iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** III **Month of publication:** March 2015

DOI:

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Intelligent Web System by Integrating Domain and Web Usage Knowledge

Sangeetha B [1], Saranya A [2], Revathi K [3]
[1,2]UG Scholar, Computer Science and Engineering, Dhanalakshmi College of Engineering, Kanchipuram, India
[3]Associate Professor, Computer Science and Engineering, Dhanalakshmi College of Engineering, Kanchipuram,India

Abstract— The growth of the web has created a big challenge for directing the user to the Web pages in their areas of interest. This paper proposes a method to provide better web page recommendation through semantic- enhancement by integrating domain and web usage knowledge. Recommendations for Web pages were not introduced in the existing system whereas in proposal system it is being implemented and enhanced. Two new models are used to enhance the page recommendation, one is ontology based model and other is semantic network. Pre-Order Linked WAP-Tree Mining (PLWAP-Mine) algorithm is used to discover web usage knowledge resulting in frequently viewed term patterns. Another new model called conceptual prediction model is purposed to automatically generate the frequently viewed terms. The "new-page problem" is over-comed in this proposed system. It produces significantly higher performance than the advanced Web Usage Mining (WUM) method.

Index Terms—Web usage mining, site-page recommendation, domain ontology, semantic enhanced network, learning representation.

I. INTRODUCTION

Site page proposal has ended up progressively prominent, and is indicated as connections to related stories, related books, or most saw pages at sites. At the point when a client peruses a site, a grouping of went to Web-pages amid a session (the period from beginning, to existing the program by the client) can be created. This arrangement is composed into a Web session S = d1d2 ... dk, where di (i = [1 ... k]) is the page ID of the ith went by Web-page by the client. The destination of a Web-page recommender framework is to successfully foresee the Web-page or pages that will be gone to from a given Web-page of a site.There are various issues in building up a successful Web-page recommender framework, for example, how to viably gain from accessible recorded information and find valuable learning of the area and Web-page route pat- terns, how to model and utilize the found learning, and how to make powerful Web-page proposals in light of the found information.

A lot of exploration has been committed to determine these issues over the previous decade. It has been reported that the methodologies taking into account tree structures and probabilis- tic models can productively speak to Web access arrangements (WAS) . These methodologies gain from the preparation datasets to construct the move connects between Web-pages. By utilizing these methodologies, given the current went to Web-page (alluded to as a state) and k previ- ously went to pages (the past k expresses), the Web-page(s) that will be gone by in the following route step can be anticipated. The execution of these methodologies relies on upon the sizes of preparing datasets. The greater the preparation dataset size is, the higher the expectation exactness is. On the other hand, these methodologies make Web-page suggestions exclusively in light of the Web access groupings learnt from the Web utilization information. In this manner, the anticipated pages are restricted inside the found Web access game plans, i.e., if a customer is heading off to a Web-page that is not in the discovered Web access progression, then these philosophies can't offer any recommendations to this customer. We allude to this issue as "new-page issue" in this study.

This paper introduces a novel technique to give wager ter Web-page proposal in view of Web use and space information, which is upheld by three new learning representation models and a set of Web-page suggestion methodologies. The principal model is a metaphysics based model that speaks to the area information of a site. The development of this model is semi-computerized so that the advancement endeavors from engineers can be decreased. The second model is a semantic system that speaks to area learning, whose development can be completely computerized. This model can be effortlessly incorpo- evaluated into a Web-page suggestion process due to this completely robotized gimmick. The third model is a reasonable forecast model, which is a route net- work of space terms in light of the every now and again saw Web-pages and speaks to the incorporated Web use and area learning for supporting Web-page expectation. The development of this model can be completely mechanized. The proposal procedures make utilization of the space information and the expectation display through two of the three models to anticipate the next pages with probabili- ties for a given Web client in view of his or her current Web-

page route state. All things considered, this new system has mechanized the information base development and mitigated the new-page issue as said above. This strategy yields better execution contrasted and the current Web utilization based Web-page suggestion frameworks. This paper is structured as follows: Section 2 briefs the related work. Section 3 presents the first model, i.e. an ontology-based domain knowledge model. Section 4 describes the second model, i.e. a semantic net of domain terms. Section 5 presents the third model, i.e. a conceptual prediction model. For each of the models pre- sented in Sections 3-5, the corresponding queries that is used to retrieve semantic information from the knowledge models have been presented.

