S. Bødker, N. O. Bouvin, W. Lutters ,V. Wulf and L. Ciolfi (eds.)
ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work,
24-28 September 2011, Aarhus, Denmark

Group Crumb: Sharing Web Navigation by Visualizing Group Traces on the Web

Qing Wang[1], Gaoqiang Zheng[2], Ya Li[3], Huiyou Chang[1], Hongyang Chao[1]
School of Software, Sun Yat-sen University
[1](wangq79, isschy, isschhy)@mail.sysu.edu.cn, [2]kenkofox@qq.com,
[3]gzhu_liya@gzhu.edu.cn

Abstract. Although the sharing of Web navigation experiences can be useful, it is not supported by contemporary browsers. The Web has been constructed along the lines of a spatial metaphor, but with a flaw of not being able to share navigation experiences, that is, group traces, as is possible in a physical space. This paper shows that from the viewpoint of Information Foraging Theory, sharing Web navigation experiences among group members can increase their information foraging performance. To verify this, a simple prototype, the Group Crumb Prototype (GCP), has been designed. The GCP visualizes *group Web traces* by altering the appearance of links on a Web page according to their Group Crumb Scents, which are calculated from the recentness and times of group navigations to corresponding links. A longitudinal user study has been conducted to compare user performance and experience when surfing the Web with and without the aid of the GCP. Results show that making group navigation traces available on Web pages to group members increases their Web information foraging performance, promotes group collaboration, and enhances their Web browsing user experience as well.

Introduction

Web browsing was originally envisioned as a solitary activity, and most contemporary browsers were designed for that purpose, as were Web sites in the era before Web 2.0. However, the Web has evolved into the era of Web 2.0, which is more concerned with collaboration between Web users than solitary Web activities. Many Web sites and applications have been designed to facilitate various kinds of sharing and cooperation, but ironically the browser itself has not. Almost everything is shared on the Web, except for the most basic activity, Web navigation or browsing path.

The Web navigation system, including Web page links and browser navigation functions, is built along the lines of a spatial metaphor and provides an important foundation for the Web (Bertel, 2001; Marshall & Shipman, 1993; Stanton & Baber, 1994). Unfortunately, its imitation of the spatial metaphor is incomplete due to the absence of a trace-sharing function (Wexelblat & Maes, 1999). In other words, in a physical space, before making a navigation choice, a user can observe other user traces, both current and past; on the current Web, however, this is not possible.

In fact, as stated in Vannevar Bush's famous 1939 article "As We May Think," which is often cited as an early source of hypertext ideas, proposed not only the idea of links between information, but that people might share the "trails" they create through information space. Group traces can be useful in making navigation choices, irrespective of where we are, either in a physical or the artificial Web space. Previous studies have shown that collaborative Web navigation is an integral part of users' information retrieval practices in many domains, particularly education (Amershi & Morris, 2008; Large et al., 2002; Twidale, Nichols, & Paice, 1997) and knowledge acquisition (Fidel et al., 2000; Hansen & Järvelin, 2005; Morris, 2008). For example, school children work together to find information for group homework assignments (Amershi & Morris, 2008), and academics collaborate on literature searches for jointly-authored publications (Morris, 2008).

HCI (Human-Computer Interaction) and IR (Information Retrieval) researchers have designed several tools aimed at facilitating collaborative Web browsing and navigation, specifically collaborative Web search tools (Amershi & Morris, 2008; Diamadis & Polyzos, 2004; Freyne & Smyth, 2006; Morris & Horvitz, 2007; Pickens et al., 2008) and collaborative Web browsing tools (Anupam et al., 2000; Brandenburg et al., 1998; Graham, 1997; Greenberg & Roseman, 1996; Yeh et al., 1996). These tools provide support for activities such as group query histories, shared views of searching result lists, identical browsing experience, and collaboration on Web activity awareness. Nevertheless, each of these tools either only focuses on a particular specific domain of Web usage or only provides pieces of navigation information in a particular scenario for group collaboration. They seldom address the flaw in current Web navigation systems of a lack of trace-sharing. Consequently, all these tools and approaches have limited usage in terms of common group information foraging on the Web.

Here, we define a group Web trace as the accumulated group visitations on a particular Web page (URL), and we believe that if we make the current Web navigation system more consistent with the spatial metaphor by visualizing group Web traces properly, it can promote Web users' information foraging performance, user experience and group awareness at the same time. Our study attempts to answer two research questions:

- Theoretically, how will information about group Web traces affect a Web user's information foraging performance?
- In practice, how will the presentation of group Web traces impact Web users' information foraging performance and their user experience and group awareness?

In the rest of this paper, we first review related works on navigation, collaboration, and information foraging on the Web. We then analyze the impact of sharing group Web traces from the viewpoint of Information Foraging Theory (Pirolli, 2007). Thereafter, we introduce the design concerns of visualization of group Web traces on a Web page. Next, we describe the design and implementation details of our prototype, the Group Crumb. We report on a longitudinal user study conducted to test the impact of sharing group Web traces by comparing users' Web information foraging performance and experience with and without the assistance of the Group Crumb. We analyze the data collected to ascertain research results. Finally, we conclude by discussing the contributions of this paper and suggesting possible future work.

Related work

Flaws in the current spatial metaphor have been discussed in several previous studies. Stanton and Baber (1994) have already criticized the absence of group traces and the unquestioned assignment of the spatial metaphor, further elaborated by Bertel (2001). According to the latter, the lack of signs of other users' visiting traces and too many unwanted and unsuitable properties of the physical space, such as direction and distance, are motivated by spatial representations of the Web. These issues can potentially lead to incorrect conclusions. Of all the flaws, the absence of group traces is critical now that the Web has evolved into the Web 2.0 era, which focuses more and more on collaboration among Web users, requiring Web space or hyperspace to be a social or group space rather than a solitary space.

Web navigation is an ongoing topic in Web research literature. According to previous studies, following links is the most important navigation action, accounting for 45.7% (Catledge & Pitkow, 1995), 43.4% (Cockburn & Mckenzie, 2001), and 43.5% (Weinreich et al., 2006) of all navigation actions. Hence an improvement in link following, which may be realized by sharing group traces, could substantially increase the efficiency of navigation on the Web.

