ISSN: 2321-8169
3754 – 3758

eNavigate: A Survey On Effective and Efficient User Website Navigation

Mr. N. B. Kasar
ME II Computer
DGOI,FOE, Duand
Pune University (MH), India.
Duand, Pune, India
nkumar.kasar@gmail.com

Prof. Priyadarshi Amrit P
Assistant Professor
DGOI,FOE, Duand
Pune University (MH), India.
Duand, Pune, India
amritpriyadarshi@gmail.com

Mr. S. D. Chopade
ME II Computer
DGOI,FOE, Duand
Pune University (MH), India.
Duand, Pune, India
sagarchopade30@gmail.com

Mr. S. P. Ghode
ME II Computer
DGOI,FOE, Duand
Pune University (MH), India.
Duand, Pune, India
shyamghode@gmail.com

Abstract- Web Structure Improvement has attracted much attention now days. The large websites such as E-commerce, universities etc. have lots of pages visited by users daily. The users needed to be navigated effectively and efficiently throughout the website. Each user has its own set of target pages where the stay time is larger than that of other pages. While making website structure improvements the web master must consider the set of target pages of the users. The users' web log must be maintained at the server side which further has to be divided into session and mini sessions. These mini sessions provides the inputs to extract the target pages of a specific user. The improvement in website must be done under some circumstances. The newly added links must satisfy some criteria such as how many number of links can be added.

Previous literature provides some meaningful considerations about static and dynamic websites. The effective website structure improvements along with the set of target pages can be efficiently designed with static and informative websites while it's difficult to consider the target pages in dynamic one. The structure optimization can be done with minimization or maximization strategies. Here in this survey paper we studied some previously published journals to get better knowledge about the web structure improvements for effective user navigation.

Keyword- *Web Navigation, Static and Dynamic Websites, Target pages, Website Optimization, Stay time, Mini Session*

*****

1. Introduction

Designing a complex Web site so that it readily yields its information is tricky. First, different visitors have distinct goals. Second, the same visitor may seek different information at different times. Third, many sites outgrow their original design, accumulating links and pages in unlikely places. Fourth, a site may be designed for a particular use, but may be used in unanticipated ways in practice; the designer's *a priori* expectations may be violated. All too often Web sites are fossils cast in HTML. We challenged the AI community to address this problem by creating *adaptive*

Web sites: *sites that automatically improve their organization and presentation by learning from visitor access patterns.* Many AI advances, both practical and theoretical, have come about in response to such challenges. The quest to build a chess-playing computer, for example, has led to many advances in search techniques (e.g., [4]). Similarly, the autonomous land vehicle project at CMU resulted not only in a highway-cruising vehicle but also in breakthroughs in vision, robotics, and neural networks. We believe that the adaptive Web sites challenge will both drive AI advances and yield valuable technology.

While adaptive Web sites are potentially valuable, their feasibility is unclear *a priori*: can nontrivial adaptations be automated? Will adaptive Web sites run amok, yielding chaos rather than improvement? What is an appropriate division of labor between the automated system and the human webmaster? To investigate these issues empirically, we focus on a case study the problem of *index page synthesis*. An *index page* is a page consisting of links to a set of pages that cover a particular topic (e.g., electric guitars) at a site; index pages are similar to the "hub pages". Index page synthesis is the problem of automatically generating index pages to facilitate efficient navigation of a site, or to offer a novel "view" of the site. While most previous work has focused on creating *customized* versions of a site for individual visitors, our goal is to *transform* the site into a better one; many visitors to the site should benefit from these transformations. Index page synthesis is a first step in this direction; new index pages represent new views of the site that may accord with the way visitors view the information available.

Our basic approach is to analyze the Web site's access logs to find groups of pages that often occur together in user visits; we assume these groups of pages represent coherent topics in users' minds. We desire a reliable technique for finding such sets of pages and turning them into index pages.

Figure 1: A Conventional SQL User Web Navigation

ISSN: 2321-8169
3754 – 3758

Previous work in *concept learning* and *clustering* is relevant to this endeavor. The goal of a concept learning algorithm is to find a conceptual description of a class of objects from a data collection based on examples of the class (positive examples) and examples of objects not in the class (negative examples). Essentially, the algorithm must determine what the objects have in common that distinguishes them from other objects in the collection. However, concept learning algorithms are *supervised* learning algorithms: they require that their inputs be classified in advance. In our domain, all we have is a collection of objects (Web pages) and some data about usage patterns. Clustering algorithms, on the other hand, are *unsupervised*: they take a collection of objects as their input and produce a *partition* of the collection a classification of each and every object into exactly one class, or "cluster". However, clustering has its own drawbacks for our domain: we do not want a partition of all the Web pages, merely to find small sets of pages that belong together. Furthermore, in our domain, pages may be placed into more than one cluster.