II. RELATED WORK

We by and large describe the investigation work recognized with Web Page proposition into going hand in hand with two classes.

A. Traditional Approaches that utilization Sequence Learning Models

In applying arrangement learning models to Web-page recommendation, affiliation rules and probabilistic mod- els have been generally utilized. A few models, for example, successive demonstrating, have demonstrated their noteworthy effec- tiveness in proposal era [3]. Keeping in mind the end goal to model the moves between diverse Web-pages in Web sessions, Markov models and tree-based structures are solid applicants [3], [12]–[15]. A few studies [16], [17] have demonstrated that tree-based calculations, especially Pre- Order Linked WAP-Tree Mining (PLWAP-Mine for short) [14], are exceptional in supporting Web-page recommen- dation, contrasted and other succession mining calculations. Besides, the incorporation of PLWAP-Mine and the higher-request Markov model [13] can fundamentally improve mining execution [18].

B. Semantic-Enhanced Approaches

The semantic-upgraded methodologies coordinate semantic data into Web-page suggestion models. By making utilization of the philosophy of sites, Web-page recom- mendation can be improved and enhanced fundamentally in the frameworks [19], [20]. In the frameworks, a space philosophy is regularly helpful for bunching records, grouping pages or seeking subjects. A space cosmology can be gotten by manual or programmed development approaches, for exam- ple, ontologies have been produced for separation learning courses [21], course content [22], customized e-learning [23], contracts [24], and programming [25]. Contingent upon the space of enthusiasm for the framework, we can reuse some exist- ing ontologies or fabricate another cosmology, and afterward coordinate it with Web mining. Case in point, philosophy ideas are utilized to semantically improve Web logs in a Web personalization framework. In this framework, a philosophy is constructed with the ideas extricated from the reports, so that the docu- ments can be grouped taking into account the similitude measure of the metaphysics ideas. At that point, utilization information is coordinated with the metaphysics so as to deliver semantically upgraded navigational examples. Therefore, the framework can make suggestions, contingent upon the data designs seman- tically coordinated with the delivered navigational examples. Liang Wei and Song Lei [19] utilize philosophy to speak to a site's space learning utilizing the ideas and signif- icant terms removed from reports. They create online suggestions by semantically coordinating and scanning for continuous pages found from the Web use min- ing procedure

III. DOMAIN ONTOLOGY OF A WEBSITE

In the connection of Web-page proposal, the info information is Web logs that record client sessions on a day by day premise. The client sessions incorporate data about clients' Web- page route exercises. Every Web-page has a title, which contains the magic words that grasp the semantics of the Web-page. In view of these certainties, we mean to find space learning from the titles of went to Web-pages at a web- website and speak to the found information in an area philosophy to help powerful Web-page proposal. A space philosophy is characterized as a theoretical model that points out the terms and connections between them expressly and formally, which thus speak to the area information for a particular space. The three principle segments are recorded as takes after [22]:

Domain terms (ideas),

Relationships between the terms (ideas), and

Features of the terms and connections.

Ontologies are frequently executed in a rationale based lan- guage, for example, OWL/RDF, to end up reasonable to programming specialists or programming frameworks. Thusly, cosmology based information representation permits offering and between changing semantic data among Web frameworks over the Internet. It likewise empowers the reuse of the space knowl- edge, and thinking the semantics of Web-pages from the current truths. Besides, ontological representation of found learning from distinctive sources can be effectively coordinated to help Web-page suggestion successfully.

A. Construction Of Domain Ontology

The structure for building up the region Ontology are of three stages.

1) *Step 1: Collect The Domain Terms:* To gather the terms, we will: (i) gather the Web log record from the Web server of the site for a time of time (no less than seven days), (ii) run a preprocessing unit to dissect the Web log document and produce a list of URLs of Web-pages that were gotten to by clients, (iii) run a product specialists to slither all the Web-pages in the URL rundown to concentrate the titles, and (iv) apply a calculation to concentrate terms from the recovered titles, i.e., single tokens are separated first by expelling prevent words from the titles, some single tokens are then consolidated into composite terms if these single terms regularly happen at the same time and there is never any token shows up between these tokens, and the staying single tokens will get to be single word terms. In view of the removed terms, we can sum them up to space con- cepts in Step 2.
2) *Step 2: Define the ideas:* It is feasible for some separated terms to have the same peculiarities, so it is better for them to be cases of an idea, as opposed to standalone ideas. In this step, the area ideas will be characterized for the given site in light of the removed terms. In this paper, we show the MS web- webpage as a sample. This site concentrates on the application programming, for example, MS Office, Windows Operating System, and Database. Thusly, the perceived region thoughts of this webpage are Support, Manufacture,News, product,Misc, and SemPage, where the con- cept SemPage insinuates the class of Web-pages, and exchange thoughts insinuate the general terms in the MS website.
3) *Step3:Define Taxonomic And Non-Taxonomic Connections:* As indicated by Uschold and Gruminger[26]. there are three conceivable ways to add to the taxonomic connection boats, for example, (5) a top-down advancement methodology begins from the most general ideas in the area and after that distinguishes the ensuing specialization of the general con-cepts, (6) a base up improvement process begins from the most particular ideas as the leave hubs in the con- cept various leveled structure/tree structure, then gatherings these most particular ideas into more general ideas, (17) a crossover advancement procedure is the mix of the top- down and base up metho0dologies. We recognize the center con- cepts in the area first and afterward sum up and practice them fittingly.

The non-taxomonic connections can be the connection boat sorts utilized as a part of a social database aside from the connections between a super-set and a sub-set, for example, referencing toward oneself, 1-M and M-N connections. In the MS web- website case, the fundamental sorts of non-taxonomic connections are recorded as beneath.

The "gives" connection depicts the M:N connection send between idea Manufacturer and ideas Product, Solution, Support, and News. The 'isPro- vided' connection is the converse of the "gives" connection.

The "has" connection portrays the M:N relationship between idea Application and ideas Product, Solution, Support, and News. The "isAppliedFor" rela- tion is the opposite of the "has" connection.

The "hasPage" connection portrays the M:N connection deliver between an idea, for example, Application and product.

IV. SEMANTIC NET OF A WEBSITE

The second model, i.e. new semantic net of a website, is generated which is a graph of concepts representing domain terms, Web-pages, and relations enclosing the collocations of domain terms, and the associations between domain terms and Web-pages. Initially, the domain terms are gathered from the Web-page titles based on the assumption that a generally composed Webpage ought to have an instructive title; then the relations among these terms are extracted from the accompanying two perspectives: (a) the collocations of terms which are dictated by the co-occurrence relations of terms in Web-page titles; and (b) the associations among terms and Web-pages. Additionally, the domain terms and co-occurrence relations are weighted to give a rough idea of how much these terms are connected with one another semantically. Taking into consideration the relations between the terms and Webpages, it can be inferred how nearly the Web-pages are semantically identified with one another. Utilizing this model, it is queried about the among terms and Web-pages, case in point, the related pages for a given page, the key terms for a given page, and the pages for offered terms, to translate the semantics of Web-pages to perform semantic enhanced Web-page proposals.This semantic network is alluded to as SemNetWeb for the accommodation in the explanation.

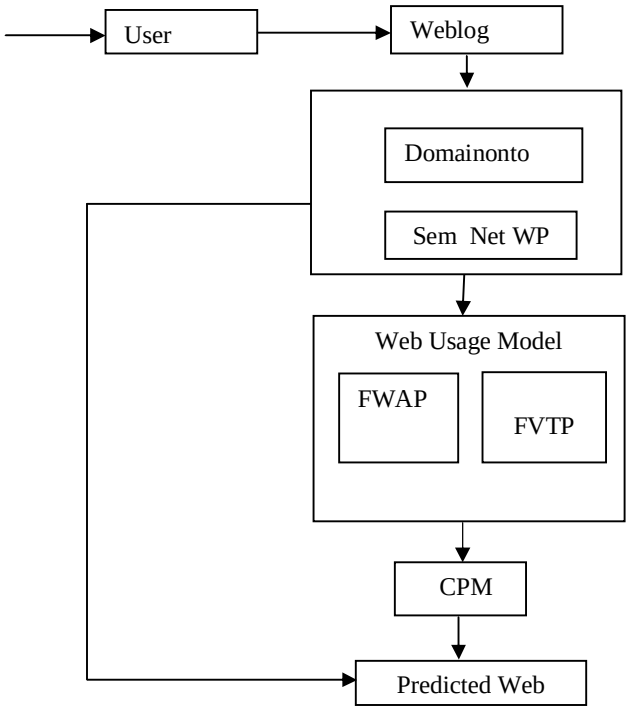
A. Process For Automatic Construction Of Semnetweb

To generate SemNetWeb, we implement the following procedure :

1) *Step 1:Collection Of Visited Web-Pages Titles:* To collect the titles, collect the Web log file from the Web server of the website for a certain length of time, run a pre-processing unit that examines the Web log file thus producing a list of URLs of Webpages that were approached by users, and run a software agent to crawl all the Web-pages in the list to extract the

titles.

2) *Step 2: Extraction Of Sequence Terms:* The algorithm used in the domain ontology construction is applied to extract the terms from the retrieved titles. The extracted terms are presented in the order as they appear in each title, namely they are collected as sequence terms.
3) *Step 3: Build The Semantic Net Semnetweb:* In SemNetWeb, each node represents a term in the extracted sequence terms and the order of sequence determines the 'from-Instance' and 'to-Instance' relations of a term between other terms.
4) *Step 4: Implemention of automatic construction of SemNetWeb:* The SemNetWeb is incorporated in RDF to enable reusability and liability of the domain term network by other parts of a Web-page recommendation system. The table shows the algorithm to automatically construct a SemNetWeb



FVTP (Frequently Viewed Term Patterns)
FWAP (Frequent Web Access Patterns)
CPM (Conceptual Prediction Model)

Figure 1. Architecture of Web Page Recommendation System

V. KNOWLEDGE REPRESENTATION MODEL OF A WEBSITE

To make efficient Web-page recommendations, usage of semantic Web usage knowledge is employed by integrating the domain knowledge model (DomainOntoWP) or the semantic network (SemNetWeb) with Web usage knowledge that can be developed from Web log files using a Web usage mining technique. To discover the Web usage knowledge, an advanced Web usage mining technique, namely PLWAPMine, is implemented which is in the form of frequent Web access patterns (FWAP), i.e patterns of frequently visited Web-pages.This FWAP is coordinated with DomainOntoWP or SemNetWeb forming a set of frequently viewed term patterns (FVTP). This is the semantic knowledge of Web usage of a website as shown in the fig. 1

A. Conceptual Prediction Model

The semantic Web usage knowledge is gained by a novel model called conceptual prediction model (CPM) which automatically generates a weighted semantic network of frequently viewed terms. This navigation system consists of domain terms depended on the as often as possible visited Webpages and characterizes the incorporated Web usage and domain knowledge for supporting Web-page prediction.The semantic network thus formed is named as WebNav. According to the Markov model, a

kind of model to describe the collection of navigation records, the CPM is developed as a self-contained and compact model. The basic elements are: state nodes, and the relations between the nodes. One node presents the current state, e.g. current viewed term, and may have some previous state nodes and some next state nodes.

The additional states used are: a start state, representing the first state of every term pattern; and a final state, representing the last state of every term pattern. The implementation of this model can be completely automated.

The probability of a transition is estimated by the ratio of the number of times the corresponding sequence of states was traversed and the number of times the anchor state occurred. Here we also consider into account the first-order and second-order transition probabilities.

B. Instinctive Construction Of Webnav Using Cpm

Given a set of frequently viewed term patterns, namely FVTP, the WebNav is generated by populating the CPM schema with FVTP. The CPM schema is designed using the formal ontology web language, RDF. An algorithm is also accomplished to perform this task. The transition probabilities is upgradable based on the first-order or second-order probability formula depending on the applied CPM's order. Thus, 1st or 2ndorder WebNav is obtained by using the 1st or 2nd-order CPM, respectively. For a given current Web-page or a combination of the current and previous Web-pages, the next Web-pages is recommended differently depending on which knowledge representation model and the order of CPM are used as mentioned earlier. These recommendation methods make utilization of the domain knowledge and the prediction model through two of the three models to forecast the enclosing pages with probabilities for a given Web client depended on his or her current Web-page navigation state. All things considered in this new system is fully automated. The knowledge base implementation has improved the new-page issue as specified previously. This technique yields better performance contrasted with the current Web usage based Web-page recommendation frameworks.

VI. PERFORMANCE

The execution of this proposal framework is evaluated by significant execution measurements specifically: Precision and Satisfaction. Here we consider the dataset of Microsoft site. Utilizing the MS dataset, we get the Web-page titles and ways. This site concentrates on the application programming, for example, MS Office, Windows Operating System, and Database. On the off chance that same peculiarities are imparted by the terms they are dealt with as occurrences of terms. The accuracy is valuable to quantify how plausible a client will get to one of the suggested Web-pages. Really, the following page got to by a client may not be the target page that client needs. In such cases, a client needs to get to a couple of halfway pages before arriving at the target page. Consequently, the fulfillment is important to give the exactness that the prescribed pages will be gotten to sooner rather than later

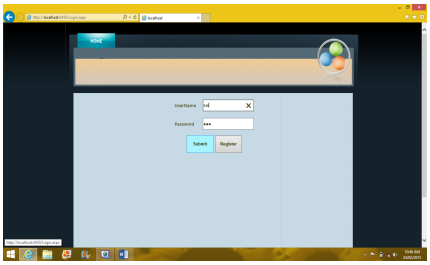


Fig.2 User Login page

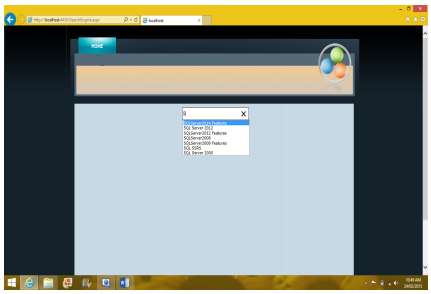


Fig.3 display of domain terms

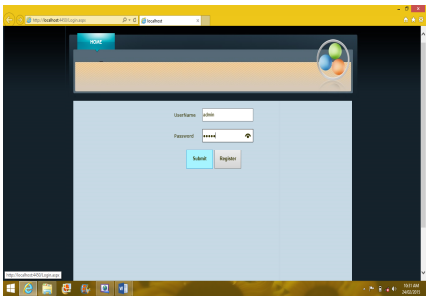


Fig.4 administrator login page

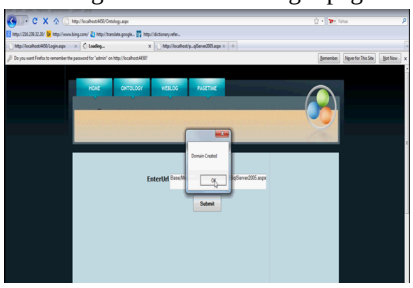


Fig.5 domain upload by admin

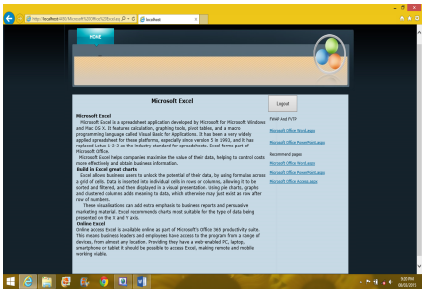


Fig.6 Recommendation provided for the existing use

VII. CONCLUSION AND FURTHER STUDY

This paper has Acquainted framework with an alternate offer better web-page proposal through semantic changes representation by three new learning models. Two new models have been proposed for repre- sentation of space learning of a site. One is a metaphysics based model which can be semi-consequently developed, specifically DomainOntoWP, and the other is a semantic system of Web-pages, which can be automati- cally built, in particular SemNetWP. A calculated forecast model is likewise proposed to coordinate the Web utilization and space information to structure a weighted semantic net- work of every now and again saw terms, in particular webNav. Various Web-page proposal methods have been proposed to anticipate next Web-page appeals of clients through questioning the information bases. The test results are guaranteeing and are demonstrative of the handiness of the proposed models.

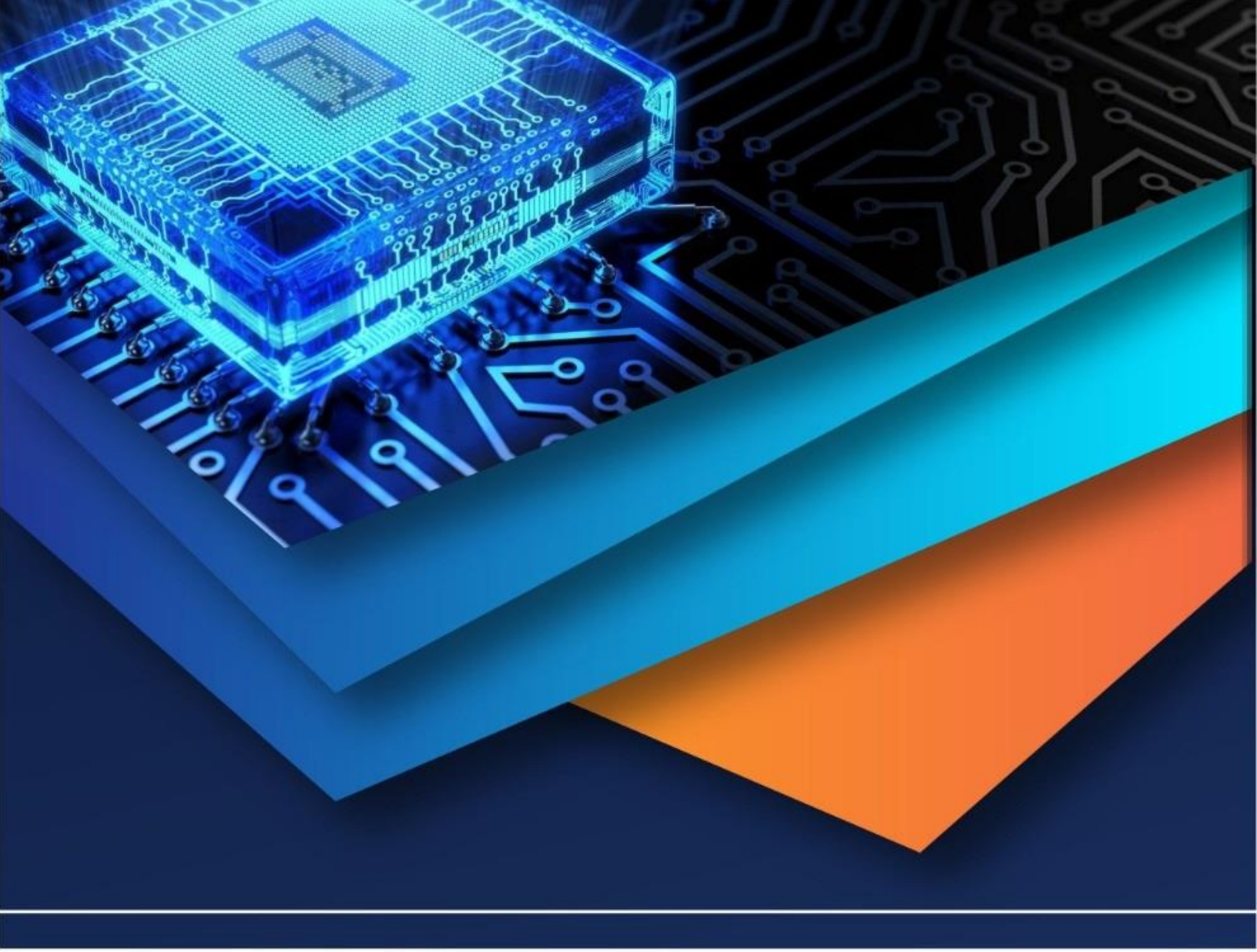
Contrasted and a standout amongst the most progressive Web use mining strategy, i.e. PLWAP-Mine, the proposed system can significantly upgrade the execution of Web-page rec- ommendation as far as accuracy and fulfillment. All the more importnatly, this system has the capacity assuage the "new-page" issue said in the presentation on the grounds that it taking into account the Web utilization information, as well as the semantics of Web-pages.