Social navigation is a stream of research that explores the ways of organizing users' explicit and implicit feedback to support information navigation (Dieberger et al., 2000). In this research domain, several approaches have been developed. Knowledge Sea II is an E-Learning system which proposes a social adaptive navigation support by summing group traffic/traces of Web pages in the corpus, and then visualizing these traces by multiple shades of blue on a dedicated navigation map/index (Brusilovsky et al., 2004). Juggle is a tool of recognizing URLs in text and then automatically opening them in a browser (Georgia & Dieberger, 1997). By this means it simplifies and facilitates explicit social navigation via communication tools, such as email and instant messengers. Compared with these approaches, the approach proposed here is more implicit, generic and supporting multiple groups.

In the research domain of Collaborative Web Browsing, many approaches have been proposed to provide a synchronous browser for tightly coupled group collaboration. Some approaches use a special browser that allows users to control each other's browsers, to know what they are doing, and to know at which page they are looking. Examples include GroupWeb (Greenberg & Roseman, 1996), Albatross (Weinreich et al., 2006) and GroupScape (Graham, 1997). Other approaches use a common browser, but in different ways. The WebVCR system (Anupam et al., 2000) and Artefact framework (Brandenburg et al., 1998) use client side technology, such as a Java applet, to synchronize or share Web page browsing and to set up an additional communication channel for users. These Collaborative Web Browsing methods provide a totally synchronous browser experience between group members. This is the opposite of what we are proposing here, that is, only sharing group traces, the navigation experiences, among group members to promote their information foraging performance and enhance their group awareness.

Utilization of group traces has been observed in many Web research domains. In collaborative Web searches, both the approach of Morris et al. (2008) and that of Sun et al. (2006) use group traces to filter and rank search results. Another interesting tool is WCSA (Diamadis & Polyzos, 2004), in which group traces, called group member URL traversal awareness (GMUTA), are used to help filter search results. In the semantic Web approach of GroupWeb (Grcar et al., 2005), user profiles are mined from group traces. In personal Web searches, the group traces are used to build better queries (Tan et al., 2006), as well as to filter and rank search results (Teevan et al., 2005). Our research differs from the above approaches in that its utilization of group traces is for a different purpose - as a visual heuristic to promote information foraging performance and enhance group awareness.

Mining Web users' navigation patterns is a popular research topic in Web mining. Several studies (AlMurtadha et al., 2010, Ting et al. 2009, Borges and Leven, 2008) have been undertaken in recent years. All these studies mined Web users' navigation patterns and then made predictions to improve users' Web navigation efficiency. Compared with what we are proposing here, these studies are complex and seldom take into account the similarity of group members, whereas, fellow Web users or group members working in the same context can be better predictors than a computer or a complicated algorithm. Meanwhile, these studies predict on patterns mined, without any reasonable explanation of these patterns, which is also different to what we do in this study. We show group navigation traces based on Information Foraging Theory, and explain the effectiveness thereof from a cognitive science viewpoint.

In fact, the sharing of group traces is also an interesting research topic out of the Web studies. Hill et al. (1992) shared the history of interactions between author and reader within documents. Deline et al. (2005) found sharing source code navigation among programmers can improve their comprehension on code. These studies showed sharing group traces promoted users' work performance and enhanced their group awareness. In this paper we extended these studies and found similar results in the information foraging on the Web.

Information foraging

Information Foraging Theory has been used to explain and predict Web users' information foraging actions, especially their navigation choices, on the Web. It models the Web as a space with many information patches, from which Web users can forage information and between which they can travel. It also recognizes that a forager's information gain within a patch is a diminishing function of her within-patch time. Then according to Charnov's Marginal Value Theorem (Charnov, 1976), the within-patch time t_W, between-patch time t_B, user's information gain function g, and average information foraging rate R satisfy,

$$R = g'(t_w) \qquad (1)$$

Furthermore, Information Foraging Theory shows that there are two effects of *between-patch enrichment,* in which a Web user's traveling cost (time) decreases. First, the *within-patch time* also decreases; and second, the *overall information foraging performance* increases (Figure 1).

From an Information Foraging Theory's viewpoint, sharing *group Web traces* among members on the Web is a *between-patch enrichment*. For a Web user with a certain information foraging goal, a Web page may serve her in two ways, either as an *information patch/destination,* from which she forages information, or as a *navigation spot* which is merely by-passed as she finds a path to her destination. Obviously, the *between-path time* of her information foraging on the Web depends on the probability of choosing the right path; in other words, clicking on the correct links on *navigation spots*. If a user has a higher *correct choice probability*, she will have a smaller expected *between-patch time*, which eventually leads to *between-patch enrichment*.

The SNIF-ACT (Fu & Pirolli, 2007) model of Information Foraging Theory predicts a user's navigation choice with a goal G and candidate links C, as a probability function of a given link L's utility $U_{L|G}$,

$$\Pr(L \mid G, C) = \Pr(U_{L|G} > U_{K|G}, \forall K \in C) \qquad (2)$$

According to the Random Utility Model (McFadden, 1974; McFadden et al., 1978), L's utility $U_{L|G}$ in the context of goal G is defined as the sum of all activations received by cognitive chunks representing a user's goal from proximal cues associated with a link, plus some stochastic noise ε.

$$U_{L|G} = \sum_{i \in G}(B_i + \sum_{j \in L} W_j S_{ji}) + \varepsilon \qquad (3)$$

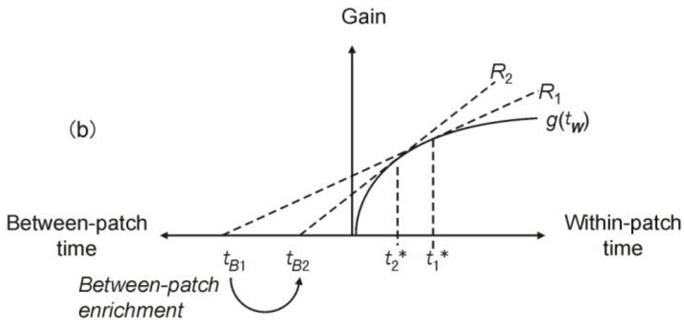

Figure 1. A between-patch enrichment of information patch model. The average rate of information gain, R, increases with a decrease in between-patch time, while simultaneously decreasing the within-patch time (Pirolli, 2007).

Here B_i is the base-level activation of chunk *i*, and S_{ji} is the association strength between an associated cue *j* and chunk *i*. Each proximal cue *j* emits a source activation W_j, which reflects the Web user's attention at the beginning. These source activations spread across cognitive Spreading Activation (Anderson & Pirolli, 1984) to feature *i* that is part of the information goal *G*.

