

Introducing Navigation Graphs as a Technique for Improving WWW User Browsing

Christos Bouras^{1,2}, Agisilaos Konidaris^{1,2}, Eleni Konidari³, and Afrodite Sevasti^{1,2}

¹ Computer Technology Institute-CTI, Kolokotroni 3, 26221 Patras, Greece
{bouras, konidari, sevastia}@cti.gr

² Department of Computer Engineering and Informatics, University of Patras, 26500 Rion, Patras, Greece

³ Child Development Centre, Department of Psychology, Heraclion, Greece
kondel@otenet.gr

Abstract. In this paper we show that users move from page to page, in the WWW, by following similar browsing patterns when they have similar browsing experience. Similar browsing patterns do not absolutely imply that the same URL's are visited. They only mean that users react in similar ways to similar web sites and follow almost foreseen routes. In order to prove this conclusion we introduce the Navigation Graph technique to trace users' web browsing patterns. We also categorize sites according to their structure by using a commercial tool. Having done these, we analyze the results of the navigation graphs and present them, pointing out the problems that users may encounter. Finally, in order to answer to these problems, we propose enhancements that can improve the performance of web browsers and web servers.

1 Introduction

The extensive use of the World Wide Web (WWW) in our days has established it as the dominant mean for accessing information. The WWW users comprise a heterogeneous group, the members of which have diverse social and professional backgrounds. The information provided and circulated through it, is the element that makes the WWW popular and imminently concerns all users.

It is interesting, however, to examine how the WWW users react to the information presented to them through WWW pages. Depending on the Web sites' structure, organization, presentation, design and appearance, users with different backgrounds and experiences seem to exhibit different reactions. These reactions concern the level of users' perception, their suppleness in navigation, their competence to access the specific information required and the degree of browsing optimization they can achieve. Another significant factor that is worth investigating in combination with the users' behavior, is the structural configuration of Web sites. This configuration is a factor that substantially affects the users' behavior and use of the WWW. It is also directly related to the efficiency and popularity of each Web site.

In this work, the above issues are thoroughly investigated. Initially, an analysis of the structural configurations of Web sites and how this configuration affects the users' behavior, is presented. Consecutively, the term of 'Navigation Graphs' is presented

and thoroughly defined. WWW users' behavior is then examined with respect to the corresponding Navigation Graphs' formation. Based on the latter, the navigating behavior of groups of users with diverse characteristics is being traced, categorized and analyzed. Finally enhancements and improvements to the current organization of Web Sites and the WWW, with respect to the users' needs and gain are suggested.

2 Web Site Mapping

The WWW as we know it today consists of an impressive amount of Web Sites, each one of which comprises of a set of Web pages, containing a varying number of hyperlinks to other resources. From the early stages of the WWW's expansion, there have been attempts to visualize the internal organization of Web Sites and their interconnection in the WWW. These attempts were initiated by the ascertained problems of Web navigation and aspired to conclude to methods for moderating those problems. Here, a short reference will be made to those navigation problems that consist common knowledge among the users and developers of Web Sites.

According to [1], the two major navigation problems on a World Wide Web site are:

- **disorientation:** the tendency to lose sense of location and direction in a non-linear environment
- **cognitive overhead:** the additional effort and concentration necessary to maintain several tasks or trails at one time

In [2] more problems of Web navigation are mentioned. These are:

- **absence of physical context:** the reader sees only one page at a time
- **increased need for graphical context cues:** the reader's idea of what the web site contains must be presented and reinforced on each page
- **lack of control over the arrivals at one page:** the reader can arrive at any page by a variety of methods

Table 1. The results of a global WWW usability study

Percentage of WWW users	Result reported
17.8%	Difficulty in finding pages already visited
8.8 %	Difficulty in visualizing previous and future WWW locations
6.4 %	Difficulty in determining current location
87.7 %	Browsing the Web in an opportunistic manner

Many usability studies support the literature developed about Web navigation problems: according to [3], many users of the WWW cannot find pages already visited and cannot visualize where they are, or where they have been browsing. The results of a global WWW usability study [4] are shown in Table 1.

Another study ([5]) has shown that the 'Back' function of browsers is heavily used to return to a page, but that the history list is not. After all, the use of the history list doesn't resolve most navigation problems and does not contribute at all to the user's orientation within the pages of a Web site or within many visited Web sites.

In [6], it is clearly stated that people follow web links not only because of the page that might immediately be brought up but also because of some eventual set of pages they wish to see which may include, or may be found, inside the next linked page. This ascertainment shows that users often do not behave as the designer of a Web Site would want them to, and this causes poor utilization of the WWW.