2. Sources of Data

Now a day there are n number of data sources are available for User Web Navigation. Users' web log is a major criterion for navigation log database.

2.1 User Web Navigation Log

Websites maintains their users' navigation log in a log file, called as web Log. It contains the web pages visited by the user along with the stay time for each web page visited.

Figure 2: User Web Log.

This data source further can be divided into sessions and mini sessions for effective log maintenance.

The main source of information we rely on is the site's Web server log, which records the pages viewed by each visitor to the site. We rely on the *visit-coherence assumption* : the pages that a user visits during one session at the site One visitor may be interested in electric guitars, and view several pages on that topic; another may be looking for inexpensive keyboards and view all the pertinent pages. We do not assume that *all* pages in a single visit are related. After all, the information we glean from individual visits is noisy; for example, a visitor may pursue multiple distinct tasks in a single visit. However, if a large number of visitors continue to visit and re-visit the same set of pages that provides strong evidence that the pages in the set are related. Thus, we accumulate statistics over many visits by numerous visitors and search for overall trends.

2.2 Real Data set

The real data set was collected from the Music Machines website (http://machines.hyperreal.org) and contained about four million requests that were recorded in a span of four months. This data set is publicly available and has been widely used in the literature. The number of pages in the website that had out-degrees within a specified range. This website has in total 916 pages, of which 716 have an out-degree of 20 or less, with the majority (83 percent) of the remaining pages having 40 links or less.

3. Approaches

3.1 The Conventional Approach

Our survey of other work in this area has led us to formulate five desiderata for an adaptive Web site.

(1) *Avoid creating work for visitors* (*e.g., filling out questionnaires*). Visitors to a site are turned off by extra work, especially if it has no clear reward, and may opt out rather than participate. If the site cannot improve itself without feedback, it will fail if users do not assist.

(2) *Make the Web site easier to use for everyone, including first-time users, casual users, etc.* Customization can be genuinely useful for repeat visitors, but does not benefit first-time users. In addition, one user's customizations do not apply to other users; there is no sharing or aggregation of information across multiple users. Transformation has the potential to overcome both limitations.

(3) *Minimize additional work for the webmaster.* Although human-authored meta information (e.g., XML annotations) may facilitate site adaptively, we should weigh the benefits of this additional information against the cost of authoring it.

(4) *Protect the site's original design from destructive changes.* For safety, we limit ourselves to *nondestructive transformations*: changes to the site that leave existing structure intact. We may add links but not remove them,

ISSN: 2321-8169
3754 – 3758

create pages but not destroy them, add new structures but not scramble existing ones.

(5) *Keep the human webmaster in control.* Clearly, the human webmaster needs to remain in control of the Web site in the foreseeable future both to gain her trust in automatic adaptive techniques and to avoid "disasters".

3.2 Re-organizing Link Model

We consider websites as graph and each page as node and redirecting URL between pages as edges figure-3 shows our website with 15 pages and many links, links as edges represent as 1 or 0 say we have source node is i and node on which out-link connect to j node with link X_{ij}.

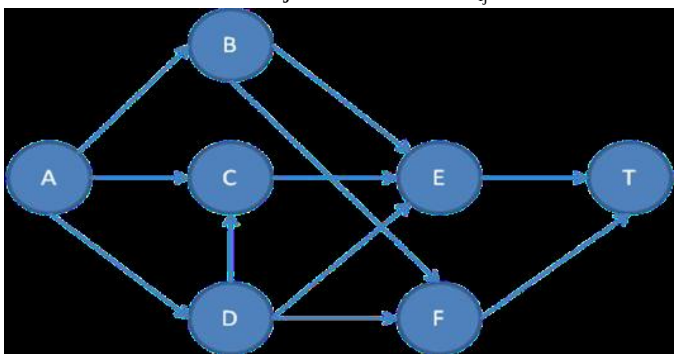

Figure 3: An Example

Here we will form cluster for which we have object are URL on which user is active and URL which satisfies threshold criteria for cluster of links between node i and j for which cluster can be represent as Kij, to perform clustering we will use the similarity measures by farthest distance and we use distance measures is Euclid , here we have two parameter session „S" and Clicks „C" on which cluster is performed for most far value from mean.

4. Towards An Adaptive Web

In order to cope with an ever-increasingly large and complex World Wide Web, there is a demand for intelligent tools and structures which can simplify the experience and make navigation of the sites easier for users and yet maximize the quality and completeness of the experience. These tools and structures should provide sufficient intelligence so that one can sense the environment, perceive and interpret the situations in order to make decisions and to control actions. From our point of view, this is possible through the integration of techniques from multiple disciplines and research areas. One reasonable approach is to combine artificial intelligence, user modelling, graph theory, and information mining techniques to create websites and website browsing tools which are adaptive. *Adaptive* refers to the ability of the website or tool to change its behavior or responses in reaction to the way it is used. An adaptive website should have the ability to recognize users and events, to reason about, and plan for the future.