In the *group Web traces* sharing scenario, we assume that there is a *group navigation similarity*, which means that typically group members trust each other and they may have a greater probability of choosing the same navigation link(s) on a Web page. This assumption has been verified in our study described in the *User Study* section.

Therefore, links with *group Web traces* are more likely to be links a Web user may choose to follow, and showing *group Web traces* on a Web page both provides additional heuristics and attracts more attention to these links, which in turn increases the source activation W_j for each proximal cue *j* of these links. Moreover, these additional cues and increased source activations then spread through cognitive Spreading Activation and hence, lead to a higher *correct choice possibility*, which eventually decreases the *between-patch time*, and as a final result, *between-patch enrichment* occurs.

At this point, it is clear that sharing *group Web traces* among group members realizes *between-path enrichment* of Web information foraging. According to Information Foraging Theory, we therefore, predict that this sharing will lead to a smaller *within-patch time* and higher *overall information foraging performance* for each group member on the Web.

Visualization of group Web traces

Although according to Information Foraging Theory, sharing *group Web traces* increases group members' information foraging performance, given here is actually an implied premise that while sharing group trace information, this sharing will not sabotage Web users' normal cognitive processes on other information within corresponding Web pages. In other words, we must visualize *group Web traces* with little disruption and proper informativeness to a user's cognitive process. Hence, from a cognitive science viewpoint, it is better to show *group Web traces* as some visual hint(s) on links, such as font color, font weight, background color, text decorations, etc., and any explicit additional elements, such as a direct text description or icon, is not an option.

In fact, there is already a similar visual hint, the Bread Crumb (Bernstein et al., 1991), in all contemporary browsers. The BC alters the color of the text and underscore of recently visited links, discriminating them from unvisited links. This visualization has a long history, since about 1991, and is an indispensable part of the spatial metaphor by providing signs of where we have been. Actually, regarding *group Web traces*, the BC is a kind of personal trace on the Web. So, it is reasonable to design the visualization of *group Web traces* in the same way as the BC, which has already been accepted by Web users. Hence we called the group trace visual hint, Group Crumb (GC). Like the Bread Crumb, the GC shows *group Web traces* by altering the appearance of links. Specifically, it alters the font weight and font size of links and adds group visitation information to the tooltips. Here the design decision is based on the principle of no additional elements deduced from cognitive science in the previous paragraph. It does not consider image links, since the visualization here is only for a research prototype verifying our prediction on sharing group Web traces, and this omission does not affect the result.

System design and implementation

We have designed a simple research prototype, called the Group Crumb Prototype (GCP) to evaluate the impact of sharing group Web traces on the information foraging performance and user experience of group members. It consist of two parts, a Firefox extension, called the Group Crumb Extension (GCE), which visualizes Group Crumbs by altering the font size and font weight of links on a Web page, and a server on Django, called the Group Crumb Django (GCD), which stores group members' visitation data in a MySQL database and answers queries from the GCEs (Figure 2).

In practice, when a user opens/enters[1] a Web page, the GCE sends the Web page's URL to the GCD, and queries the GCD about Group Crumb Scents (GCS) of links on that Web page. The GCE then alters the font size and font weight of links on the Web page according to the query result, and adds group visitation information on tooltips (Figure 4). Generally speaking, a link visited recently by more group members will be shown in a bigger font size and a stronger font weight (Figure 4).

[1] In contemporary multi-window and multi-tab browsers, a user can either open/close a Web page, or enter/leave a Web page in another window or tab.