In all the above cases, problems of Web navigation are being investigated and the Web users' behavior is analyzed. In most cases users have been asked questions and statistics have been produced according to their answers. In our case study, in the next paragraphs, users have only been asked one question. The question is:

– Do you consider yourself experienced or inexperienced?

The first step towards our analysis is site mapping and is described in the following section.

2.1 Web Site Mapping Results

In order for our experimental results to be produced, we used the log files of an HTTP proxy server over a period of one week. The population of the users that were connected to the WWW through our proxy server was dissimilar, with respect to the users' navigational experience, age and interests. The emerging log files contained approximately 1500 different Web Sites and 100 different users.

From these log files, all different Web Sites were extracted and their Site Maps were created using the I/O/D 4 Web Stalker [11]. The diagrams that occurred were then categorized according to their graphical representation and properties. From this indicative, extended sample of Web Sites, four categories of Web Sites were identified:

1. **Linear Web Sites.** This was one of the most rare cases of Web Site structure. In this category of Web Sites, all the site's pages contain a unique hyperlink to another page. Therefore, there is only one way to 'arrive' to a page and one way to 'depart' from it.
2. **Star-like Web Sites.** In this category of Web Sites, a central Web page (which is usually the home page of each Web Site) contains hyperlinks to all the other pages of the Site. The rest of the pages are only one 'hop' away from the home page and two 'hops' away from one another.
3. **Coherent Web Sites.** In this category of Web Sites, all pages contain hyperlinks to many of the other pages of the Site. The degree of coherency increases with the number of links existent among the pages of the Web Site.
4. **Non-structured Web Sites.** This category contains Web Sites that do not have a special pattern of Site Map. For these cases of Web Sites, the hyperlinks among pages do not follow any rules and have been incidentally created, according to the Web Site's developer needs. Such a Web Site can have a random Site Map layout, of any shape and structure.

Of course several existent Web Sites have a hybrid structure. In fact, the most popular configuration is that of the coherent Web Site with star-like components.

3 Navigation Graphs

Based on the categorization of Web Sites already presented, we can now proceed to the analysis of the users' navigational behavior. There are several already implemented tools that can be used to follow the trail of a Web user and record it. Most of these tools "translate" the user's actions into a directed graph that consists of nodes and arrows.

In [7] the authors have developed a prototype system called 'Footprints' which allows the tracing of users behavior while navigating in existent Web Sites. Internet Cartographer [8] is an application that works alongside the Web browser to classify and map all pages visited by the user. The map can also be used to find Web sites. SurfSerf [9] is another helper application that produces two-dimensional graphs of the Web browsing procedure. It demonstrates the structure of the sites visited and how they are linked together. Natto View [10] is a dynamic 3D visualization tool of the Web and finally, WebPath is another tool that visualizes the users' navigation, thoroughly described in [3].

3.1 Navigation Graphs' Definition

We have introduced a special methodology in order to visualize the vague concept of navigation in the WWW. This methodology consists of the use of what we call "Navigation Graphs" (NG).

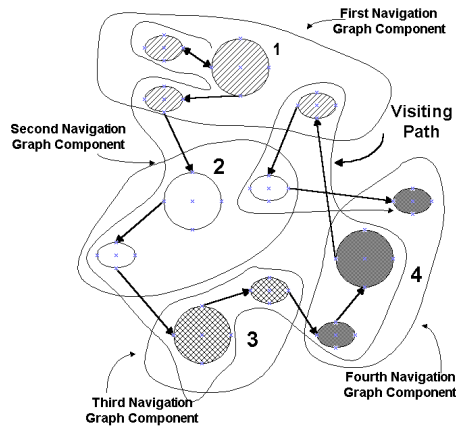


Fig. 1. A Navigation Graph

NGs are graphs that are able to picture the whole procedure of WWW navigation. For the formation of a NG, the URLs written in the navigation bar of the users' browser are taken into consideration in order of appearance.

In our NG notation, visits to pages within a site are represented as ovals, visits to new sites as circles and movements as directed lines (arrows). In Figure 1, all visited pages within a certain site are enclosed inside a perimeter line. Everything inside this

line is a NG component that can be examined separately. Another element of Figure 1 is the curved line that connects the visited pages and is defined as the Visiting Path (VP). The arrows that connect NG components are defined as the backbone of the VP.

3.2 Connections between NG Components

For our analysis in this work, we have chosen to study a NG that consists of all four possible Web site structures, starting with a component with hybrid Web Site structure (coherent and star-like). This NG was chosen due to the evident popularity of this hybrid Web Site, together with the randomness of the rest of components of a NG. In the following section we will examine the VPs created by both inexperienced and experienced users in a NG that consists of the same components.

