue, and so forth.), and in addition social setting factors (e.g., outside occasions, encompassing individuals, and so forth.), could be construed from client's profile, motivation, and outside specialist organizations, what's more, limited with the got to pages. Along these lines, not just the client him/herself could profit by such rich relevant signs amid re-discovering process, yet in addition, other

clients with comparative access reason and foundation could share the more coordinated page get to Several studies are carried out to enhance the accuracy in the browsing results.

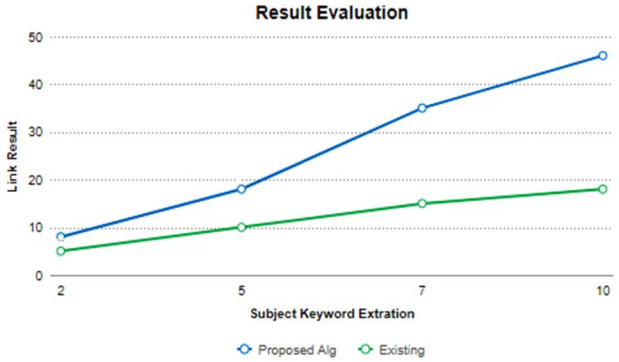


Figure 5: Subject keyword extraction comparison

As per the results obtained from the semantic similarity assessment a graph can be plotted based on the results obtained from [21] and our proposal. Figure 5 shows the comparison between the two methods.

Essayeh Aroua, Abed Mourad[22] in their paper used an approach based on ontology and ontology modularisation enhance the search results. In Table II there shows a comparison with the results shown in the concept of Essayeh Aroua, Abed Mourad[22] and our proposal.

TABLE II An extract of results of semantic measure comparison

No	Word pair		Essayeh Aura[22]	Proposed system
1	tiger	cat	0.93	0.82
2	tiger	tiger	1.0	1.0
3	book	Paper	0.88	0.76
4	autograph	Shore	0.29	0.16
5	autograph	Signature	0.55	0.82
6	automobile	Car	0.64	0.80
7	computer	Keyboard	0.98	0.55
8	plane	car	0.90	0.30
9	boy	Lad	0.66	0.82
10	boy	Sage	0.51	0.37
11	cost	Forest	0.36	0.36
12	Cost	shore	0.76	0.80
13	smart	student	0.10	0.08
14	smart	stupid	0.50	0.00
15	glass	tumbler	0.65	0.82
16	student	professor	0.88	0.40
17	Midday	noon	1.0	1.0
18	pet	dog	0.56	0.48
19	cushion	pillow	0.66	0.82
20	cord	string	0.68	0.82
21	John is very nice	Is John very nice?	0.97	0.88
22	Red alcoholic drink	A bottle of wine	0.58	0.30
23	Red alcoholic drink	Fresh orange juice	0.61	0.27
24	Red alcoholic drink	An English dictionary	0.00	0.028
25	I have a pen	Where is ink?	0.12	0.07
26	A glass of cider	A cup of apple juice	0.67	0.34
27	……..	……..	……..	…….

V. CONCLUSION

Drawing on the attributes of human mind memory in organizing and exploiting episodic events and semantic words in data review, this paper introduces an individual web revisitation method in light of setting and content keywords.

Customized recovery data is an important issue in the intelligent framework. It quality rally relies on the client's inclination and necessity. In this paper, here propose another way to deal with find user's interest. This later is executed as a modular ontology to diminish the larger than average and the many-sided quality of information. The objective is to find implicit interests and suspect new outcomes. This reasoning phase is mutually connected with a Spreading activation

technique. The ontology modularization helps in enhancing the efficiency of the system. The provision for finding the previously visited web pages also makes the system more user-friendly

We propose in future work, to incorporate our approach into a

context-aware framework and in this manner suggest new outcomes that rely on the given context

VI. ACKNOWLEDGMENT

This work has begun as a part of Master's Degree project. In this respect, the authors would like to thank the staffs of the college. Along with that lots of thanks to the authors who gave enlightenment to this survey through their papers.