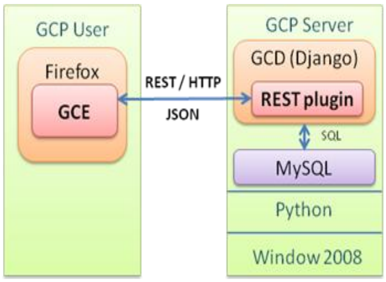

Figure 2. Group Crumb architecture

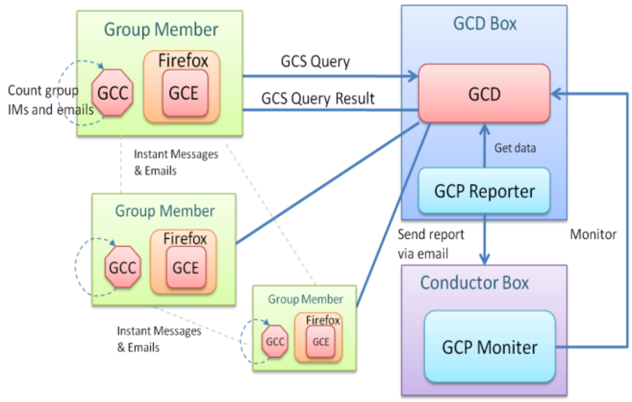

Figure 3. GCP study environment

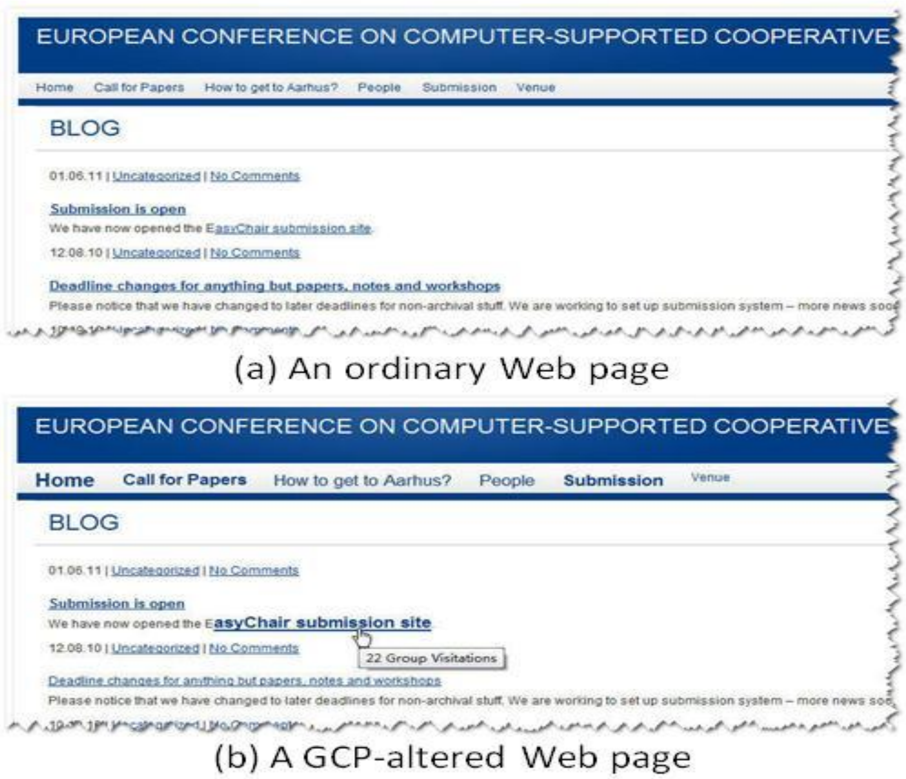

Figure 4. A partial Web page rendering

Group Crumb Scent

The GCP calculates and then uses the GCS of a link as the dominant factor determining its font size and font-weight. In practice the GCS is calculated in the following way.

The *group visitation scent* on a certain link is an imitation of human memory, which follows the rough formula of a *forgetting curve* (Ebbinghaus, 1913), where t is the number of days since the visitation andλ is the relative strength of memory, which is set to 7 here according to the study of Obendorf et al. on Web page revisitation patterns (Obendorf et al., 2007). In this study, Web users' revisitations older than 7 days (one week) are categorized as 'long-term' revisitations, which despite being valuable, are not well supported in contemporary browsers.

$$s = e^{-\frac{t}{\lambda}} \tag{4}$$

For a particular link l, we sum all the *group visitation scents* S_v as the *group link scent* S_l.

$$S_l = \sum_{v \in V_l} S_v = \sum_{v \in V_l} e^{-\frac{t_v}{\lambda}} \tag{5}$$

Then, the GCS of a certain link is calculated by normalizing its *group link scent* S_l in the range [-1, 1], with regard to the *group link scents* S of all links on the same Web page.

$$G_l = \frac{2 \times S_l - \max(S) - \min(S)}{\max(S) - \min(S)} \tag{6}$$

Finally, GCS is used to alter the font size and font weight of links. In equations (7) and (8), f_s and f_w represent the original values of the link's font size and font-weight, respectively, and f'_s and f'_w are the respective GCP-altered values.

$$f_s^{'} = (1 + \mu \times G_l) \times f_s \tag{7}$$

$$f_w^{'} = \begin{cases} \min(100 \times [\mu \times G_l \times 4] + f_w, 900), G_l > 0 \\ \max(100 \times [\mu \times G_l \times 4] + f_w, 100), G_l < 0 \end{cases} \tag{8}$$

In the above formulas, μ is a configurable parameter used to control the strength of GCS, and f_w is restricted to integer multiples of 100, no greater than 900 according to CSS 2[2].

Foraging vs. Passing

As stated in the *Information Foraging* section, the Group Crumb should separate group members' foraging visitations on information patches from their passing visitations on navigation spots, and only calculate GCS from these foraging visitations.

However, this separation can hardly be done precisely without direct identification of the user, which is obviously impossible in the GCP's implementation. Therefore, we have to find an approximate solution. Fortunately, according to the study of Cockburn et al. (Cockburn & Mckenzie, 2001), Web pages viewed by web users for less than 10 seconds are often passing-through pages, and not pages they are really interested in. Hence we use this heuristic in the GCP to filter these *passing visitations*; that is, the GCE sends all the user's visitation data to the GCD, and the GCD tracks the current viewed Web page of the user, and only considers pages viewed for more than 10 seconds when calculating the GCS.

2 CSS 2 Fonts, http://www.w3.org/TR/CSS21/fonts.html#font-boldness.

GCS Query

The GCP stores group foraging visitations and calculates GCSs on its server, the GCD. When a user opens/enters a Web page, the GCE sends a *GCS query* with GCE_ID, GROUP_ID, URL, ACTION_TYPE[3], and VISIT_TIME to the GCD. The GCD then calculates the GCSs of all links on the Web page according to the URL and GROUP_ID and wraps them all in a corresponding *GCD answer*. Moreover, the GCD caches the URL and VISIT_TIME, and will add a visitation to the former URL if the interval between the two VISIT_TIMEs is more than 10 seconds[4].

According to (8), before altering a Web page, a GCE must know all the GCSs of the links on it. This means that a GCE cannot render *group Web traces* on a Web page properly before the *GCD answer* arrives. Obviously, this delay may cause a user's browser to freeze and thus, have a negative effect on the user's experience. To resolve this issue, the GCP adopts two strategies. First, a *GCE query* is sent asynchronously without blocking the normal page rendering and the GCE visualizes GCs by means of animated changes on the link appearance when the *GCD answer* arrives. Second, instead of sending all links of a Web page in a *GCS query*, only the URL of the queried Web page is sent, and then the GCD figures out all the other links on its own. This is more efficient than sending all links of a Web page since the GCD caches/stores the links of queried Web pages, thus reducing the retrieval time of links in follow-up queries.

Group Crumb Server

We implemented the GCD as a simple Django application deployed in an Apache HTTP server using the mod_python module. We chose Python and Django for their simplicity, easy and rapid development, and maintainability. The GCD stores all group visitation data in a MySQL database, and answers GCE queries via REST Web Services with data in JSON format.

Although scalability and performance issues were not our primary concern, in practice, the scalability of the GCP is good enough for our research purposes. A GCD can handle at least 200 GCEs concurrently, while each GCE issues a GCS query every 10 seconds on average, and without any data loss at the server or frozen browsers at the client.

Privacy

Another important issue we have to consider here is how to protect user privacy when sharing *group Web traces*. We made two decisions about privacy. First, the GCE only shares *group Web traces* among group members; and second, these *group Web*

---

[3] A Web page can be opened by following a link, typing a URL, reloading a bookmark/history, and these actions are distinguished by the ACTION_TYPE parameter.

[4] In the current GCD, other visitations are also stored for research purposes.

traces are only shown in the containing Web page as a Group Crumb without details of group visitations, that is, a user only knows that a link has been visited by some of the group members, but has no idea exactly who. According to our user study results (see *Privacy* section of *Results and Discussion*), these decisions effectively solved the privacy issues.