3.3 The VP of an Inexperienced User

The NG that we chose to examine is the one shown in Figure 2 (A). In the NG of Figure 2 (A) Site A is a Web Site that has the hybrid, coherent and star-like structure (a search engine in our case), Site B is a Web Site that is non-structured, Site C is a Web Site that has the linear structure and Site D is a Web site that has the coherent structure.

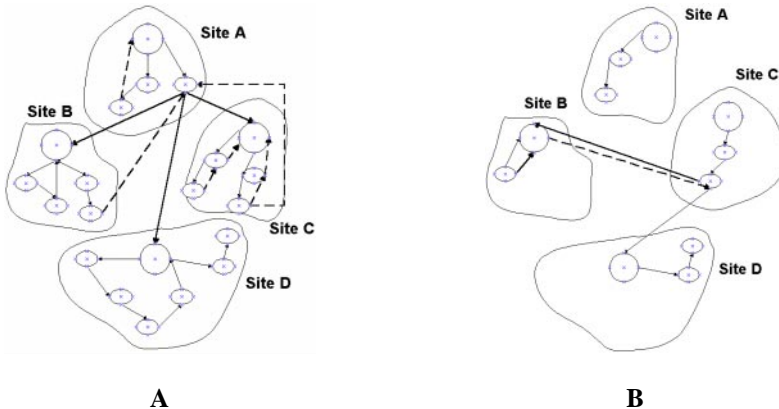


Fig. 2. The NG with the VP created by an inexperienced (A) and an experienced (B) user

The first observation made is that the VP consists of many loops. After observing similar loops in VPs of many inexperienced users, we have concluded that loops are a basic characteristic of Web behavior of inexperienced users. Of course loops are not always due to the user's lack of experience. It is clear that the structure of Site C (linear) is a structure that invokes loops independently of the user's experience. Consequently, many loops in non linear structured sites, and mostly loops in the backbone of the VP, are an indication of a Web user's lack of experience.

The topology of a NG's backbone is a very interesting result in our study. It is an indicator of user experience but also provides valuable hints on how a Web Site should be structured. In Figure 2 (A), one can see that the topology of the backbone of

the NG is star-like and the user is using the BACK button in order to return to the central node. This indicates that the user is using a certain page (most likely a Web search engine result page) as a basis for his/her navigation. This behavior is very common amongst WWW users (experienced or not). There are certain Web pages, or sites, that are the basis of many users' VPs. These sites may change from session to session, or even during a session. In our opinion this is an issue that has to be examined very carefully. It is commonly accepted that a substantial number of users, use certain pages as "milestones" in their navigation. These pages must be treated with special care both from the server's side and the client's side. A suggestion for this problem is given in following sections.

The NG of Figure 2 (A), also gives us an indication of the user's knowledge on browser capabilities. The usage frequency of the "Back" button is a metric of user inexperience. Regarding Figure 2 (A), it is obvious that the user takes advantage of the functionality of the "Back" button a lot (dashed lines). It is obvious that the user's browser belongs to the fourth browser generation, because it provides more than one-step-back capabilities, and the user is aware of that.

The use of the "Back" button is not always positive, in determining user experience. Inexperienced users tend to use the "Back" button of a browser more than it should be used. This can lead to loops and insufficient use of Web resources.

3.4 The VP of an Experienced User

The VP of an experienced user will be examined briefly, and only for cross-reference purposes, namely to show the difference with the path derived from the inexperienced user. In the NG of Figure 2 (B), Web Sites A, B, C and D belong individually to the same categories as the ones represented in Figure 2 (A).

By comparing the NGs of Figures 2 (A) and (B), we can immediately conclude that the experienced user has fewer arrows, fewer loops, fewer pages visited. In other words, the user's VP is restricted with respect to that of the inexperienced users. Another interesting feature of the experienced user's NG is that it is not fully connected. The user has not used a hyperlink to move from Site A to site B. This means that the user has typed the URL of Site B directly into the Location Bar of his browser. This is another indication of the user's experience.

The only similarity that we can derive from the two NGs is the use of a page as a navigation "milestone". This is very important and leads us to believe that such pages are used by almost all users and, as already mentioned above, should be given special attention.