REFERENCES

[1] Chandni Saini,, and Vinay Arora, "Information Retrieval in Web crawling a survey," Conference on Advances in Computing, Communications, and Informatics (ICACCI) , Jaipur, India, Sept. 21-24, 2016

[2] Farhin Mansur, Vibha Patel and Mihir Patel A Review on Recommender Systems, International Conference on Innovations in information Embedded and Communication Systems (ICIIECS), March 17-18 2017

[3] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommendation systems: principles, methods and evaluation" , Elsevier, 2015

[4] Vandana A. Patil, Lata ragha, "Comparing the performance of

collaborative filtering algorithms", IEEE, 2011.

[5] Joel pinho lucas, Saddys segrera, Maria N. moreno , "Making

use of associative classifiers in order to alleviate typical drawbacks in recommender systems", Elsevier, 2011

[6] Artem Chebotko, Literature Survey on the Semantic Web and Search,2005, TR-DB-032005-C

[7] Eliane M Aalouf and Maria Sokhn, Enhancing Business Process Modelling for E-Government Processes using Semantic Web Technologies and Linked Data Principles, challenges e-2014 Conference Proceedings Paul Cunningham and Miriam Cunningham (Eds)IIMC International Information Management Corporation,2014

[8] S.R Rashmi,R.Krishanan,"Domain Ontologies And Their Use In Building Intelligent Systems,A Comprehensive Survey",International Conference On Innovative Mechanisms For Industry Applications,2017

[9] Sudha Ramkumar,Poorna,"Ontology based semantic search:an introduction and a survey on current approaches",International conference on intelligent computing applications,2014

[10] N.F. Noy, R.W. Fergerson, and M.A. Musen, "The knowledge model of protege-2000: combining interoperability and flexibility", in: 12th International Conference in Knowledge Engineering and Knowledge Management (EKAW00), Lecture Notes in Artificial Intelligence, vol.1937, Springer, Berlin, 2000, pp. 17–32.

[11] http://protege.stanford.edu

[12] S.R Rashmi,R.Krishanan,"Domain Ontologies And Their Use In Building Intelligent Systems,A Comprehensive Survey",International Conference On Innovative Mechanisms For Industry Applications,2017

[13] Sudha Ramkumar,Poorna,"Ontology based semantic search:an introduction and a survey on current approaches",International conference on intelligent computing applications,2014

[14] Paul Doran, Valentina Tamma, and Luigi lannone, "Ontology ModuleExtraction for Ontology Reuse: An Ontology Engineering Perspective,"in 1 6th ACM International Conference on Information and KnowledgeManagement, 2007 , pp. p. 61 -7 0

[15] N.F. Noy, R.W. Fergerson, and M.A. Musen, "The knowledge model of protege-2000: combining interoperability and flexibility", in: 12th International Conference in Knowledge Engineering and Knowledge Management (EKAW00), Lecture Notes in Artificial Intelligence, vol.1937, Springer, Berlin, 2000, pp. 17–32.

[16] http://protege.stanford.edu

[17] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, Bernardo Cuenca Grau, and James A. Hendler, "Swoop: A Web Ontology Editing Browser". J. Web Sem. 4(2): 144-153 (2006