Initialization and configuration

The GCE_ID, GCD_URL, and GROUP_ID of a GCE must be set or configured properly before executing the GCE. In the current GCE, for the purpose of our research, the GCD_URL is preconfigured to our GCD server, and GROUP_ID can be set via a prompt window when the GCE is first executed, and later changed in the Firefox preferences. A GCE also issues an INITIAL request to the GCD when executed initially. It retrieves Web page visitations longer than 10 seconds during the past 7 days[5] from Firefox's history records, and wraps these in the INITIAL request to the GCD. The GCD stores and allots these visitations to the group identified by the GROUP_ID in the request, and answers the request with a unique GCE_ID generated. In this way, each GCE receives a unique GCE_ID when initially executed and uses it in the subsequent communications with the GCD.

For the purpose of our user study, the GCD was configured as a 'dummy' server to a certain group(s). This means that the GCD did not answer *GCS queries* from GCEs of the configured group(s), and hence these GCEs did not alert Web pages. However, all visitations of the group(s) were still stored for our study.

Table I. Details of groups

Group in Exp.	Team in Comp.	Members	Project
G1	T1	5	**Campus Map**, a mashup Web application
	T2	6	**QReader**, an E-book reader on Android
G2	T3	6	**AnswerIt**, an answering machine on Android
	T4	5	**Happy Hospital**, an SNS game

Table II. Groups' GCD configuration

	Phase 1	Phase 2
G1	active	dummy
G2	dummy	active

User study

We conducted a 2-month longitudinal user study on the GCP to test our prediction, based on Information Foraging Theory, that sharing group Web traces leads to a smaller *within-patch time* and higher *overall information foraging performance*. Hence the study was designed to capture changes in participants' information foraging

[5] According to (4), visitations older than 7 days have little *scent*.

before, during and after exposure to their Group Crumbs. The study lasted two months from November 10th, 2010 to January 9th, 2011 with two phases each lasted one month (30 days). Participants are grouped into two groups, namely Group 1 (G1), and group 2 (G2). During the experiment participants' browsers were configured either with or without showing Group Crumbs on Web pages they visited, and their performance data were collected to verify our prediction.

Participants, apparatus and environment

We recruited 22 participants (15 male, 7 female), aged between 19 and 21, from the sophomore class of the Software School of Sun Yat-sen University. All the participants had majored in Software Engineering, and were taking part in a one-year software development competition conducted by the Software School (beginning Oct. 1st, 2009). This made these participants perfect study subjects, since they were already grouped and were working collaboratively on the Web in foraging useful information for their competition projects. The number of members and project for each team are shown in Table I.

We separated the four teams into two groups, G1 and G2. All participants were required to use Firefox, with the GCE Extension installed, as their only browser during the study. In the 1st phase of the experiment, the GCD was configured active for G1' GCEs while dummy for G2's, and vice versa in the 2nd phase (Table II).

All participant groups used agile methods in their software process. They all made use of 2-week iterations, and therefore released their software four times during our study.

Because the duration of the study was long, we developed tools to guarantee it ran effectively and smoothly. The first tool was the GCP Monitor, a long-running process that monitored the GCD. It sent a warning email to experimenters when the GCD's service was unavailable. The second was the GCP Reporter, which was deployed as a scheduled task, executed daily, to report on data collected (Figure 3).

To analyze the GCP's impact on group collaboration, all participants were also required to install a tool called the Group Conversation Counter (GCC), which was only run at the end of each phase. This tool counted the group instant messenger dialogs and email conversations of each participant by retrieving them from configured email addresses and instant messenger accounts of her teammates. The group email conversations of a participant were retrieved directly from the POP server(s) of her email box(es), and the instant messenger dialogs were retrieved from the local message cache file of Tencent QQ[6], which is a popular instant messenger and independently chosen by all participant groups as their primary daily conversation tool. Then the GCC generated a GCC result by counting the numbers of email conversations and instant messenger dialogs each day during the longitudinal user study. A GCC result is a plain

6 http://www.qq.com/

text file containing only the numbers of email conversations and instant messenger dialogs. It did not include anything about the content of the conversations, thereby protecting the participants' privacy.

Surveys and data collected

During the longitudinal study, we collected Web information foraging activities for all participants based on the following fields: GCE_ID, GROUP_ID, VISIT_TIME, URL of the Web page, and ACTION_TYPE (Open, Close, Enter, Leave). At the end of the study, participants were required to submit their GCC results and complete an online survey designed to collect free form comments about users' positive and negative experiences with the GCP and to ascertain the GCP's impact on their group awareness and collaboration. The survey also provided information about which situations participants found the Group Crumb most useful (Table III).

Results and discussion

During the longitudinal study, the participants visited a total of 30,691 Web pages, about 22 pages per day per participant ($\mu=22.3$, $\sigma=9.7$), and dwelled more than 10 seconds on 18,722 (61%) of these, about 14 pages per day per participant ($\mu=14.2$, $\sigma=10.2$). The latter pages were regarded as *information patches,* and the others *navigation spots* (39%). 82% of all Web visitations (20,582) were from following links, but the figure drops to 73% if considering only *information patch* visitations. Among all visitations, each participant spent on average 72.3 sec ($\sigma=8.7$) on a Web page, with the figure increasing to 102.3 sec ($\sigma=8.7$) when excluding visitations to *navigation spots*. Participants conducted a total of 2,798 email conversations and 5,766 instant messenger dialogs during the study. All 22 participants completed the online survey, and some provided addition comments on the usability and their experiences with the GCP.

In the following sections, we compare the data in two collections, Collection Active (CA) and Collection Dummy (CD). CA summarizes participants' performance with Group Crumb's help (GCD active), that is, the data of G1 in phase 1 and G2 in phase 2; and CD summarizes participants' performance without Group Crumb's help (GCD in dummy), that is the data of G2 in phase 1 and G1 in phase 2 (Table II). By this cross summing-up, the biases of time and group were very likely been swept off. We also cite survey results to analyze the GCP's impact on group Web information foraging and collaboration. Significance is reported using one-tailed paired t-tests.