4 Results

In this section we will present the results that were the outcome of the Proxy files analysis process, the site categorization technique and the NG formation of most of the user navigation patterns found in the proxy files. The results provided by the NG patterns are the following:

- Inexperienced users often fall into browsing loops
- Inexperienced users use the "Back" button more often than experienced users
- Experienced users type URL's almost as often as they use bookmarks
- Both experienced and inexperienced users make the same browsing "mistakes" over and over
- Both experienced and inexperienced users follow the same or similar browsing patterns many times. This means that they often visit the same site and move to the same "next" site

The results of our research would not be the same if we did not know in advance if a user was considered experienced or inexperienced. Browsing patterns can heavily rely on every user's perception abilities. It is obvious that a user can not be easily categorized as experienced or inexperienced. We were able to do this by asking the user in advance if he considered himself experienced or inexperienced. After forming the experienced users NGs we realized that most of them followed similar patterns. The same happened with inexperienced users. Only these NGs were taken into consideration in the results that were mentioned above. This was done in an attempt to eliminate any possible false answers by users.

5 Proposed Enhancements of the Web Navigation Procedure

Our study of Web Navigation and Web Site structure with the use of NGs has led to some interesting ideas on how the whole WWW client-server model could be improved in order to facilitate the WWW users. The proposed enhancements apply to both clients and servers in the WWW. All the proposed enhancements aim at eliminating the users' browser problems that were the result of the Navigation Graph study. According to this general rule we propose enhancements that aim at making the user's browsing pattern as linear as possible since "loops" in navigation are not considered as a good practice. The proposed enhancements are the following:

- **Automated bookmark creation.** In the case of inexperienced users help could be provided in the form of automatic creation of bookmarks based on NGs. A browser that used NGs to record users' browsing patterns would be able to create bookmarks, for the user. These bookmarks could be based on "milestone" pages and help users in future web browsing sessions.
- **Help feature enhancement.** The use of NGs could lead to the provision of personalized help and tips. This would be possible since the browser would "know" by examining the user's NG about the problems that he/she may be encountering. This would result in the creation of personalized help and navigation tutorials.
- **NG archiving.** This could lead to the provision of "reminders" that would examine the current NG and when finding a relation with an older one, pass information concerning possible next links to the user enabling him to choose.
- **Client sided personalized statistics.** With the use of NGs browsers could provide personalized statistics, like how long a user has remained on a page or how often he/she has visited it.

- **Provision of site-category specific help.** By using NGs the browser could define a site's categories (as presented previously in this paper) and present useful information on how to navigate in the specific site.
- **Special treatment of "milestone" pages by the clients and servers.** Since "milestone" pages are generally considered to be very important to users, their "special treatment" by WWW clients and servers could be potentially beneficial. "Special treatment of "milestone" pages may include: Client automatic bookmarking, Client caching, Server caching, Server response priority, Server script pre-execution.

6 Conclusions

The basic conclusion of our work is that the Web navigation and browsing procedures are highly individualized but also follow common patterns. A general remark though, is that most users would be happy if this procedure (visualized as the VP in this work) would become more linear. People do not like to go around in circles or come up with the same results. Diversity and variety is a popular demand in the case of the WWW.

NGs may be used to achieve two goals, the first being that of decoding the vast world of the WWW and the other being that of decoding the user navigational patterns within it. By defining and using NGs, we have come to certain conclusions that may help developers improve their Web Sites and allow users to utilize the full potential of the WWW.

References

1. Conklin, J.: Hypertext: An Introduction and Survey, *IEEE Computer* (1987) 20(9): 17-41
2. Kahn, P.: Mapping Web Sites: Planning Diagrams to Site Maps, A seminar of a series from Dynamic Diagrams, Inc found at <http://www.dynamicdiagrams.net/seminars/mapping/maptoc.htm>
3. Frécon, E. Smith, G.: WebPath - A three-dimensional Web History, *IEEE Symposium on Information Visualization (InfoVis '98)*, Chapel Hill, NC, USA
4. GVU's WWW Surveying Team.: GVU's 8th WWW User Survey', October-November 1997, found at http://www.cc.gatech.edu/gvu/user_surveys/survey-1997-10/
5. Cartledge, L., Pitkow, J.: Characterising Browsing Strategies in the World-Wide-Web, *Proc. of the 3rd international World Wide Web Conference*, Germany, 1996.
6. Furnas, G.: Effective View Navigation, *Proceedings of CHI'97*, ACM Press, 1997
7. Wexelblat, A., Maes, P.: Footprints: Visualizing Histories for Web Browsing, MIT Media Lab, 1998
8. Inventix Software: Makers of Internet Cartographer, found at: <http://www.inventix.com/>
9. SurfSerf, found at: <http://www.surfserf.com/>
10. The Natto View, found at: <http://www.mos.ics.keio.ac.jp/NattoView/>
11. I/O/D Index, found at: <http://www.backspace.org/iod/>