Web logs are the main source of user behavior data used to continuously tune and adapt the site to it's users. Adaptation may be done in the form of temporarily altering text, links or page format. Creating new pages and adding or removing links are also possible permanent adaptations that may be considered. Although adaptive web technologies belong properly as a subset of the field of adaptive hypermedia, the rapid growth of the World Wide Web has magnified its importance and prevalence to the point where it currently overshadows the larger field. This paper focuses on issues and technologies specifically relating to the Web, mentioning those that belong to the larger field only for historical reasons or when there are useful abstractions or generalizations.

The broadest definition of an adaptive website is a *website which changes based on the way it is used*. Changes can take on many forms (as described in section 3), may either be immediate (as in the case of recommendation systems) or gradual (as in the case of systems which suggest changes to a website administrator). An adaptive website technique may apply to smaller, closed-corpus websites, where the entire site is known in advance, or to the Web in general, where it is virtually impossible to know all the web pages even casually. In the latter case, we might describe it as an adaptive web browsing tool rather than an adaptive website.

5. A Hybrid Web Personalization Model:

The overall process of Web personalization generally consists of three phases: data preparation and transformation, pattern discovery, and recommendation. In traditional collaborative filtering approaches, the pattern discovery phases (e.g., neighborhood formation in the k-nearest-neighbor) as well as the recommendation phase are performed in real time. In contrast, personalization systems based on Web usage mining, perform the pattern discovery phase offline. Data preparation phase transforms raw web log files into

Click stream data that can be processed by data mining tasks. A variety of data mining techniques can be applied to the click stream or Web application data in the pattern discovery phase, such as clustering, association rule mining [1,2], and sequential pattern discovery [3].The recommendation engine considers the active user session in conjunction with the discovered patterns to provide personalized content. The personalized content can take the form of recommended links or products, or targeted advertisements tailored to the user's perceived preferences as determined by the matching usage patterns. In this paper, our focus is specifically on association rule mining and sequential pattern discovery, and the suitability of the resulting patterns for personalization.

6. Challenges for the Recommendation System

Website structure: The way that the website is physically laid out can be useful toward understanding usage behavior and interpreting system suggestions. Furthermore, the *semantic* information about the reasons why the structure exists in the way that it does may also inform the system. In closed-corpus systems, it is possible to know this in detail, but in Web-spanning systems this information is not immediately available, and retrieving it fully is an enormous problem of scale (and may not even be possible). Note that an additional problem occurs when one considers the changeability of web pages: the website structure is easily changed, and is subject to change over time.

Website content: The content of web pages themselves is essential toward determining particular topical interests and understanding the relationships between pages, but suffers even more greatly from the problems of scale, availability and changeability.

User data: Information about the people using the system can help in understanding their interests, or in finding common groups of users which share interests. User profiling is a technique by which user data is gathered from one or more sources, processed and analyzed in order to better understand a user's characteristics and browsing trends. User profile data may be gathered from the client side, server side or from a proxy, either through direct interview or through observed behavior such as purchases or dialogue acts. The data may be categorized as *demographic*, *behavioural*, *attitudinal* or *click stream* data. There are really two types of user data: those that describe individuals and those that describe groups of users. Each individual user profile is based on contextual relevance observed during the user information access.

Information may include demographics, goals and interests, browsing behaviour patterns, browsing capabilities, shopping behaviors, connection speed and type, and human relationships. One of the key problems with user information is the difficulty in obtaining it, and another is the difficulty of verifying the veracity of the data [6]. It has been suggested that users are neither interested in providing this information, nor are they necessarily even willing to provide it for privacy concerns.

Website usage data: Perhaps the most important data set is the recording of interactions of users with the website, in other words, the way that the website is used. This data set may be described in terms of simple page views, transactions (which are "significant" events, and may combine multiple page views), and sessions (which are a combination of page views or transactions that together represent an individual users' experience). In addition to the simple sequence of events, information about time of access and frequency of access is also useful.

7. Applications of recommendation systems

7.1 SEO (Search Engine Optimization)

Google likes consistent navigation. It is good to have consistent navigation not only for users to understand and get the idea of how to navigate through your website but also for search engines to index your website. Search engine robots will crawl through your website in order to index your website and put the links in the search engine results page. If you want to be visible, pay attention to good navigation design and get more traffic.

7.2 Indicate Where You Are

It is crucial to let the user know where he is at all times. You can do this by changing the link's background, color of the page name or turn the text bold in the navigation menu to make it different from others.

7.3 Provide Web Context

To be consistent with your content and navigation provide some context for website users to find things they need quickly. You can place small icons related to the content you link to or short descriptions to give an overview of what the page is about.

8. Conclusion

In this paper we have studied the survey of Website structure improvement over different issues which include different techniques used for Website structure improvement, challenges of the Website structure improvement and the applications of it. Website structure improvement is a prominent field of the data mining used to extract the essential knowledge from a enormous amount of users web navigation logs, stay time on each page. Users' mini sessions are most helpful to recognize and predict the users' target pages. Big organizations use the Website structure improvement for search engine optimization applications. A lot of work is done in the field of Website structure improvement in the form of collaborative and non-collaborative. Still lots of drawbacks and challenges are there like detecting the fake links and logs so on. Now a day it is founded that the Website structure improvement is always be a ongoing research field for future researchers.

Acknowledgement

I express great many thanks to Prof. Priyadarshi Amrit P and for his great effort of supervising and leading me, to accomplish this fine work. To college and department staff, they were a great source of support and encouragement. To my friends and family, for their warm, kind encourages and loves. To every person who gave me something too light along my pathway. I thanks for believing in me.

References

[1] M. Perkowitz and O. Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study," Artificial Intelligence, vol. 118, pp. 245-275, 2000.

[2] M. Kilfoil et al., "Toward an Adaptive Web: The State of the Art and Science," Proc. Comm. Network and Services Research Conf., pp. 119-130, 2003.

[3] M. Nakagawa and B. Mobasher, "A Hybrid Web Personalization Model Based on Site Connectivity," Proc. Web Knowledge Discovery Data Mining Workshop, pp. 59-70, 2003.

[4] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proc. ACM SIGMOD Conference on Management of Data, 1993, pp. 207–216.

[5] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. Verkamo, in: Fast Discovery of Association Rules, MIT Press, Cambridge, MA, 1996, pp. 307–328.

[6] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proc. 20th VLDB Conference, 1994.

[7] T. Anantharaman, M. Campbell, F. Hsu, Singular extensions: Adding selectivity to brute-force searching, Artificial Intelligence 43 (1) (1990) 99–109

[8] Altavista's Babel Fish translator. http://babelfish.altavista.com/.

[9] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. *Journal of the ACM SIGMOD*, pages 207–216, 1993.

[10] L. Ardissono, L. Console, and I. Torre. An adaptive system for the personalized access to news. *AI Communications*, 14(3):129–147, 2001.

[11] M. Balabanovic. An adaptive web page recommendation service. In *Proceedings of the 1st International Conference on Autonomous Agents*, pages 378–385, Marina del Rey, CA, USA, 1997

[12] R. Agarwal, C. Aggarwal, and V. Prasad. A tree projection algorithm for generation of frequent itemsets. In *Proceedings of the High PerformanceData Mining Workshop*, Puerto Rico, April 1999.

[13] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*, Santiago, Chile, Sept 1994.

[14] R. Agrawal and R. Srikant. Mining sequential patterns. In *Proceedings of the International Conferenceon Data Engineering (ICDE'95)*, Taipei, Taiwan, March 1995.

[15] B. Berendt, B. Mobasher, M. Nakagawa, and M. Spiliopoulou. The impact of site structure and user environment on session reconstruction in web usage analysis. In *Proceedings of the 4th WebKDD 2002 Workshop, at the ACM-SIGKDD Conference on Knowledge Discovery in Databases(KDD'2000)*, Edmonton, Canada, July 2002.

Authors

Prof. Priyadarshi Amrit. received his B.Tech. degree from University of Kochin in 2005 and M.Tech Degree in 2010. He has 09 years of teaching experience at undergraduate and postgraduate level. Currently he is working as Assistant Professor in Department of Computer Engineering of DGOI, FOE, Swami-Chincholi, Daund, Pune. His research interests are Data Mining, Image processing and Cloud Computing.

Mr. Kasar Nitin B. received his B.E. degree in Computer Engineering from Shivaji University, Kolhapur in 2011. He is currently working toward the M.E. Degree in Computer Engineering from University of Pune, He has more than 02 years of industrial experience. His research interests lies in Data Mining, Query Recommendation, and Natural Language Processing.

Mr. Chopade Sagar D. received his B.E. degree in Computer Engineering from University of Pune in 2011. He is currently working toward the M.E. Degree in Computer Engineering from University of Pune, Pune. He has more than 02 years of industrial experience. His research interests lies in Data Mining, Query Recommendation, and Natural Language Processing.

Mr. Ghode Shyam P. received his B.E. degree in Information Technology from University of Pune in 2012. He is currently working toward the M.E. Degree in Computer Engineering from University of Pune. He has 02 years of industrial experience. His research interests lies in Data Mining and Natural Language Processing.