Group navigation similarity

The *group navigation similarity* assumption in the *Information Foraging* section was supported by the user study. Either with or without the GCP's help, participants tended to choose those links also visited by their group mates. In the CD, participants selected a total of 6,670 links to follow, about 1% (μ *=.9%,* σ *=.2%*) of all 717,423 links presented on all the visited Web pages. However, regarding in *group navigation similarity*, they selected significantly more links, 21% (μ *=19.7%,* σ *=4.3%*), to follow from the 6,670 links visited by the group ($p<.01$). The same trend was seen in the CA where participants selected 1% (μ *=.9%,* σ *=.2%*), 8,124 links in total, to follow from all 876,852 links presented, and selected significantly more links, 37% (μ *=37.3%,* σ *=4.4%*), to follow from the 8,124 links visited by the group ($p<.01$).

GCP effectiveness

We found that the use of the GCP influenced participants' navigation decisions. As shown in the previous section, with the aid of the GCP participants followed significantly more group visited links in the CA (μ_1*=19.7%,* σ_1*=4.3%,* μ_2*=37.3%,* σ_2*=4.4%,* $p<.01$). Furthermore, in the CD, participants clicked on average 1.2 times (μ *=1.18,* σ *=.13*) on a group visited link; but in the CA, this figure increased significantly to 1.3 (μ *=1.33,* σ *=.12,* $p<.03$). Moreover, the survey results also confirmed that the GCP effectively influenced participants' link following as they preferred to click on a link with a Group Crumb (see 1 in Table II). One participant reported her feelings on the GCP's influence:

> "The GCP gives me great hints about what might be interesting. I prefer to select a GCP link rather than others since my buddies have already polled it."

Between-patch time

We predicted sharing group traces using the GCP as a kind of between-patch enrichment from the viewpoint of Information Foraging Theory. This is consistent with the user study, where both the proportion and dwell time of *navigation spots* dropped after participants were exposed to Group Crumbs. In the CD, without the GCP, 43% (μ *=42.5%,* σ *=3.7%*) of the entire visited Web pages were *passing pages* or *navigation spots*; but in the CA, this figure dropped significantly to 36% (μ *=35.7,* σ *=3.6%,* $p<.03$). Moreover, in the CD, participants spent on average 8 sec (μ *=8.3,* σ *=.7*) on a navigation page; yet in the CA, they only spent 6 sec (μ *=6.7,* σ *=1.1,* $p<.02$). Combining these two facts, it is clear that participants' between-patch time in their information foraging activities on the Web decreased using the GCP, and this confirms the proposition that using the GCP is a *between-patch enrichment*.

Within-patch time

In the CD, participants spent on average 80 sec (μ *=79.5,* σ *=4.2*) on a Web page; but in the CA, they only spent on average 65 sec on a Web page (μ *=65.1,* σ *=4.7,* $p<.01$). Taking only these information patches into account (*within-patch time*), participants spent on average 113 sec per page in the CD (μ *=112.7,* σ *=4.2*), and

92 sec per page in the CA (*μ =92.3, σ =5.2, p<.01*). This is consistent with the indication from Information Foraging Theory that a decrease in *within-patch time* is a consequence of *between-patch enrichment*.

It is also worth mentioning here, that in the CD the average dwell time on group-shared Web pages, visited by more than one group member, was 84 sec (*μ =84.4, σ =4.1*), about 8 sec longer than that for other visited Web pages, 77 sec (*μ =76.8, σ =3.8, p<.04*). In the CA, although the average *within-patch time* decreased, the difference between the dwell-time on group-shared Web pages and others increased significantly to about 14 sec (μ_g*=72.7,* σ_g*=5.2*; μ_o*=59.2,* σ_o*=4.8*; *p<.01*).

Information Foraging Theory predicts that an information forager will leave an information patch when its diminished gains drop to the average information foraging rate, and hence dwell time on a rich information patch with higher diminished gains is longer than on a poor one (see Figure 5). According to this, the longer dwell time on group-shared Web pages suggests that these group-shared Web pages had higher diminished gains than the others, and the increase in the difference in dwell time between group-shared Web pages and others can be interpreted as participants choosing to find more valuable information patches with GCP's help.

Overall information foraging performance

In the CD, participants visited 3,811 information patches (*μ =173.2, σ =18.6*) and spent 429,217 sec on them (μ =19,509, σ =2,030). The average information foraging rate was 0.53 patches per minute (σ =.04). In the CA, participants visited 5,443 information patches (μ =247.4, σ =28.5) and spent 502,607 sec on them (μ =22,846, σ =2,577). The average information foraging rate increased to 0.64 patches per minute (σ =.05, p<.02). This result supported that sharing group traces using the Group Crumb significantly increased participants' *overall information foraging performance*.

GCP users' comments in the survey also confirmed this finding. Participants said:

"I really like the GCP, It makes me more efficient in finding things I am looking for…"

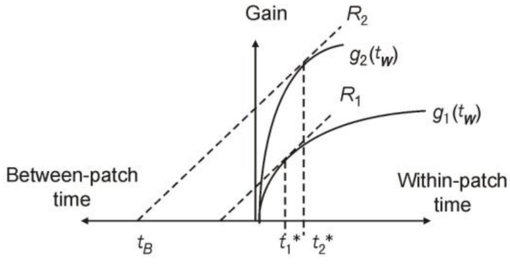

Figure 5. Dwell time difference (time spent in a poor patch t1*, and in a rich patch t2*)

"GCP is a good link filter and recommendation engine for me. It filters those irrelevant links and recommends other interesting links. With its help, I can get information from the Web with less time and errors."

Group awareness and collaboration

Another interesting observation is that the GCP enhanced participants' group awareness and collaboration. In the CA, participants' Web visitations showed a greater group tendency, indicated by a significantly higher proportion of *group-shared* Web pages in *all visited* Web pages than in the CD ($\mu_1=57.5$, $\sigma_1=3.5$, $\mu_2=64.3$, $\sigma_2=3.7$, $p<.02$). This occurred without a significantly longer dwell time on group-shared Web pages than other Web pages ($\mu_1=7.61$, $\sigma_1=3.9$, $\mu_2=13.5$, $\sigma_2=5.1$, $p=.06$).

The same trend was also found in both email conversations and instant messenger dialogs. Together the participants conducted 1244 email conversations in the CD ($\mu=56.5$, $\sigma=4.9$), which increased significantly to 1554 in the CA ($\mu=70.6$, $\sigma=7.3$, $p<.01$). At the same time, they conducted 2365 instant messenger dialogs in the CD ($\mu=107.5$, $\sigma=8.9$), which also increased significantly to 3401 in the CA ($\mu=154.6$, $\sigma=12.3$, $p<.01$). Participants' comments also confirmed this enrichment. One of them said:

"The GCP shows me a group vision of the Web. Knowing your buddy has been somewhere is definitely cool. As you known, following your buddies makes you feel safe and somehow protected in your incoming findings."

Usefulness and usability

Despite the fact that the GCP is only a simple research prototype, our survey results confirmed that it is useful and usable in assisting participants in their information foraging on the Web (Table III).

Table III. Survey Results (Likert-Scale, 1=strongly disagree, ..., 5=strongly agree)

Question	$\tilde{x}$
I prefer to click on a link with a Group Crumb.	4.0
The Group Crumb is useful for navigation on the Web.	4.0
The Group Crumb is useful for highlighting useful links.	4.5
The Group Crumb is useful for filtering irrelevant links.	4.0
I experienced less navigation cost/time with Group Crumb's help.	4.0
My Web information foraging efficiency increased with Group Crumb's help.	3.5
I felt better group awareness with Group Crumb's help.	4.5
The Group Crumb clutters original Web pages.	2.5
The Group Crumb is annoying to me.	1.5
The Group Crumb should be more conspicuous.	3.5
I do not mind showing my entire browsing history to my group mates.	1.5
The Group Crumb compromises my privacy.	1.5

Privacy

Recent research (Brush et al., 2009) shows that only about 20% of Web users are comfortable sharing the URLs of Web pages they recently visited. This applies in our experiments as well, but we also found some more specific results. Participants were not happy to show all URLs they visited to their teammates, but they did not mind showing their group Web traces/visitations using the Group Crumb. According to our survey, only 3 of the 22 participants were happy to show a list of all URLs they visited to their teammates, but none of them thought that the GCP compromised their privacy or that sharing *group Web traces* through GCP was dangerous (Table III).

Limitations of the study

There are some limitations to this study. First, we used a selected, small sample population consisting of university students who were working on a specific collaborative task, programming and developing software. This means that this study is limited in educational arena now and we cannot expect our results to generalize to a larger, less homogeneous population on other tasks. Instead, the results of this study provide insight into how GC impacts a skilled Web user's information foraging on the Web for his/her software developing tasks. Second, we applied the factor found by Cockburn to label visitations shorter than 10 seconds as passing visitations, and this approach might be too simple. Further studies with other sophisticated methods are needed. Third, due to previous teammates' biases or mistakes on their information forging, GC may augments links wrongly and then leads the user missing useful links. How to resolve this problem is still an open problem now. Finally, the small sample size in this study may not provide sufficient statistical significance to draw strong conclusions and a longitudinal study using a larger sample size is required to verify our results.

Conclusions and future work

We found that group Web traces complement the current spatial metaphor in contemporary browsers. We predicted that sharing group Web traces would increase Web users' information foraging performance from the viewpoint of Information Foraging Theory.

We presented the design of the Group Crumb, which follows the Bread Crumb's convention and shows group Web traces by altering the appearance of corresponding links on Web pages. We also presented the Group Crumb Prototype, a novel tool for showing Group Crumbs on the Web. With its help, users can see their group Web traces with little effort.

We presented a user study on the impact of showing group Web traces on users' information foraging performance and their group awareness and collaboration. This evaluation is based on the GCP. We found that with the GCP's help, the information foraging performance improved and the group awareness and collaboration was significantly enhanced. These results confirm our prediction of sharing group Web traces, and provide answers to the research questions posed in the introduction. Showing group Web traces consequently improves group members' information foraging performance and also enhances their group awareness and collaboration.

Future work includes extending this study to ascertain the impact of showing group Web traces to users neglected in this work, and a broader deployment of the GCP to explore its impact over longer periods of time and within a larger user population. Future development of the GCP will focus on a better visualization method, a more sophisticated algorithm for identifying passing pages, and perhaps more group collaboration information on the Group Crumb, such as a list of visitors and the time and dwell time of visitations, so long as privacy can be properly protected.

References

AlMurtadha, Y. M., Sulaiman, M. N., Mustapha, N., & Udzir, N. I. (2010): 'Mining Web Navigation Profiles For Recommendation System', *Information Technology Journal*, *9*(4), 790–796.

Amershi, S., & Morris, M. R. (2008) : 'CoSearch: a system for co-located collaborative web search', In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1647–1656).

Anderson, J. R., & Pirolli, P. L. (1984) : 'Spread of activation', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 791–798.

Anupam, V., Freire, J., Kumar, B., & Lieuwen, D. (2000): 'Automating Web navigation with the WebVCR', *Computer Networks*, *33*(1-6), 503-517.

Bernstein, M., Bolter, J. D., Joyce, M., & Mylonas, E. (1991): 'Architectures for volatile hypertext', In *Proceedings of the third annual ACM conference on Hypertext* (pp. 243-260). San Antonio, Texas, United States: ACM. doi:10.1145/122974.122999

Bertel, S. (2001): *Benutzerunterst\ützung im World Wide Web mit Hilfe r\äumlicher Konzepte*, Hamburg, Department for Informatics, Universit \ät Hamburg.

Borges, J., & Levene, M. (2008): 'Mining Users' Web Navigation Patterns and Predicting Their Next Step', *nato security through science series d-information and communication security*, *15*, 45.

Brandenburg, J., Byerly, B., Dobridge, T., Lin, J., Rajan, D., & Roscoe, T. (1998): 'Artefact: a framework for low-overhead Web-based collaborative systems', In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (pp. 189-196). ACM New York, NY, USA.

Brush, A. B., Meyers, B. R., Scott, J., & Venolia, G. (2009): 'Exploring awareness needs and information display preferences between coworkers', In *Proceedings of the 27th international conference on Human factors in computing systems* (pp. 2091-2094). Boston, MA, USA: ACM. doi:10.1145/1518701.1519018

Brusilovsky, P., Chavan, G., & Farzan, R. (2004) : ' Social adaptive navigation support for open corpus electronic textbooks', In *Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 176–189).

Catledge, L. D., & Pitkow, J. E. (1995): 'Characterizing browsing strategies in the World-Wide web', *Computer Networks and ISDN Systems*, *27*(6), 1065-1073. doi:10.1016/0169-7552(95)00043-7

Charnov, E. L. (1976): 'Optimal foraging, the marginal value theorem', *Theoretical population biology*, *9*(2), 129–136.

Cockburn A., & Mckenzie B. (2001): 'What do web users do? An empirical analysis of web use', *International Journal of Human-Computer Studies*, *54*, 903-922.

Deline, R., Czerwinski, M., & Robertson, G. (2005): 'Easing Program Comprehension by Sharing Navigation Data', *Visual Languages and Human-Centric Computing, IEEE Symposium on* (Vol. 0, pp 241-248). Los Alamitos, CA, USA: IEEE Computer Society. doi:http://doi.ieeecomputersociety.org/10.1109/VLHCC.2005.32

Diamadis, E. T., & Polyzos, G. C. (2004): 'Efficient cooperative searching on the Web: system design and evaluation', *International journal of human-computer studies*, *61*(5), 699–724.

Dieberger, A., Dourish, P., Höök, K., Resnick, P., & Wexelblat, A. (2000, November): 'Social navigation: techniques for building more usable systems', *interactions*, *7*, 36–45.

Ebbinghaus, H. (1913): *Memory: A contribution to experimental psychology*. Teachers College, Columbia University.

Fidel, R., Bruce, H., Pejtersen, A. M., Dumais, S., Grudin, J., & Poltrock, S. (2000): 'Collaborative information retrieval', *The New Review of Information Behaviour Research*, *1*(1), 235–247.

Freyne, J., & Smyth, B. (2006): 'Cooperating search communities', In *Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 101–110).

Fu, W., & Pirolli, P. (2007): 'SNIF-ACT: a cognitive model of user navigation on the world wide web', *Hum.-Comput. Interact.*, *22*(4), 355-412.

Georgia, A. D., & Dieberger, A. (1997): 'Supporting Social Navigation on the World Wide Web', Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.8360

Graham, T. C. N. (1997): 'Groupscape: Integrating synchronous groupware and the world wide web', In *Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction* (pp. 547-554). Chapman & Hall, Ltd. London, UK, UK.

Grcar, M., Mladenic, D., & Grobelnik, M. (2005): 'User profiling for interest-focused browsing history', In *Workshop on End User Aspects of the Semantic Web, ESWC-2005, Heraklion*.

Greenberg, S., & Roseman, M. (1996): 'GroupWeb: a WWW browser as real time groupware', In *Conference companion on Human factors in computing systems: common ground* (pp. 271-272). Vancouver, British Columbia, Canada: ACM. doi:10.1145/257089.257317

Hansen, P., & Järvelin, K. (2005): 'Collaborative information retrieval in an information-intensive domain', *Information Processing and Management*, *41*(5), 1101–1119.

Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T. (1992): 'Edit wear and read wear', *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92 (p 3–9). New York, NY, USA: ACM. doi:10.1145/142750.142751

Large, A., Beheshti, J., & Rahman, T. (2002): 'Gender differences in collaborative web searching behavior: an elementary school study', *Information Processing & Management*, *38*(3), 427–443.

Marshall, C. C., & Shipman III, F. M. (1993): 'Searching for the missing link: discovering implicit structure in spatial hypertext', In *Proceedings of the fifth ACM conference on Hypertext* (p. 230).

McFadden, D. (1974): 'Conditional logit analysis of qualitative choice behavior', *Frontiers in econometrics*, *8*, 105–142.

McFadden, D. et al. (1978): 'Modelling the choice of residential location', *Spatial interaction theory and planning models*, *25*, 75–96.

Morris, M. R., & Horvitz, E. (2007): 'SearchTogether: an interface for collaborative web search', In *Proceedings of the 20th annual ACM symposium on User interface software and technology* (p. 12).

Morris, M. R. (2008): ‘A survey of collaborative web search practices’, In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1657-1660). Florence, Italy: ACM. doi:10.1145/1357054.1357312

Morris, M. R., Teevan, J., & Bush, S. (2008): ‘Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting’, In *Proceedings of the ACM 2008 conference on Computer supported cooperative work* (pp. 481-484). San Diego, CA, USA: ACM. doi:10.1145/1460563.1460640

Obendorf, H., Weinreich, H., Herder, E., & Mayer, M. (2007): ‘Web page revisitation revisited: implications of a long-term click-stream study of browser usage’, In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 597-606). San Jose, California, USA: ACM. doi:10.1145/1240624.1240719

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back, M. (2008): ‘Algorithmic mediation for collaborative exploratory search’, In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 315–322).

Pirolli, P. L. T. (2007): *Information Foraging Theory: Adaptive Interaction with Information* (1st ed.). Oxford University Press, USA.

Stanton, N. A., & Baber, C. (1994): ‘The Myth of Navigating in Hypertext: How a" Bandwagon" Has Lost Its Course!’, *Journal of educational multimedia and hypermedia*.

Sun, J., Wang, X., Shen, D., Zeng, H., & Chen, Z. (2006): ‘Mining clickthrough data for collaborative web search’, In *Proceedings of the 15th international conference on World Wide Web* (pp. 947-948). Edinburgh, Scotland: ACM. doi:10.1145/1135777.1135958

Tan, B., Shen, X., & Zhai, C. (2006): ‘Mining long-term search history to improve search accuracy’, In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 718-723). Philadelphia, PA, USA: ACM. doi:10.1145/1150402.1150493

Teevan, J., Dumais, S. T., & Horvitz, E. (2005): ‘Personalizing search via automated analysis of interests and activities’, In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (p. 456).

Ting, I. H., Clark, L., & Kimble, C. (2009): ‘Identifying web navigation behaviour and patterns automatically from clickstream data’, *International Journal of Web Engineering and Technology*, *5*(4), 398–426.

Twidale, M. B., Nichols, D. M., & Paice, C. D. (1997): ‘Browsing is a collaborative process’, *Information Processing & Management*, *33*(6), 761–783.

Weinreich, H., Obendorf, H., Herder, E., & Mayer, M. (2006): ‘Off the beaten tracks: exploring three aspects of web navigation’, In *Proceedings of the 15th international conference on World Wide Web* (pp. 133-142). Edinburgh, Scotland: ACM. doi:10.1145/1135777.1135802

Wexelblat, A., & Maes, P. (1999): ‘Footprints: history-rich tools for information foraging’, In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99 (pp. 270–277). New York, NY, USA: ACM. doi:10.1145/302979.303060

Yeh, P. J., Chen, B. H., Lai, M. C., & Yuan, S. M. (1996): ‘Synchronous navigation control for distance learning on the Web’, *Computer Networks and ISDN Systems*, *28*(7-11), 1207-1218.