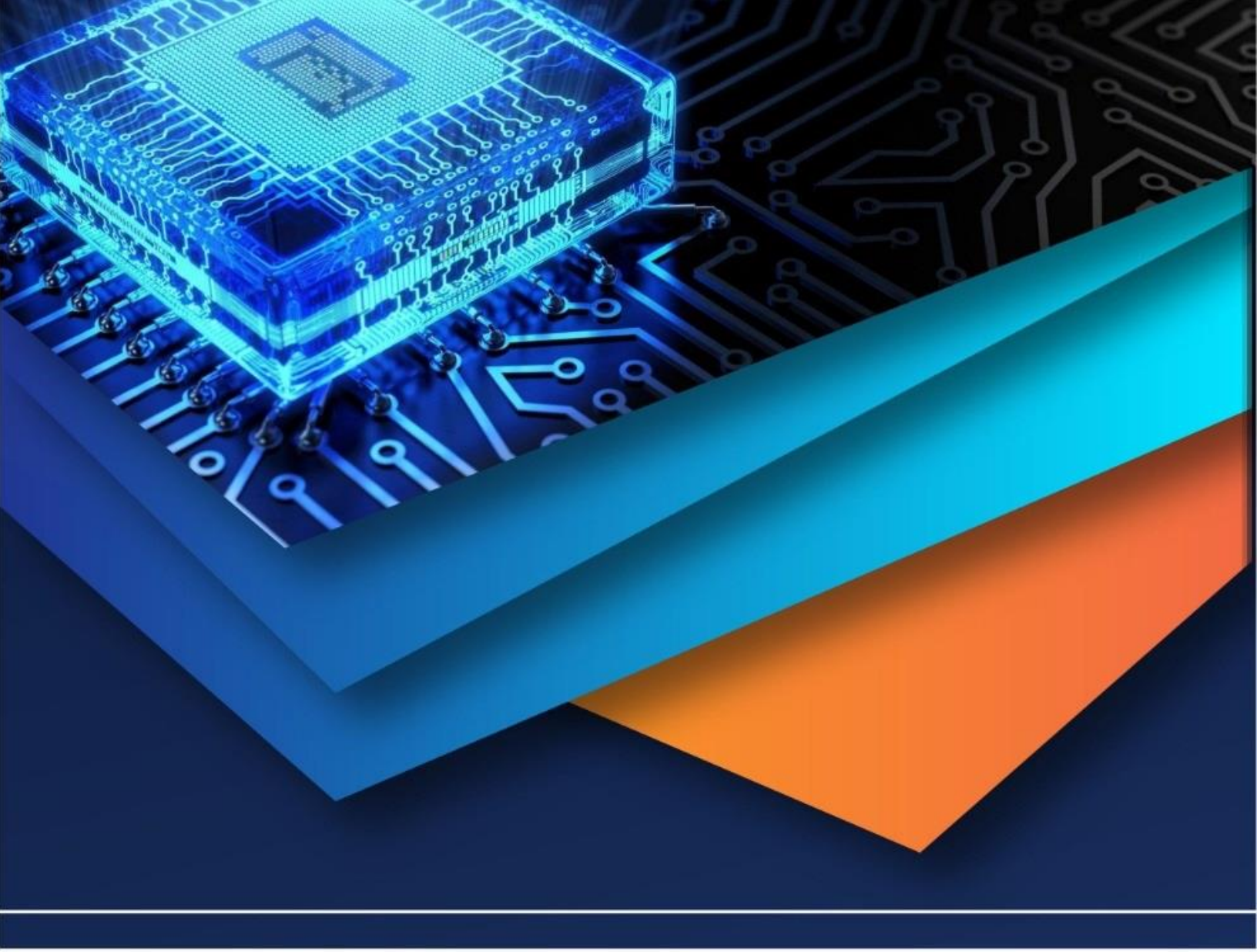
[18] http://apollo.open.ac.uk/index.html

[19] Tanya Bhattacharya1, Arunima Jaiswalet.al Web Usage Mining and Text Mining in the Environment of Web Personalization for Ontology Development of Recommender Systems, 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO),2016

[20] V.K.Reshma, B, Saravana Balaji, Cloud Service Publication and Discovery Using Ontology, International Journal of Scientific & Engineering Research Volume 3, Issue 5,2012

[21] Li Jin, Gangli Liu, Chaokun Wang and Ling Feng, Personal Web Revisitation by Context and Content Keywords with Relevance Feedback, IEEE Transactions on Knowledge and Data Engineering, DOI 10.1109/TKDE.2017.2672747,2017

[22] Essayeh Aroua, Abed Mourad, An Ontology-Based Framework For EnhancingPersonalized Content And Retrieval Information, 11th International Conference On Research Challenges In Information Science (RCIS) , 2017



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY