

Task Trail Based Enhanced User Experience in Customer Facing Website using Automatic Facet Construction Techniques.

A.R.Arunachalam, G.Michael, C.Anuradha, Sangeetha.S

Abstract: World Wide Web is the indispensable source for millions of millions user. The primary goal of the www is provide the most relevant, valid and right information for the end user to who is looking for the information. In this paper we conducted study on various term extraction method used recently by the researchers and we have made a comparative study of various term extraction techniques used in past. Finally we proposed a novel method of improving end user search experience by combining the task trail based user behavior and Automatic facet searching. Also this approach can be further taken forward by enhancing the facets search by referencing the well done Wikipedia for the descriptive based user search goal. The scope of the Automatic facets searching can be further enhanced for the product based product standard specification. Our scopes of work will predominately combining the Automatic facets with the user behavior from the task trail and adding product based standardization to improve the end search experience.

Index Terms: Facets searching, task trail, Automatic facets

I. INTRODUCTION

World Wide Web (WWW) is the one stop place for millions of million users for their day to day activity to knowledge gathering for modern sophisticated research and innovation. The primary goal of the www is provide the most relevant and valid information for the end user to who is looking for the information in the World Wide Web. WWW comprises the interlinked of complex connection of various websites, various pages, various domain and various protocols between them. The efficiency of the www is demonstrated only based on the fact how good you search engine delivers the valid results to the end user .Most of the case the input /Query supplied to the search is ambiguous in nature . So the search engine should be able provide the precise and much relevant information to the end user. There

Revised Manuscript Received on July 22, 2019.

A.R.Arunachalam, Department of Computer Science and Engineering, Bharath Institute of Higher education and research, Chennai, India

G.Michael, Department of Computer Science and Engineering, Bharath Institute of Higher education and research, Chennai, India

C.Anuradha, Department of Computer Science and Engineering, Bharath Institute of Higher education and research, Chennai, India

Sangeetha.S, Department of Computer Science and Engineering, Bharath Institute of Higher education and research, Chennai, India

are numerous ways that we can improve the performance of the search engine. The ways that can be included are faceted searching, Relevance ranking, Task trail based approaching, Feedback based approach, Semantic ontology based ranking. This paper provides and overview of some of the techniques that is used widely in recent years to optimize the search result based on the user behavior/goals. Also this paper provide the literature survey on the term extraction method used widely and how best this can align to my research area Faceted searching . [1],[3],[5]

II. RELATED WORK ON TERM EXTRACTION

ThiThanh Sang Nguyen [1]studied the new–page problem that current existing in web page recommendation system. They studied various technique used in current web recommendation system and finally found out that web usage mining and constructing domain ontology will effective provide better web page recommendation . Also they explain the various drawback and challenges in constructing semantic domain knowledge. Their proposed recommendation is based on web usage and domain knowledge. They had constructed the model in three ways. [2],[4],[6]

- (1) The ontology model that represents the domain knowledge of the website.
- (2) Semantic network that represent domain knowledge which is fully automatic
- (3) Conceptual prediction model, which is domain terms is constructed based on the web pages visited by the user.

A. Term extraction used in domain name retrieval.

The web server log file in the server is main source of input for the extraction of terms. The weblog files are preprocessed to get the valid websites URL that the user had visited the collected URL are subjected to crawl to capture the title of the page. Then preprocessing action like applying stop wordare applied to the given title and the terms are extracted from title. Now with the given terms the domain concepts are identified. The taxonomic relationships are derived from the extracted terms with relationship taxonomic approach. Such as consist of belongs to and so on. [7],[9],[11]

Zhen Liao [2] introduces the concept of tasks trail to understand the search behavior of user. The weblog is the input for this study and weblog is studied. The weblog capture the activities the user had done in the session (or)

list of task had performed in session. Also they have shown the advantage of task trail over the query trail and session trail. This dealt with determine the user satisfaction using task trail. This task trail helps user with better web page utilization. Task trail is comparable with query trail also much better than session trails. The query term from same task is consistent with query term from different task and finally the query suggestion based on task trail is better than query suggestion based on session trail and click through data. [8],[10] ,[12]

B. Term extraction used in task trail page

The task is defined as the set of activity the user performed within a session to define an atomic user information need. The primary source to perform this activity is the webpage data. The web logs collected from the servers are segmented based on the time threshold (i.e) session based. The segmented logs are analyzed based on the user search queries. The queries may sometime interleave with in the time threshold. [13], [15] ,[17]

The session based logs are further segmented based on the task according to semantic relationship between the user search queries. The user search query may not be continuation or consecutive within the task. The example could be like if the user search “pen” and further if he is searching for “ball pen” the records are displayed for the ball pen and if user click (or) further searches “blue ball pen” the intention of the user is to obtain the **result** for “blue ball pen” in that particular session . so the tasks identified is nothing but the search for “blue ball pen”, based on the relationship between the consecutive query the task can be also grouped as different task, same task. So the term/task clearly defines the more like the appearance/consists of/ resembles of the product the user is looking for. [14],[16], [18]

Weize Kong and James Allan [3] identifies the query facets groups and facets search terms from the user input search queries. The search queries are normally multifaceted (or) have different meanings. The facets are grouped based on their semantical related terms extracted from the search results. They proposed a supervised approach based on the graphical model to recognize query facets. Also they propose two algorithms for approximate inference on graphical model. Also they have defined some other approaches to find the query facet like QD miner, Topic modeling, and the QD miner is an unsupervised clustering approach and topic modeling is used to find high- quality semantic class from search result. [19],[21],[23]

C. Term extraction used in Query facets

The facet terms are extracted from the search result for the given ambiguous search query. These facet search term are part of semantic set, which is instances of semantic class. The facet terms can be extracted in many aspects .but this paper dealt with extraction from the search result. The top ‘k’ search results are considered for the facet term extraction and followed by grouping here the pattern like lexical and HTML pattern are used to collect the facet terms from the results.

Here after extracting the list from the top ranked results, first all the list terms are converted to lowercase and all alphanumeric characters are removed. Then the stop words and duplicate terms are removed. The term which are group less than two item (or) greater than 200 items are discarded and list is also noisy. Which cannot use as facet terms. So the develop graphical model to find the facet term, and how they are likely to group as query facet and capture the dependences between the two factor. Also they used semantic top modeling on the top search results to bring the facet term.

Zheng Lu, [4] the queries submitted to the search engine differ from users to user and intend to get result (or) the information needs are differs from each other the query can be always ambiguous in natural and it may differ in natural. If we analysis the user search query and goal. We can modify the search result to fulfill the user need.

The inference and analysis of user search goal can play vital role in bringing better user search experience and improve search engine relevance. This paper dealt with discovering the no of diverse user search goal for a query and depicting each goal with some keyword automatically. Here the clicked through logs (or) session logs is used to perform the analysis this paper is organized in three major contribution. (i)Feedback session clustering – the different user search goal for a query is concluded with feedback session clustering. Also it is found to be efficient than clustering search results (or) clicked URI s directly. (ii) a novel optimization method to combine the enriched URLs in a feedback session to form a pseudo – document, which can effected reflect the information need to a user.(iii)They proposed a new creation CAP to evaluate the user search goal performance based on restructuring web search results. So they will be able to determine the no of user search goal for query. [20],[22], [24]

D. Terms extraction used in CAP.

The click stream logs are used in this paper. A single session is containing only a query is considered for this analysis. The URLs that both clicked and unclicked URL which is obtained from user search query is considered the user feedback session contains unclicked URL before the last click used. [25],[27],[29]

Zhicheng Dou [5] addresses the problem of finding query facets from the query. The facet is a phrase (or) word that describes that denotes multiple items for the word. The query facets provide useful knowledge about query so that can be used to improve the user search experience. They also found out that import piece of information is usually presents in list style and repeated many times among. Top retrieved documents. They proposed a solution called QD miner to automatically mine query facets by extracting & grouping frequent list from free text, HTML tags and repeat region with in the top search result. Also they analyzed the problem of list duplication and find better query facets can be mined by modeling fine grouped similarity between lists and removed the duplicate lists the proposed models were unique website model, and the context similarity model to rank query facets. Also they made comparative study with

building facet hierarchies and found the proposed model is better unique. The experiment results also proved the context similarity model out performs the unique website model, which means that quality of query facet is improved by context similarity of the list. [26],[28],[30]

E. Term extraction technique applied.

The main input for this technique for this analysis and result is the end user search result for the given search term. The Terms (list) are extracted in mainly three ways free text pattern, HTML tag pattern, Repeated region pattern. (i) In the free text pattern all the text within the document considered and split into sentences. Then pattern matching method is used to obtain the list from the sentence. Similarly for the semi structured paragraphs the list are extracted from continuous lines that are comprised of two parts separated by a dash or a colon. The first parts of these lines are extracted as a list. (iii) In HTML tag pattern the list is formed from the HTML tag, which includes SELECT, UL, OL and Table .All the element with in the HTML tags are considered to be a list.(iii) In repeated region pattern, the repeated region in the web page is detected based on Visual –based DOM trees ,the repeat region is nothing but the set of similar items which has different values for the fields . The different values are basically the test that is used for the construction of list. [31],[33],[35]

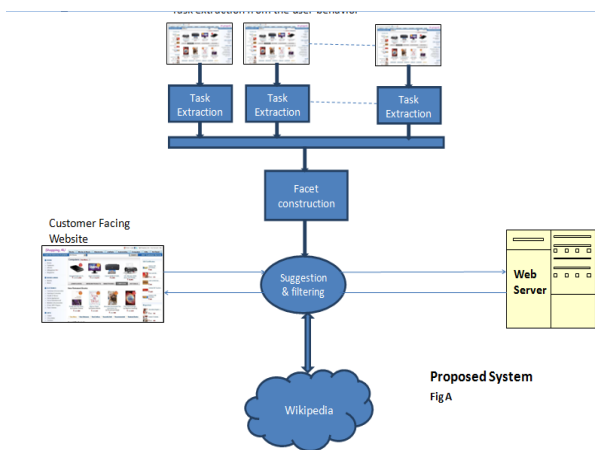


Fig :1 Overview Of Proposed System

III. PROPOSED CONCEPT

It is proved that task trail helps us to identify the user search goal and better compared to query trail. Facets techniques is use to bring ambiguous keyword to identify the user preference. The facets search will bring a better user experience and provide the drill down support to end user. It is also proved that the Automatic facets construction and mining with provide substantial support to user search experience. So the combined approach of task trail and facet technique will bring better search experience. Let us say the user search for baby doll Barbie which is nothing but a task trail. This task trail can be a faceted term for the next search irrespective of the same user or different user. Also Wikipedia provide a precise information for the product or any item the user is looking for so if the user is looking for

product and providing the descriptive search , we can refer the Wikipedia to bring the correct product or brand name which is a again the descriptive based faceting technique. Again most of the products come with standard dimension so the faceting of the dimension will also bring better faceting. [32],[34],[36]

IV. CONCLUSION

Enhanced user search experience is one of primary goal for the online user till now from past several years. The ontology based construction model and Facet searching are some of ongoing research area. In this paper we have seen some of the technique implemented in recent years to bring better search result to the End user. Also we have seen the comparative study of the different search terms techniques used by the researches to the goals. Also it found Task trail plays vital role to bring the result user search goal. The combination of the Task Trail and Faceted searching algorithms techniques will definitely improve the end user search experience. [37],[39],[41] [37],[39],[41]

REFERENCES

- [1] Kumarave A., Rangarajan K.,Algorithm for automaton specification for exploring dynamic labyrinths,Indian Journal of Science and Technology,V-6,I-SUPPL5,PP-4554-4559,Y-2013
- [2] P. Kavitha, S. Prabakaran “A Novel Hybrid Segmentation Method with Particle Swarm Optimization and Fuzzy C-Mean Based On Partitioning the Image for Detecting Lung Cancer” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019
- [3] Kumarave A., Meetei O.N.,An application of non-uniform cellular automata for efficient cryptography,2013 IEEE Conference on Information and Communication Technologies, ICT 2013,V-,I-,PP-1200-1205,Y-2013
- [4] Kumarave A., Rangarajan K.,Routing algorithm over semi-regular tessellations,2013 IEEE Conference on Information and Communication Technologies, ICT 2013,V-,I-,PP-1180-1184,Y-2013
- [5] P. Kavitha, S. Prabakaran “Designing a Feature Vector for Statistical Texture Analysis of Brain Tumor” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019
- [6] Dutta P., Kumaravel A.,A novel approach to trust based identification of leaders in social networks,Indian Journal of Science and Technology,V-9,I-10,PP--,Y-2016
- [7] Kumaravel A., Dutta P.,Application of Pca for context selection for collaborative filtering,Middle - East Journal of Scientific Research,V-20,I-1,PP-88-93,Y-2014
- [8] Kumaravel A., Rangarajan K.,Constructing an automaton for exploring dynamic labyrinths,2012 International Conference on Radar, Communication and Computing, ICRCC 2012,V-,I-,PP-161-165,Y-2012
- [9] P. Kavitha, S. Prabakaran “Adaptive Bilateral Filter for Multi-Resolution in Brain Tumor Recognition” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019
- [10] Kumaravel A.,Comparison of two multi-classification approaches for detecting network attacks,World Applied Sciences Journal,V-27,I-11,PP-1461-1465,Y-2013
- [11] Tariq J., Kumaravel A.,Construction of cellular automata over hexagonal and triangular tessellations for path planning of multi-robots,2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016,V-,I-,PP--,Y-2017
- [12] Sudha M., Kumaravel A.,Analysis and measurement of wave guides using poisson method,Indonesian Journal of Electrical Engineering and Computer Science,V-8,I-2,PP-546-548,Y-2017
- [13] Ayyappan G., Nalini C., Kumaravel A.,Various approaches of knowledge transfer in academic social network,International Journal of Engineering and Technology,V-,I-,PP-2791-2794,Y-2017
- [14] Kaliyamurthi, K.P., Sivaraman, K., Ramesh, S. Imposing patient data privacy in wireless medical sensor networks through homomorphic cryptosystems



- 2016, Journal of Chemical and Pharmaceutical Sciences 9 2.
- [15] Kaliyamurthi, K.P., Balasubramanian, P.C. An approach to multi secure to historical malformed documents using integer ripple transfiguration 2016 Journal of Chemical and Pharmaceutical Sciences 9 2.
- [16] A.Sangeetha,C.Nalini,"Semantic Ranking based on keywords extractions in the web", International Journal of Engineering & Technology, 7 (2.6) (2018) 290-292
- [17] S.V.GayathiriDevi,C.Nalini,N.Kumar,"An efficient software verification using multi-layered software verification tool "International Journal of Engineering & Technology, 7(2.21)2018 454-457
- [18] C.Nalini,ShwtambariKharabe,"A Comparative Study On Different Techniques Used For Finger – Vein Authentication", International Journal Of Pure And Applied Mathematics, Volume 116 No. 8 2017, 327-333, Issn: 1314-3395
- [19] M.S. Vivekanandan and Dr. C. Rajabhushanam, "Enabling Privacy Protection and Content Assurance in Geo-Social Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 49-55, April 2018.
- [20] Dr. C. Rajabhushanam, V. Karthik, and G. Vivek, "Elasticity in Cloud Computing", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 104-111, April 2018.
- [21] K. Rangaswamy and Dr. C. Rajabhushanam, "CCN-Based Congestion Control Mechanism In Dynamic Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 117-119, April 2018.
- [22] Kavitha, R., Nedunchelian, R., "Domain-specific Search engine optimization using healthcare ontology and a neural network backpropagation approach", 2017, Research Journal of Biotechnology, Special Issue 2:157-166
- [23] Kavitha, G., Kavitha, R., "An analysis to improve throughput of high-power hubs in mobile ad hoc network" , 2016, Journal of Chemical and Pharmaceutical Sciences, Vol-9, Issue-2: 361-363
- [24] Kavitha, G., Kavitha, R., "Dipping interference to supplement throughput in MANET" , 2016, Journal of Chemical and Pharmaceutical Sciences, Vol-9, Issue-2: 357-360
- [25] Michael, G., Chandrasekar, A.,"Leader election based malicious detection and response system in MANET using mechanism design approach", Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
- [26] Michael, G., Chandrasekar, A.,"Modeling of detection of camouflaging worm using epidemic dynamic model and power spectral density", Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
- [27] Pothumani, S., Sriram, M., Sridhar, J., Arul Selvan, G., Secure mobile agents communication on intranet,Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg No S32-S35, 2016
- [28] Pothumani, S., Sriram, M., Sridhar , Various schemes for database encryption-a survey, Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg NoS103-S106, 2016
- [29] Pothumani, S., Sriram, M., Sridhar, A novel economic framework for cloud and grid computing, Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg No S29-S31, 2016
- [30] Priya, N., Sridhar, J., Sriram, M. "Ecommerce Transaction Security Challenges and Prevention Methods- New Approach" 2016 ,Journal of Chemical and Pharmaceutical Sciences, JCPS Volume 9 Issue 3.page no:S66-S68 .
- [31] Priya, N.,Sridhar,J.,Sriram, M."Vehicular cloud computing security issues and solutions" Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016
- [32] Priya, N., Sridhar, J., Sriram, M. "Mobile large data storage security in cloud computing environment-a new approach" JCPS Volume 9 Issue 2. April - June 2016
- [33] Anuradha.C, Khanna.V, "Improving network performance and security in WSN using decentralized hypothesis testing "Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
- [34] Anuradha.C, Khanna.V, "A novel gsm based control for e-devices" Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
- [35] Anuradha.C, Khanna.V, "Secured privacy preserving sharing and data integration in mobile web environments " Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
- [36] Sundarraj, B., Kaliyamurthi, K.P. Social network analysis for decisive the ultimate classification from the ensemble to boost accuracy rates 2016 International Journal of Pharmacy and Technology 8
- [37] Sundarraj, B., Kaliyamurthi, K.P. A content-based spam filtering approach victimisation artificial neural networks 2016 International Journal of Pharmacy and Technology 8 3.
- [38] Sundarraj, B., Kaliyamurthi, K.P. Remote sensing imaging for satellite image segmentation 2016 International Journal of Pharmacy and Technology 8 3.
- [39] Sivaraman, K., Senthil, M. Intuitive driver proxy control using artificial intelligence 2016 International Journal of Pharmacy and Technology 8 4.
- [40] Sivaraman, K., Kaliyamurthi, K.P. Cloud computing in mobile technology 2016 Journal of Chemical and Pharmaceutical Sciences 9 2.
- [41] Sivaraman, K., Khanna, V. Implementation of an extension for browser to detect vulnerable elements on web pages and avoid click jacking 2016 Journal of Chemical and Pharmaceutical Sciences 9 2.

AUTHORS PROFILE



A.R. Arunachalam, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India



G. Michael, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India



C. Anuradha, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India



Sangeetha.S, Assistant Professor, Department of Computer Science & Engineering, Bharath Institute of Higher Education and Research, Chennai, India