ijecs open access www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 1 January 2015, Page No. 9836-9840

A Survey Paper on "Effective Navigation in Website

Payal R. Potarkar[1], Mayuresh A. Hedao[2], Prasad M. Dandhare[3] Samiksha D. Nakade[4],Pradnya B.Kamble[5],Vrushali S. Bhusari[6]

[1]Payal Potarkar, Bapurao Deshmukh College of Engineering Sewagram, India
Potarkarpayal@gmail.com
[2] Mayuresh A. Hedao, Bapurao Deshmukh College of Engineering Sewagram, India
Mayuresh24567@gmail.com
[3]Prasad M. Dandhare, Bapurao Deshmukh College of Engineering Sewagram, India
[4] Samiksha Nakade, Bapurao Deshmukh College of Engineering Sewagram, India
Sdnakade@gmail.com
[5] Pradnya Kamble, Bapurao Deshmukh College of Engineering Sewagram, India
Pradnya.kamble02@gmail.com
[6]Vrushali Bhusari, Bapurao Deshmukh College of Engineering Sewagram, India

Abstract: ***Well-structured website is not been design yet to facilitate effective user navigation.It has long been a challenge for developers. Mathematical model and two evaluation metrics using real data sets are to be discusses here. Here we also discusses the previous designed models and algorithms which are not so enough for effective user navigation. Developers design the website according to user's behavior so that the most disoriented user can find their path easily.we are not reorganizing website,just by making substantial changes we are making website more useful. Our goal is to provide better navigation to user by proposed model.***

Index Terms*:* Website Design, Link Analysis, User Navigation, Mathematical Model, Web Mining.

1. INTRODUCTION

Data mining plays very important role in extraction of information from large databases. As World Wide Web is every where and millions of people are using internet, it becomes very important to extract the exact information from the website. Data mining has various applications in everyday life. So it is necessary to study data mining along with its techniques. Till now various techniques are proposed but they have some limitations. So we have to either overcome these limitations or propose a new model which can help the effective user navigation.

Millions of users use the website as it has the large source of information. We can extract the any type of information through the website. Website also contains many links and pages within it. Users for World Wide Web increases day by day, so it is necessary to make web fast and easy to use. For this we need to know content of website, which are the users of website and structure of website.

Web graphics designs, interface design, authoring, including standardized code and, user experience design, search engine optimization and proprietary software are the different areas of web design. In a team individual work in covering different aspects of design process. Developers only organize pages on their own judgments. We can directly forward the user towards their destination. Generally as user search for any information, due to wrong labels to links user did not get the correct path. As a result user get back tracked to the previous links which wastes the time and user gets irritated by this and log out from the site. Hence it is very necessary to design well structured website with minimal changes so that the most disoriented user can get his path over a large warehouse.

Previous study includes extraction of templates from WebPages, mining informative structure of news website, finding relevant pages of a given page, and understanding web structure. In this paper our work is to examine how to improve website navigability by detecting user entered data. It can be classified into two categories: first personalization and second transformation.

Data sets are the important part of website. We need to understand its working in websites. We can perform some

experiments on the data sets of the website so that we can significantly improve the site structure with the minimum changes. We can't reorganize the complete website as it can change the location of familiar items, the new website can disorient the user. Our model is for the static contents of websites or not purely dynamic pages. Examples universities, tourist attractions, hospitals and sports organizations.

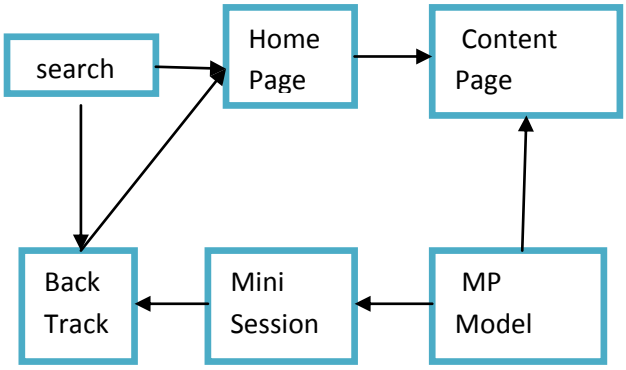

Figure1: Block diagram of searching of desired webpage using MP model

The Mathematical programming (MP) model which facilitates user navigation on website with minimal changes to the current structure of website is proposed here. MP model successfully accomplish the task and also generates the optimal solution very fast. We also model the out-degree as a cost term in objective function instead of as hard constraints. It allows a page to have more links to website if cost is reasonable. It also minimize information overload to user and we propose two evaluation metrics to confirm the validity of our model.

Organizing of the paper in this as section II describes literature survey, Section III provides existing work, Section IV provides proposed work, Section V concludes the paper.

1. LITERATURE SURVEY

The purpose of this review is to compare different models and algorithm. Our main focus is on facilitating effective user navigation through website structure improvement.

- Data has been generated in large volumes today. Extracting such large data is very difficult. Many users are using same source of extracting data. Information visualization and visual data mining to deal with large data from large warehouse of data are used [5]. It deals with one dimensional, two dimensional, multi dimensional data, hierarchies and graphs, algorithm and software. Visual data exploration involved the user in data mining processes. Numbers of visualization techniques were discussed here. Classification of information visualization and visual data mining techniques based on data type to be visualized, the interaction, the visualization technique and distortion technique. Paying by credit card or using telephone were recorded in computers, hence multidimensional data was stored. If we know very little about data and we don't know exploration target then visual data exploration is used.

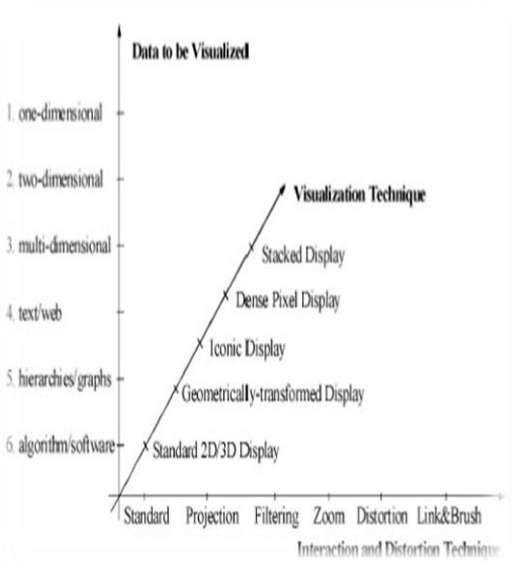

Figure2: Information visualization classification techniques.

Visualization techniques such as geometrically transformed display techniques, iconic display techniques, dense display techniques, stacked display techniques were used. Interaction and distortion techniques such as dynamic projections, interactive filtering, interactive distortion, interactive zooming, interactive linking and brushing were also used. Large data set which is very difficult is explored. It is used in fraud detection and data mining. Visual data exploration was dealing with non-homogeneous and noisy data. It did not need understanding of complex mathematical algorithms. Visual data exploration aims to present data in visual form. It was interacting with user so that human insight into data and could draw conclusions.

Classification of visual data mining techniques involved the x-y plots, line plots and histograms which were limited to small and low dimensional data sets. Data sets were to be one-dimensional, two-dimensional, multidimensional or any complex data types. It was necessary to combine visualization methods to overcome limitations of single techniques. One technique can't fulfil the requirement of user or not efficient for better data exploration.

- Decision-theoretic framework for evaluating data mining system was developed [4]. It also provides classification methods based on their utility in decision-making. This model provided value of extracted knowledge to give better payoff to company. Combined data mining systems were used for decision-making process to increase payoff of organization. Several examples were discussed for better understanding of framework. The data to be extracted should be related to designs, materials, machines, inventories, marketing and might include patterns, dependencies etc. To collect such data databases were required which could contain large number of records and attributes that were important in decision-making process. Hence KDD concept evolved (Knowledge Discovery in Databases).

  KDD process was defined as the non-trival process of identifying valid, accurate, potentially useful and understandable patterns in the data. KDD process contained some steps for data extraction. KDD(knowledge discovery in databases) process which involves understanding the application domain, selecting, integrating, and checking the target data set, data preprocessing, data mining for extracting pattern from data, interpreting and presenting results for the decision maker. Here the data was compared hence different classification models were introduced in terms of their utility in decision making. It also included the visualization techniques. It had given the relationship between investment cost and quality results. Terms such as decision-theoretic in terms of their classification model, optimal investment utility in decision in KDD, combining disjoint data mining systems are studied. Different classification methods, in terms of their value in decision-making for evaluating data mining system were used. It discussed the objectives such as KDD investments and data mining quality. Decision-maker would like to know how much to invest in KDD process. For example medical patients are to be treated and that treatment record is stored in any database, then for any new patient doctor can find the most effective treatment to treat that patient. This is done on large databases.

  In this data mining quality was the main goal to achieve. In above topic techniques were discussed for visualization. It was not so powerful and secured as this framework is. Concept of actionability that is the ability that the extracted information to suggest concrete and profitable action by decision-maker. It is generally used in marketing. As in marketing the information is not static so the previously done classifications were shown to be invalid.
- World Wide Web is exploring day by day. Many users are using WWW for their work. In website various information contents published as HTML pages on Internet.Mining the informative structure of a news website were seen to be difficult. These news websites consisted of hyperlinked documents. These structures of webpage are called as TOC (table of contents, pages). Entropy based analysis (LAMIS) mechanism was proposed based on the Hyperlink Induced Topics Search (HITS) algorithm [3]. It was used for analyzing the entropy of anchor texts and links to remove the redundancy of hyperlinked structure. By this complex structure of website was removed. Many sites published their pages with redundant information to increase value and the accessibility of page. Redundant information consists of navigation panels, copy announcements, advertisements etc. To remove such redundancy mechanism called as InfoDiscoverer had proposed. It helped to extract informative contents blocks.

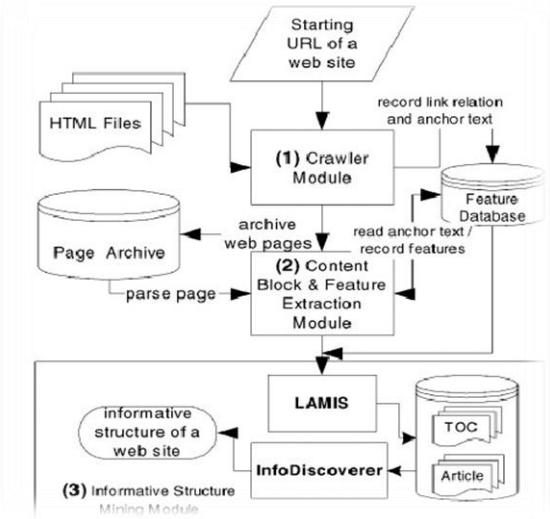

Figure3: The system architecture of LAMIS web mining system

This approach is much better than other conventional methods in mining the informative structures of news websites. LAMIS increases the accuracy by factor ranging from 122 to 257 percent.During experiments performed of LAMIS, HITS-related algorithms were not so good enough to applied in mining the informative structures. The precision and the recall of InfoDiscoverer were greater than 0.956 as compared with manual heuristics. These results were useful for search engines, extracts and navigate significant information from website. Ranking of informative pages from website was done according to the

closest match of user's destination. Data sets used here were 14 Chinese and 5 English news websites

- Development of computational cognitive model that explained navigation behaviour on the World Wide Web had been described. SNIF-ACT (scent-based navigation and Information Foraging in the ACT cognitive architecture) model is proposed here [1]. It provided that how people used information scent cues, such as text related to web links, to make navigation decisions where to go next on the web. SNIF-ACT model was motivated by Information Foraging Theory (IFT), perceived relevance of a web link which was quantified to a use's goal by spreading activation mechanism. Links on webpage evaluated sequentially by user and decided to click on a link or to go back to previous page, by a Bayesian Satisficing Model (BSM) that worked on previous and current assessments of the relevance of link to information goals. Version 1.0 of SNIF-ACT utilized the measure of utility called information.

  Making planning travel around traffic conditions, an investment, or finding a restaurant were some of the problems faced in everyday life. So number of people turned to the World Wide Web in search of above information. The model was successful in predicting users' behaviour in tasks which were most visited.

  First version of SNIF-ACT i.e SNIF-ACT1.0 was developed to make a small number of users working on a small number of tasks. SNIF-ACT 2.0 evaluated first version by including the BSM in evaluation of web links. It worked on the information scent and selected the actions which were currently attended. It did not understand the user's requirement propennnnnrly. It works on sequential links.
- Webpage structure understanding and natural language sentence processing were the two main tasks for information extraction. Some work was also done on integrated statistical model which understand webpage structure and processes sentences of natural language in the HTML elements. Two main branches for webpage understanding are template-dependent and template-independent approaches. Here template-independent approach was used. Hierarchical Conditional Random Fields (HCRF) and extended SemiMarkov Conditional Random Fields (semi-CRF) were introduced [2]. Joint model of HCRF and semi-CRF was used. HCRF model could guide the decision-making of semi-CRF model. But the drawback is that decision of semi-CRF could not be used by the HCRF model. Framework for WebNLP is approached here, which helps in bidirectional integration of page structure understanding and text understanding in an iterative manner. It consists of two components structure understanding and text understanding.

  As there are large number of webpages, links and raw data on WWW, much benifits were not obtained as extraction of information takes place accurately and organized well. Hence results of text understanding can be used to guide the decision-making of each other and it also improves the performance. To increase data coverage of the windows worked had been done on local entity extraction. The WebNLP framework closes the loop in webpage understanding. It is better than the English local entity extraction and Chinese named entity extraction appear on webpages.

Table 1: LITERATURE SURVEY

Author Name	*Topic Name*	*Year*	*Description*
Danial A. Keim	Information Visualization and Visual Data mining	March 2002	Interaction,Distortion,Visualization Techniques are used.Operation research and simulation can be done on large data warehouse
Yuval Elovici & Dan Braha	A Decision Theoretic Approach to Data Mining	January 2003	Decision-Theoretic framework,KDD Processes Complexity and cost should be overcome
Hikao,S.Lin, J. Ho & M. Chen	Mining Web Informative Structure and Contents Based on Entropy Analysis	January 2004	System composed of LAMIS &InfoDiscoverer.Data Extraction is not good enough,web domain specific studies con be done
W.T.Fu & P.Pirolli	SNIT-ACT:A Cognitive Model of User Navigation on the World Wide Web	2007	ACT-R,CoLiDeS & SNIF-ACT modules are used.Prediction of site when user first want to search information is not detected
Y.Yang, Y.Cao,Z.Nie,J. Zhou & J.Wen	Closing the Loop in Webpage Understand	May 2010	Models such as HCRF,Semi-CRF,WebNLP Framework are used.Web page structure is not understand

2. EXISTING WORK

Developers develop website according to their understanding. They don't design website as per user's point of view. So the new users don't get the desired path. Reorganization can be made to the website but it is not possible to reconstruct the whole website newly as it contains the familiar contents of website. There are different types of systems, models, frameworks and mechanisms, but they are not so effective in data mining or extraction of data in effective manner. Different models perform different works in extracting information from website. Some are effective some were not effective. Existing work only describe about the models to apply. It also had the cost factor to manage.

3. PROPOSE WORK

Algorithm for Mathematical Programming (MP) model which improve navigability of user through website with the minimum changes in its current structure. Operations on real data sets are performed. Two evaluation metrics which are use to access the performances of improved website are to be introduced. Optimal solution is quickly obtained, large size problem are solved quickly. Reorganization does not take place here. We just add some few links to website such that backtracking does not takes place and saves users time and they get the desired information.

4. CONCLUSION

Algorithm for mathematical model which helps in effective navigation through website is proposed. We can get the correct information by this and the new user can get their path easily as links are minimum. It maintains the accessibility of websites. Therefore maintanance cost reduces. Most disoriented user can use this type of website. It directly links to the desired path of user. MP model can solve large-sized problems in very less time. To evaluate the improved website, two metrics are used. Small path thresholds can give better results.

6. REFERENCES

[1] W.T.Fu and P. Pirolli, "SNIF-ACT: A Cognitive Model of User Navigation on the World Wide Web," Human-Computer Interaction, vol.22, pp.355-412, 2007.

[2] Y.Yang, Y.Cao, Z. Nie, J.Zhou, and J.Wen, "Closing the loop in Webpage Understanding,"IEEE Trans. Knowledge and Data Eng., vol.22, no.5, pp. 639-650, May 2010

[3] H. Kao, S. Lin, J. Ho and M.Chen, "Mining Web Informative Structures and Contents Based on Entropy Analysis, "IEEE Trans. Knowledge and Data Eng., vol 16, no. 1,pp.41-55,Jan. 2004.

[4] Yuval Elovici & Dan Braha, "A Decision Theoretic Approach to Data Mining", IEEE Trans. Knowledge and Data Eng., vol.33, no.1, pp.42-51,Jan 2003.

[5] Danial A. Keim, "Information Visualization and Visual Data mining" IEEE Trans. Knowledge and Data Eng., vol.8, no.1, pp.1-8, March 2002.

[6] Domenico Talia and Paolo Trunfio, "How Distributed Data Mining Tasks can Thrive as Knowledge Services" IEEE Trans. Knowledge and Data Eng., vol.53, no.7, pp.132-137, July 2010.

[7] Gonzalo Mariscal, Oscar Marban and Covadonga Fernandez, "A survey of data mining and knowledge discovery process models and methodologies", The Knowledge Engineering Review, Vol.25,no.2,pp.137-166,2010.

[8] Natarajan Meghanathan,et al.,"A survey on web pre-fetching and web caching techniques in a mobile environment",ITCS,SIP,JSE,PP.119-136,2012.

[9] John G. Vlachogiannis, Nikos D. Hatziargyriou,"Ant colony system-based algorithm for constrained load flow problem", Vol.20,No.3,pp.1241-1249,August 2005.

[10] Thair Nu Phyu,"Survey of classification techniques in data mining",Proceedings of the International MultiConference of Engineers and Computer scientists, Vol I, March 2009.