For the future work, a key data extraction algo- rithm will be created to pose as a viable rival with the term extrac- tion system in this work, and we will perform serious correlations with the exsiting semantic Web-page recom- mendation frameworks.usage mining method, i.e. PLWAP-Mine, the proposed method can substantially enhance the performance of Web-page rec- ommendation in terms of precision and satisfaction. More importnatly, this method is able to alleviate the "new-page" problem

mentioned in the introduction because it based on not only the Web usage knowledge, but also the semantics of Web-pages. For the future work, a key information extraction algo- rithm will be developed to compare with the term extrac- tion method in this work, and we will perform intense comparisons with the exsiting semantic Web-page recom- mendation systems.

REFERENCES

[1] Thi Thanh Ngugen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web usage and Domain Knowledge", in DataMining, published : Knowledge and Data Engineering,IEEE Transactions on(Volume:26, Issue:10),2014,pp.2574-2587.
[2] B. Liu, B. Mobasher, and O. Nasraoui, "Web usage mining," in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed. Berlin, Germany: Springer-Verlag, 2011, pp. 527–603.
[3] B. Mobasher, "Data mining for web personalization," in The Adaptive Web, vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 90–135.
[4] G. Stumme, A. Hotho, and B. Berendt, "Usage mining for and on the Semantic Web," in Data Mining: Next Generation Challenges and Future Directions. Menlo Park, CA, USA: AAAI/MIT Press, 2004, pp. 461–480.
[5] H. Dai and B. Mobasher, "Integrating semantic knowledge with web usage mining for personalization," in Web Mining: Applications and Techniques, A. Scime, Ed. Hershey, PA, USA: IGI Global, 2005, pp. 205–232.
[6] S. A. Rios and J. D. Velasquez, "Semantic Web usage mining by a concept-based approach for off-line web site enhancements," in Proc. WI-IAT'08, Sydney, NSW, Australia, pp. 234–241.
[7] S. Salin and P. Senkul, "Using semantic information for web usage mining based recommendation," in Proc. 24th ISCIS, Guzelyurt, Turkey, 2009, pp. 236–241.
[8] A. Bose, K. Beemanapalli, J. Srivastava, and S. Sahar, "Incorporating concept hierarchies into usage mining based recommendations," in Proc. 8th WebKDD, Philadelphia, PA, USA, 2006, pp. 110–126.
[9] N. R. Mabroukeh and C. I. Ezeife, "Semantic-rich Markov models for Web prefetching," in Proc. ICDMW, Miami, FL, USA, 2009, pp. 465–470.
[10] M. O'Mahony, N. Hurley, N. Kushmerick, and G. Silvestre, "Collaborative recommendation: A robustness analysis," ACM Trans. Internet Technol., vol. 4, no. 4, pp. 344–377, Nov. 2004.
[11] G. Stumme, A. Hotho, and B. Berendt, "Semantic Web mining: State of the art and future directions," J. Web Semant., vol. 4, no. 2, pp. 124–143, Jun. 2006.
[12] B. Zhou, S. C. Hui, and A. C. M. Fong, "CS-Mine: An efficient WAP-tree mining for Web access patterns," in Proc. Advanced Web Technologies and Applications. vol. 3007. Berlin, Germany, 2004, pp. 523–532.
[13] J. Borges and M. Levene, "Generating dynamic higher-order Markov models in Web usage mining," in Proc. PKDD, Porto, Portugal, 2005, pp. 34–45.
[14] C. I. Ezeife and Y. Lu, "Mining Web log sequential patterns with position coded pre-order linked WAP-tree," Data Min. Knowl. Disc., vol. 10, no. 1, pp. 5–38, 2005.
[15] B. Zhou, S. C. Hui, and A. C. M. Fong, "Efficient sequential access pattern mining for web recommendations," Int. J. Knowl.-Based Intell. Eng. Syst., vol. 10, no. 2, pp. 155–168, Mar. 2006.
[16] C. Ezeife and Y. Liu, "Fast incremental mining of Web sequential patterns with PLWAP tree," Data Min. Knowl. Disc., vol. 19, no. 3, pp. 376–416, 2009.
[17] T. T. S. Nguyen, H. Lu, T. P. Tran, and J. Lu, "Investigation of sequential pattern mining techniques for Web recommendation," Int. J. Inform. Decis. Sci., vol. 4, no. 4, pp. 293–312, 2012.
[18] S. T. T. Nguyen, "Efficient Web usage mining process for sequential patterns," in Proc. IIWAS, Kuala Lumpur, Malaysia, 2009, pp. 465–469.
[19] L. Wei and S. Lei, "Integrated recommender systems based on ontology and usage mining," in Active Media Technology, vol. 5820, J. Liu, J. Wu, Y. Yao, and T. Nishida, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 114–125.
[20] A. Loizou and S. Dasmahapatra, "Recommender systems for the semantic Web," in Proc. ECAI, Trento, Italy, 2006.
[21] D. Dzemydiene and L. Tankeleviciene, "On the development of domain ontology for distance learning course," in Proc. 20th EURO Mini Conf. Continuous Optimization Knowledge-Based Technologies, Neringa, Lithuania, 2008, pp. 474–479.
[22] S. Boyce and C. Pahl, "Developing domain ontologies for course content," Educ. Technol. Soc., vol. 10, no. 3, pp. 275–288, 2007.
[23] J. M. Gascuena, A. Fernandez-Caballero, and P. Gonzalez, "Domain ontology for personalized e-learning in educational systems," in Proc. 6th IEEE ICALT, Kerkrade, Netherlands, 2006, pp. 456–458.
[24] Y. Yalan, Z. Jinlong, and Y. Mi, "Ontology modeling for contract: Using OWL to express semantic relations," in Proc. EDOC'06, Hong Kong, China, pp. 409–412.
[25] D. Oberle, S. Grimm, and S. Staab, "An ontology for software," in Handbook on Ontologies, vol. 2, S. Staab and R. Studer, Eds. Berlin, Germany: Springer, 2009, pp. 383–402.
[26] M. Uschold and M. Gruninger, "Ontologies: Principles, methods and applications," Knowl. Eng. Rev., vol. 11, no. 2, pp. 93–36, Jun. 1996.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY