Modelling User Behaviour on Page Content and Layout in Recommender Systems

Kin Fun Li and Kosuke Takano

Abstract. With the exponential growth of information on the Web, recommender systems play an important role in many service applications such as e-commerce and e-learning. Recommender systems are used to assist users in navigating the Web or propose items that the users are likely interested in. Most of the currently prevalent approaches use collaborative filtering based on the preference of a group of similar users. In the past decade, there has been some but rather limited research in personalized recommender systems incorporating an individual user's explicit and implicit feedbacks. In our previous work, a personalized recommender system that extracts an individual user's preference and the associated Web browsing behaviour such as print and bookmark, has been designed and implemented. In this chapter, Web browsing behaviour reflecting a user's preference on layout and design is investigated. We postulate that when a user browses a page, her actions on the content and links could be associated with personal preference on an object's location, icon shape, colour scheme, etc. Furthermore, tags and labels of selected objects contain valuable information to facilitate the recommendation process. Consequently, systematic and automatic analysis of the relationship between information preference and Web browsing behaviour based on structure and schema learning could be exploited to complement recommendation utilizing content similarity. Survey and related work on personal recommender systems that model Web browsing behaviour are presented. A proof-of-concept system is designed with the objective to study whether there is a correlation between browsing behaviour, both in the content and visual aspects of a Web page, and user preference.

Kin Fun Li
Department of Electrical and Computer Engineering, University of Victoria, Canada
Tel.: +1 250 721 8683, Fax: +1 250 721 6052
e-mail: kinli@uvic.ca

Kosuke Takano
Department of Information and Computer Sciences, Kanagawa Institute of Technology, Japan
Tel: +81 (0)46 291 3266, Fax: +81 (0)46 242 8490
e-mail: takano@ic.kanagawa-it.ac.jp

M. Biba and F. Xhafa (Eds.): Learning Structure and Schemas from Documents, SCI 375, pp. 289–313.
springerlink.com

1 Introduction

The World Wide Web has become an indispensable resource for people to gather information as the Web provides a quick search capability to its rich data and abundant services. To retrieve desired and relevant information effectively, many techniques have been proposed. In particular, recommender systems have been designed to assist a user in navigating the myriad of information on the Web and suggest items that the user is most likely interested in.

Most of the current recommender systems use the collaborative filtering approach [4] that predicts the interest of a user by analyzing preference information collected from a group of similar users. Collaborative filtering has been widely used in many applications such as e-commerce [35], netnews [33], and hobby-sharing in music, movie, etc. [41].

1.1 Personalization and Browsing Behaviour

Many search engine users have the experience that the returned results of a query are not exactly what they are looking for. Teevan et al. termed the "large gap between how well search engines could perform if they were to tailor results to individuals, and how well they currently perform by returning a single ranked list of results designed to satisfy everyone" as the potential for personalization [40]. With the advance of monitoring and measuring techniques, implicit personal preference information can be collected easily and harnessed effectively in recommender systems. A user's Web browsing behaviours such as dwell time, mouse click, scroll action, and search query, together with site visit history and personal document collection, are often used in usage and content mining to assist in making personal recommendation.

In the "Stuff I've Seen" system [10], personal contextual items, such as authors and thumbnails from the documents that the user has already seen, are used to search for relevant information. The SEARCHY system [28] filters and re-ranks the Web search results by exploiting the user's profile as obtained from her Web browsing behaviour. Morita et al. proposed an information reminder system [29] where a user's action such as printing, copying and pasting, are recorded during a Web browsing session. This user profile is then utilized to provide personalized information to the user. Chirita et al. proposed a personalized query expansion method to retrieve Web information based on the personal collection of text documents, emails, cached Web pages, etc. [8].

In a previous work [39], we proposed an adaptive personalized recommender system using a preference-thesaurus constructed based on Web browsing behaviour and user feedback. This system is personalized for an individual user by capturing her browsing behaviour into a preference-thesaurus. Moreover, the system can adapt to different users as well as their changing behaviour and interest through direct feedback and continuous update to each individual's preference-thesaurus. Explicit user preference information based on user feedbacks and implicit measures such as browsing history are being used in interest prediction and information filtering. The browsing behaviours captured are the ones associated with actions such as bookmark, print, and save. The contents of the pages associated with these actions are analyzed and used to predict future interest.

1.2 Motivation and Objectives

Current personalized information provision systems recommend or navigate to preferable information based on the implicit assumption that a user's preference is strongly correlated to her browsing behaviour. However, to the best of our knowledge, this assumption has only been investigated in a few limited studies. Moreover, recommendation is formulated using only content-based information filter.

Recommender systems based on browsing behaviour have also been used successfully in assistive technology for the mobility or visually impaired [38]. While researching literature in assistive technology for Web browsing, we came upon a recent study conducted by Francisco-Revilla and Crow [13]. They investigated how users interpret the layout of news and shopping pages. Their study reveals that users look for familiar structural elements and use them as references and entry points, before even looking at the main content. Although the target application of their work is in the assistive technology area, this prompts us to postulate that the layout and design of Web pages may also be used as an information filter for establishing user preference.

The objectives and contributions of this chapter are multi-fold. First, an extensive survey on Web browsing behaviour and user behavioural models is presented. Then, previous work on structure and layout of Web pages, though almost all targeted towards the facilitation of Web page design, is discussed. Our proposed recommender system architecture is introduced to reflect how the various modules are integrated. The implementation of the recommender is shown with the important inner working details. Finally, a system design to capture layout and structure information of Web pages, together with how such information can be used for recommendation, are presented.

Specifically, this chapter aims to (1) reinforce the notion that there is a positive correlation between Web browsing behaviour and information preference; (2) strengthen the concept that content-based information filtering is a valid approach for recommendation; (3) promote the novel idea that the layout and design of a Web page is a plausible visual information filter for establishing user preference and thus is useful for recommendation.

2 Literature Review and Survey

User Web browsing behaviour is used in many applications including recommender systems and Web page re-design. Much research work has been done in exploiting information derived from various browsing and navigation behaviour to predict and improve the Web search process. By analyzing a user's Web browsing behaviour, her personal preference can be inferred and utilized in recommending information [2, 5, 11, 36]. For example, if a user spends a considerable amount of time on some Web pages, it is reasonable to regard the user is more interested in the contents of these Web pages than other pages. Also, search engine keywords and results are very important factors to detect personal interest [11, 28]. By analyzing such browsing behaviour information, a user's preference can be established and used to recommend information that suits one's individual taste. For instance, if a user browses some Web sites related to Mozart over a long period of time or repeatedly, it can be inferred that the user prefers classical music and she may also be interested in Beethoven and Bach. Likewise, if a user prints Web articles about digital cameras and MP3 music players, it is highly likely that she is also interested in some related electronic devices such as DVD players and mobile phones.

We postulate that some specific actions performed during Web browsing are positively correlated to a user's preference. Also, specific actions could be correlated to particular interest genres. There are many Web browsing behaviours and it is difficult to identify which ones are influential to a specific user's preference, since each user has her peculiar browsing behaviour that may not be universally held by others. For example, although selecting terms on a Web page by mouse-clicks seems to be an important Web browsing behaviour, however, there are some people who click and select items without much thoughts or intentions. In general, bookmarking is a useful resource but old bookmarks may not reflect a user's current interest. It is therefore important to filter out non-influential browsing behaviours in order to make recommendation. We have carried out an extensive literature review on browsing behaviour and found answers to some of the above questions in our survey.

As stated in Section 1, we are interested in the relationship between a user's preference and the layout and design of a Web page, in addition to the correlation between user interest and browsing actions. Therefore, we view a browsing behaviour consisting of two identifiable components: action and visual. Actions are the browsing interaction a user has with the browser such as bookmark and print. Each action also has a visual aspect related to the layout and design of the page, for example, a user most likely prefers a Q & A type of document if she bookmarks such type of pages frequently.

2.1 Browsing Behaviour: Action

Browsing behaviour actions are the ones that a user interactively enters into a browser including bookmark, print, save, etc., and the derived ones such as dwell time. These are implicit measures for recommendation effectiveness collected during a browsing session as opposed to explicit measures that require users to state their preference or rank a list of items.

Teevan et al. developed a prototype system that makes use of three sources to improve relevance and search personalization: (1) Explicit ratings; (2) implicit click-through behaviour; and (3) implicit content-based measures including information created, copied, or viewed by an individual [40]. Furthermore, they found that implicit behaviour-based measures are useful in capturing relevance while content-based measures are more suitable for capturing an individual's variation.

Similarly, Seo and Zhang learned a user's preference by observing the user's browsing behaviour implicitly [36]. In their system, a user's implicit feedbacks are profiled including time for reading, bookmarking, scrolling, and following up the hyperlinks in a document.

2.1.1 Dwell Time

In one of the early studies in user behaviour and relevance judgment, Morita and Shinoda performed extensive experiments on user behaviour and emphasized that reading time is an important behavioural indicator [30]. Since then, many researchers have established that user browsing time is a major parameter to determine a user's interest of the content [9].

Based on the assumption that the more an object contains the information needed, the longer the viewing time, Liang and Lai [25] presented a time-based approach to determine user interest in news services. In addition, keywords are identified and their position and frequency in the document are analyzed.

Recently, Liu et al. proposed to model the dwell time, the time spent on a document, using the Weibull distribution [27]. They also demonstrated the possibility of predicting dwell time distribution.

2.1.2 Other Actions

Many researchers have revealed that click-through is the second most important browsing behaviour behind dwell time [27]. Meanwhile, Claypool et al. established a strong positive correlation between dwell time and mouse scrolling [9]. On the other hand, Seo and Zhang found in their studies of implicit user feedbacks that bookmarked URL reflects a user's strong opinion of relevance [36].

Kumar and Tomkins performed a large-scale study of user online behaviour based on Yahoo toolbar logs [20]. They developed a taxonomy of pageviews consisting of three high-level classes: content, communication, and search; moreover, they found that the ratios of all online pageviews for the three classes are half, one-third, and one-sixth, respectively.

Multiple tabs in browsers have also been a subject of research study. Viermetz et al. investigated the impact of multiple-tab browsing on Web usage mining and its relevance to business applications [43]. Huang et al. examined the effect of parallel browsing sessions on design implication for Web sites, browsers, and search interface [15].

Weinreich et al. conducted an extensive long-term client-side Web usage study [42]. They discovered that users do not use backtracking in Web navigation as frequent as previously thought. One reason for this is due to the usage of multiple windows and tabs. They concluded that Web designers must consider the limited real estate space provided by the browser. This points to the importance and the effect of the layout and design of a Web page on user's browsing behaviour and experience.

2.1.3 Classification of Browsing Behaviour

Oard and Kim [32] developed a framework that categorizes observable behaviour into broad classes. Objects at different levels of abstraction, such as a term, paragraph or a document, can be examined, retained, referred, or annotated. The *examination* category consists of the actions view, listen and select. The *retention* category has behaviours, such as bookmarking, that indicate possible future use of an object. Activities that relate two objects, such as linking, form the *reference* category. The last category *annotation* consists of actions, such as highlighting, that intentionally add value to an object. The objective of this framework, however, is for modelling information content using observable browsing behaviour.

Kelly and Teevan later added *create* as a fifth broad category of observable behaviour that includes editing and authoring [19]. They reviewed and classified research work on implicit feedback using this framework of observable behaviour. However, they concluded that "what can be observed does not necessarily reflect the user's underlying intention". This assertion agrees with the fact that the visual aspects of browsing behaviour are implicit feedbacks that cannot be observed directly but can only be estimated statistically.

2.1.4 Browsing Behaviour Models

Zheng et al. developed a user interest model based on the following five behaviours: save page, print page, bookmark page, frequency of visit and dwell time on a page [45]. Li and Feng proposed a page interest estimation model based on information found in Web access log, including page size, frequency of access, date of visit and the time spent of each visit [24]. They purposely did not ask for user feedback nor collect any user identifiable information to avoid privacy issues.

Yu and Liu proposed a 'Short-term User Interest Model' for personalized recommendation to accommodate changes in user's interests over time [44]. Using the assertion that a user's interests are related and concentrated in a short period of time, they concluded that Web pages visited are semantically associated. Furthermore, they used a semantic link network to represent these similar pages.

Burklen et al. presented their 'User Centric Walk' algorithm as the basis for modelling browsing behaviour [5]. Their system consists of two models. The Web graph model includes parameters on the structure and the size of the document, while the access behaviour model considers Web page popularity, path length, viewing time, revisiting, link choice, and jump probability.

Sah et al. proposed an architecture to generate dynamic link and personalization using linked data, and the user's browsing strategies [34]. The user strategy model includes search/purposive browsing that looks for specific information, general purpose/explanatory browsing that stemmed from interests, and serendipity/capricious browsing which is undirected browsing.

2.2 Browsing Behaviour: Visual

A Web page's structure includes objects such as text bodies, images, videos, and their associated tags and labels. A Web page's layout includes elements such as background colour, font size, font style, font colour, style sheets, in addition to the locations of objects.

Lerman et al. believed a Web page contains many explicit and implicit structures in the layout and content. They presented an automatic approach for record extraction and segmentation from Web tables [23].

Song et al. asserted that a Web page can be partitioned into several blocks and the importance of those blocks is not equivalent [37]. They found that users do have a consistent view about the importance of blocks on a web page. Using machine learning techniques, they managed to find functions to describe the correlations between page blocks and importance values. Lim et al. described an algorithm for selecting the main content of a Web page automatically [26]. This is done by first segmenting the page into several blocks and then extracting the main content from the important blocks.

Cai et al. argued that traditional link analysis algorithms ignore the fact that a Web page contains multiple semantics [6]. They treated a Web page as a set of blocks and linkages are from blocks to pages rather than from pages to pages.

Layout of a Web page is an important part of the Web site design. Most Web pages are designed using either standardized layout templates or some logical placement based on the nature of the site. Individuals do have their own favourite layouts, therefore, layout is an important factor in capturing user's preference.

Fiala et al. used a component-based XML document format to enable Web contents and adaptive presentations to be automatically adjusted to user's preference [12]. Kawai et al. developed a content fusion system that displays news items in the user's favourite layout format [18].

Lam and Chan proposed a graph mining algorithm to study how and what specific patterns and features of layout can affect advertising click rate [21]. Examining a page's five general areas: header, footer, left sidebar, right sidebar, and body, they investigated how the layout influences click rate, either positively or negatively.

Karreman and Loorback conducted a study to investigate the visual effect of text structure on users' browsing behaviour [17]. Their results showed that users prefer text structured as list than as paragraphs. Moreover, they found that sites with text lists have their pages visited and appreciated more by the users.

2.2.1 Multimedia Objects

In addition to semantic information from the text body, the structure and semantics of images and videos are also useful in the modelling of user browsing behaviour.

Lee et al. proposed a keyword extraction method for videos by analyzing the distance of text blocks to a video [22]. This 'layout distance' is an indication of how relevant a text block is to the video, and thus important keywords can be extracted from the relevant text blocks.

Textual and link information such as labels and tags of images can be obtained easily and exploited in modelling user behaviour. He et al. presented a method to segment a Web page into blocks and obtain textual and link information of images extracted from blocks that contain those images [14].

Song and Lim partitioned a page into blocks to extract important contents [26, 37]; we, however, use the partitioned blocks to extract visual information. Citing the fact that a page can have semantics associated with the different areas within it, we exploit this further in our proposed system [6]. These are elaborated in Section 4.

We concur with Karreman and Loorback's finding that emphasizes the visual aspect (i.e., the structure and layout) of a page, is highly relevant to a user's browsing behaviour. Furthermore, we believe a user's preference on the layout of a page could be useful in recommender systems.

3 A Recommender System Based on Browsing Behaviour

Our proposed recommender system suggests items that match a user's preference in content, layout and design. A user's preference on the two aspects of browsing behaviour (i.e., action and visual) is monitored, extracted, and stored in a preference-thesaurus which is updated continuously. In this section, we describe the overall architecture and implementation details of our recommender system, based on browsing behaviour of actions. Representation and recommendation of the visual aspects of browsing behaviour are presented in the next section.

3.1 Recommender System Architecture

Figure 1 shows the architecture of our system. It consists of three iterative phrases. During the first phase, a user's Web browsing behaviours are monitored and an important term set is extracted for each behaviour. An initial personal preference-thesaurus is constructed based on each behaviour's term set and its term score. In the second phase, Web documents to be recommended are ranked by the similarity between the preference-thesaurus term set and each document. During the final learning phase, the preference-thesaurus is updated based on the user's evaluation feedback on the most recent recommended items.

3.1.1 Web Browsing Behaviour Monitor

The user's Web browsing behaviours, such as the typical ones shown in Table 1, are monitored continuously, and important term set from each behaviour is extracted as shown in Figure 1 (P1). The behaviour term sets and their scores are stored in the

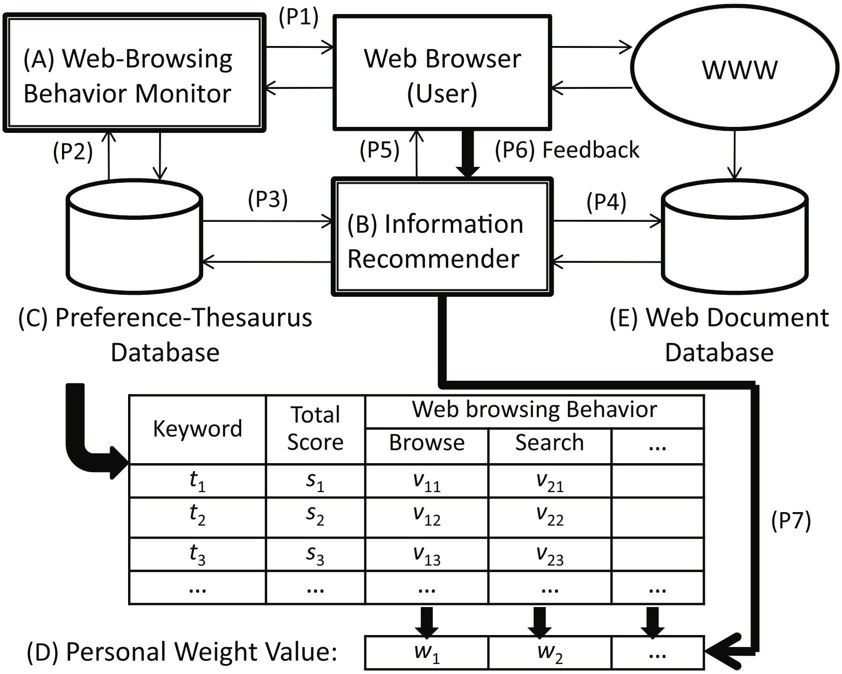

Fig. 1 Recommender Architecture

preference-thesaurus database (P2). For example, for *Web pages browsed* as shown in Table 1, terms appeared on a Web page are regarded as important terms related to a user's preference. The score of each extracted term is the accumulated browsing time of the Web pages that contain the term. For *clipboard copy*, terms copied onto the clipboard are extracted and their scores are the frequency of copies.

3.1.2 Information Recommender

The candidate documents or their URLs for recommendation are stored in the Web document database. These Web documents are collected by the user through various means such as Web crawling, RSS feeds, search engine results, etc. Recommended documents are ranked by calculating the similarity between the preference-thesaurus made up of weighted behaviour term sets and each document in the Web document database (P3, P4), and presented to the user (P5).

3.1.3 Evaluation Feedback

In this phase, the user evaluates whether Web documents recommended are relevant or not (P6). The top-*n* Web browsing behaviours associated with the relevant

Table 1 Typical Web browsing Behaviour

ID	Web browsing behavior	Term set to be extracted
I_1	Web pages browsed	Terms appeared on the Web pages
I_2	Terms on Web pages selected by mouse-click	Terms selected
I_3	Terms on Web pages copied onto the clipboard	Terms copied onto the clipboard
I_4	Keywords searched within Web pages	Search keywords
I_5	Web pages saved	Terms appeared on the saved Web pages
I_6	Web pages printed	Terms appeared on the Web pages printed
I_7	Web pages bookmarked	Terms appeared on the Web pages bookmarked
I_8	Search keywords input to the Web search engines	Search keywords input to the Web search engines
I_9	Web pages browsed from search results	Terms appeared on the returned Web pages browsed

documents as indicated by the user are identified by the similarity between the behaviours' term sets and the relevant documents. Here, n is the number of behaviours which term set's score is greater than zero, and these n behaviours are deemed to be influential on user's preference. The personal weights associated with the top-n behaviours are increased (P7) to reflect the most recent changes in browsing behaviour and preference.

3.2 Recommender Implementation

3.2.1 Extraction of Influential Browsing Behaviour

In evaluation feedback, a user evaluates an item on the recommended list, by navigating to the linked page or giving it a score according to her preference. The recommender then associates the specific item with certain influential browsing behaviours, as shown in Figure 2.

For example, Table 2 shows the ranking of browsing behaviours for three explicit feedbacks: EFB-1, EFB-2, and EFB-3. In the table, each I_x corresponds to a Web-browsing behaviour. For instance, I_5, I_6, I_7, and I_9 refer to *save*, *print*, *bookmark*, and *browsed from search results*, respectively. In this example, the user prefers a music-related document in EFB-1 and EFB-2, and a politics-related document in EFB-3.

For EFB-1 and EFB-2, I_9 (*browsed from search results*), I_6 (*print*), and I_7 (*bookmark*) are identified as the most influential Web-browsing behaviours. Thus, one

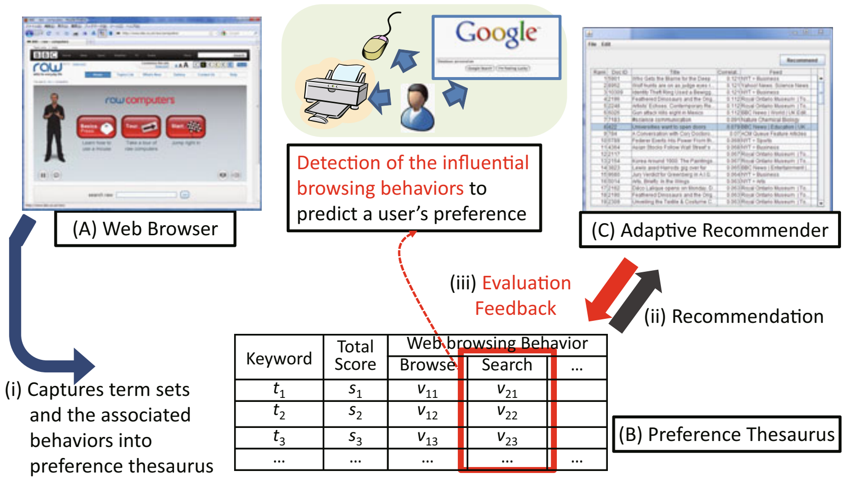

Fig. 2 Extraction of influential browsing behaviour

Table 2 Example of Web-browsing behaviour ranking

Rank	EFB-1 (music)	EFB-2 (music)	EFB-3 (politics)
1	I_9	I_9	$\underline{I_7}$
2	$\underline{I_6}$	$\underline{I_6}$	$\underline{I_6}$
3	$\underline{I_7}$	$\underline{I_7}$	I_5
4	I_2	I_5	I_3
5	I_1	I_1	I_1

can assert that the specific user searches, prints, and bookmarks music related documents regularly. In addition, based on EFB-3, I_7 (*bookmark*), I_6 (*print*), and I_5 (*save*) are deemed to be strong influential behaviours. This user seems to prefer bookmarking, printing, and saving politics related articles. One can also deduce that in general, this user bookmarks her preferred documents.

The most influential browsing behaviours for each individual user can be extracted from the positive items selected from the recommended list via this evaluation feedback mechanism.

3.2.2 Personal Preference-Thesaurus Construction

Typical Web browsing behaviours and their corresponding term sets as shown in Table 1 are extracted by monitoring the user's browsing behaviour and are used to construct a personal preference-thesaurus.

Let a Web browsing behaviour be I_x. Let the term set be T_x that includes the m terms extracted from behaviour I_x.

	I_1	I_2	I_3	$\cdots$	I_x
t_1	s_{11}	s_{21}	s_{31}	$\cdots$	s_{x1}
t_2	s_{12}	s_{22}	s_{32}	$\cdots$	s_{x2}
t_3	s_{13}	s_{23}	s_{33}	$\cdots$	s_{x3}
t_4	s_{14}	s_{24}	s_{34}	$\cdots$	s_{x4}
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$t_{\|T\|}$	$s_{1\|T\|}$	$s_{2\|T\|}$	$s_{3\|T\|}$	c	$s_{x\|T\|}$

Fig. 3 Personal preference-thesaurus matrix

$$T_x = \{t_{x1}, t_{x2}, t_{x3}, \cdots, t_{xm}\} \tag{1}$$

where t_{xi} $(i = 1, 2, \cdots, m)$ is a term included in T_x. The total term set T of all terms appeared in each T_x is represented as follows:

$$T = \bigcup_{i=1}^{x} T_i \tag{2}$$

By using each term included in the term set T and each behaviour I_x, a term-behaviour matrix is created as shown in Figure 3, which is referred to as the *personal preference-thesaurus* matrix. In Figure 3, each element s_{ij} is the score of a term t_j for a behaviour I_i.

Most of the score s_{ij}'s are defined as the frequency of a behaviour. For instance, the behaviour *clipboard copy*'s score s_{ij} indicates how many times a user copied the term t_j to the clipboard. Other scores are expressed in different units such as the browsing time in behaviour *Web page browsed*. Since the units of each behaviour may be different (frequency, time, etc.), the score of each behaviour I_i is normalized in the manner of (5), (8), and (11) as described in the next sections.

3.2.3 Web Documents Recommendation

Typical Web browsing behaviours and their corresponding term sets as shown in Table 1 are extracted by monitoring the user's browsing behaviour and are used to construct a personal preference-thesaurus. The recommendation of Web documents is based on the similarity between the personal preference-thesaurus and each Web document.

First, a document vector space S using the term set T is created. Each document d_i in the Web document set D is represented as a vector $\mathbf{d}_i$ based on term frequencies appeared in the document as follows:

$$\mathbf{d}_i = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{|T|} \end{bmatrix} \tag{3}$$

where, e_k is the term frequency of t_k ($t_k \in T$) in the document d_i, and $|T|$ is the number of terms in the term set T.

Second, in order to realize personal document retrieval for the document set D, personal ranking is performed in the following steps.

Step-1: A query based on user's Web browsing behaviours is created, first by representing each behaviour I_k as a vector $\mathbf{I}_k$. Each element s_{kj} is that behaviour's term score in the personal preference-thesaurus matrix.

$$\mathbf{I}_k = \begin{bmatrix} s_{k1} \\ s_{k2} \\ \vdots \\ s_{k|T|} \end{bmatrix} \tag{4}$$

Step-2: By summing the behaviour vectors $\mathbf{I}_k$'s, a query vector $\mathbf{q}$ is created as follows:

$$\mathbf{q} = \sum_{k=1}^{x} w_k \cdot \frac{\mathbf{I}_k}{|\mathbf{I}_k|}, \quad \sum_{k=1}^{x} w_k = 1 \tag{5}$$

where, x is the number of Web browsing behaviours, w_k is a weighing value for each behaviour I_k and is normalized with 1-norm. Since the relative importance of each behaviour cannot be pre-determined, therefore the initial values of w_k's are set as follows:

$$w_k^{init} = \frac{1}{x} \tag{6}$$

Step-3: The similarity between the query $\mathbf{q}$ and each Web document d in D is calculated. Various similarity measures such as asymmetric measure, Jaccard measure and extended Jaccard measure can be used for this purpose. Here, the commonly used Cosine measure is employed:

$$sim(\mathbf{q}, \mathbf{d}) = \frac{(\mathbf{q} \cdot \mathbf{d})}{|\mathbf{q}||\mathbf{d}|} \tag{7}$$

Then, each document is ranked according to its Cosine similarity score.

3.2.4 Document Evaluation Feedback and Re-recommendation

In (5), it is assumed that each weight w_k of behaviour I_k differs for each person due to individual's Web browsing habit. Therefore, it is necessary to set w_k adaptively based on the characteristic of each user's Web browsing behaviour.

When a user selects a document of her interest from the recommended rank list, the behaviour I_k that strongly affect the similarity score of the selected document can be identified. The weight corresponding to the behaviour I_k is then increased, and a new personal query $\mathbf{q}^{new}$ is formed. This feedback process is performed in the following steps.

Step-1: From the recommended rank list, a user selects a document d_f ($\in D$) which she is interested in.

Step-2: In order to detect the influential behaviour I_k that strongly affects the similarity score between each behaviour I_k and d_f, the following Cosine similarity measure is used:

$$score(\mathbf{I}_k, \mathbf{d}_f) = \frac{(\mathbf{I}_k \cdot \mathbf{d}_f)}{|\mathbf{I}_k||\mathbf{d}_f|} \tag{8}$$

The behaviour I_k is ranked according to the similarity score.

Step-3: The weight of a behaviour I_k whose score in (8) is greater than 0 is increased:

$$w_k = w_k + \alpha_k \tag{9}$$

where, α_k is an incremental value of w_k. Let the rank of the behaviour I_k be r. Each α_k is set according to its rank in Step 2 as follows:

$$\alpha_k = w_k^{init} \cdot \frac{1}{r} = \frac{1}{x \cdot r} \tag{10}$$

Step-4: The new personal query vector $\mathbf{q}^{new}$ is represented as follows:

$$\mathbf{q}^{new} = \sum_{i=1}^{x} w_i^{new} \cdot \frac{\mathbf{I}_i}{|\mathbf{I}_i|}, \ \sum_{i=1}^{x} w_i^{new} = 1 \tag{11}$$

where, w_i^{new} is the weight for each behaviour I_k, and is normalized with 1-norm after Step-3 when (9) and (10) are processed.

Using the new query $\mathbf{q}^{new}$, the ranking process as described in Section 3.2.3 is performed again. The Web documents with the top-n similarity scores are recommended to the user. This feedback is an iterative process so that even if a user changes her information preference and browsing behaviour over time, appropriate recommendation can still be made with the adaptive capability of our system.

4 Structure, Layout, and Schema Learning

4.1 Profiling Layout and Design

Similar to content preference as described in the last section, our system profiles personal preference on the visual aspects (i.e., layout and design) of a Web page continuously by monitoring a user's action browsing behaviour such as printing and bookmarking. Common layout and design attributes are extracted from the Web page, analyzed, and stored in the preference-thesaurus. Attributes of interest include background and foreground font style, size, and colour, shape and colour of icons, link position and colour, video and image position, etc.

4.2 Formalizing Layout and Design

When a user is browsing a Web page W, its layout and design scheme WS is profiled. We define WS as a 4-tuple:

$$WS = \langle A, O, C, S \rangle \tag{12}$$

A is a set of $n \times m$ square-shaped areas $A_{x,y}$ partitioning the page as shown in Figure 4. The numbers n and m depend on, and are adjusted to the width and height of W.

$$A = \{A_{1,1}, A_{2,1}, \cdots, A_{n,1}, A_{1,2}, A_{2,2}, \cdots, A_{n,2}, \cdots, A_{1,m}, A_{2,m}, \cdots A_{n,m}\} \tag{13}$$

The second WS element O is a set of object type ot_p for each object appeared in W.

$$O = \{ot_1, ot_2, \cdots, ot_p\} \tag{14}$$

For example, table, list, link, image, video, icon, background-area with boundary, and so on are elements of object type O. HTML objects such as table ⟨TABLE⟩ $\cdots$ ⟨/TABLE⟩, list ⟨UL⟩$\cdots$ ⟨/UL⟩, link ⟨A⟩$\cdots$ ⟨/A⟩, and image ⟨IMG / ⟩ can be extracted by parsing DOM nodes of a HTML document. Similarly, object node information can be obtained from DOM nodes of an XML document. For icon detection, a small image is recognized as an icon, if its size is smaller than a threshold value.

There are two typical ways to create background-area with boundary: (1) a simple boxed-area using CSS description and (2) a complex-shaped background-area using image files. The first type of bounded area can be detected by parsing HTML tags and their corresponding CSS descriptions. To detect complex-shaped bounded area, some image processing techniques for pattern recognition can be used.

The third element C is a set of colour c_q used in W, and the fourth element S is a set of shape feature s_r for each object,

$$C = \{c_1, c_2, \cdots, c_q\} \tag{15}$$

$$S = \{s_1, s_2, \cdots, s_r\} \tag{16}$$

Among many options, 8-bit colours can be used in the colour set C. In the shape feature set S, basic shapes such as box, rectangular-box, circle, solid line, and dotted line can be used.

4.2.1 Representing Layout and Design

In order to represent how many objects belonging to object type ot_p that are located in area $A_{i,j}$, we define a matrix M_o,

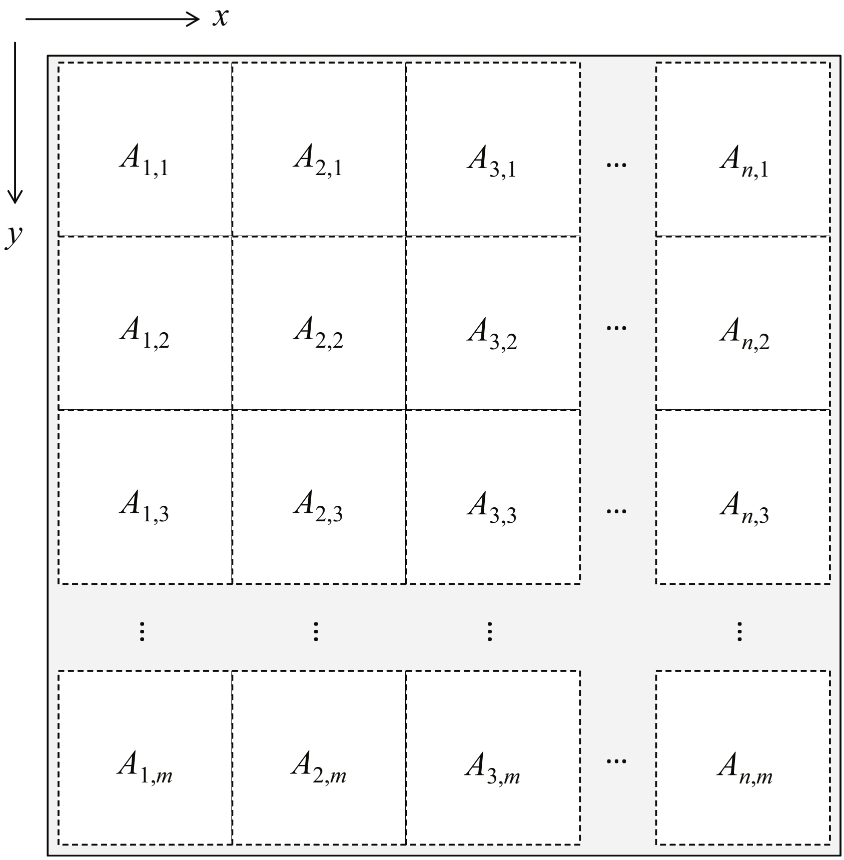

Fig. 4 $n \times m$ square-shaped areas on a partitioned page

	1	2	3	$\cdots$	$N = (n \times m)$
	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$\cdots$	$A_{n,m}$
o_{t1}	f_{11}	f_{21}	f_{31}	$\cdots$	f_{N1}
o_{t2}	f_{12}	f_{22}	f_{32}	$\cdots$	f_{N2}
$\vdots$					
o_{tp}	f_{1p}	f_{2p}	f_{3p}	$\cdots$	f_{Np}

where, f_{xy} is the frequency of ot_p in $A_{i,j}$.

In addition, the colour scheme of objects belonging to ot_p is represented using c_q. We define a matrix M_c as follows:

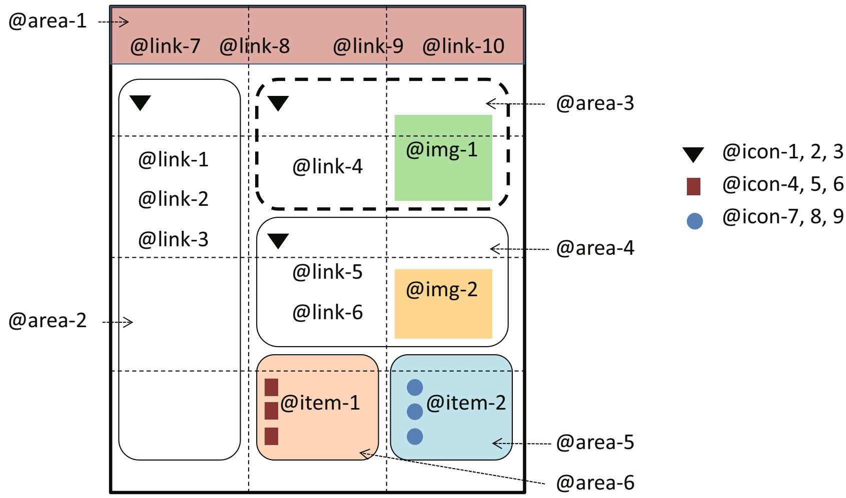

Fig. 5 Example of the layout and design of a Web page

$$
\begin{array}{c|c|c|c|c|c|}
 & c_1 & c_2 & c_3 & \cdots & c_q \\
\hline
o_{t1} & v_{11} & v_{21} & v_{31} & \cdots & v_{N1} \\
\hline
o_{t2} & v_{12} & v_{22} & v_{32} & \cdots & v_{N2} \\
\hline
\vdots & & & & & \\
\hline
o_{tp} & v_{1p} & v_{2p} & v_{3p} & \cdots & v_{Np} \\
\hline
\end{array}
$$

where, v_{xy} is the number of pixels, or the colour histogram using RGB value, for rendering objects belonging to ot_p.

We define a matrix M_s to represent the shape feature of objects belonging to ot_p using s_r as follow:

$$
\begin{array}{c|c|c|c|c|c|}
 & s_1 & s_2 & s_3 & \cdots & s_r \\
\hline
o_{t1} & u_{11} & u_{21} & u_{31} & \cdots & u_{N1} \\
\hline
o_{t2} & u_{12} & u_{22} & u_{32} & \cdots & u_{N2} \\
\hline
\vdots & & & & & \\
\hline
o_{tp} & u_{1p} & u_{2p} & v_{3p} & \cdots & u_{Np} \\
\hline
\end{array}
$$

where u_{xy} is the number of objects belonging to ot_y that has a shape feature s_x.

4.2.2 A Layout and Design Example

Figure 5 shows the layout and design of a Web page. Using this example, the following steps illustrate how the three matrices M_o, M_c, and M_s are generated:

Step-1: The Web page is partitioned into 12 square-shaped areas $A_{1,1}$, $A_{2,1}$, $A_{3,1}$, $A_{1,2}, \cdots, A_{3,4}$, proportional to the page's height and width.

Step-2: Objects are extracted from the Web page. In this example, five types of objects are extracted: 10 hyperlinks (@link), 2 images (@img), 2 items (@item), 6 boundary areas (@area), and 9 icons (@icon).

Step-3: The matrix M_o represents the number of objects within each area. When an object appears in more than one area, it contributes to the frequency count of each of the overlapped areas.

M_o:

	$A_{1,1}$	$A_{2,1}$	$A_{3,1}$	$A_{1,2}$	$\cdots$	$A_{1,4}$	$A_{2,4}$	$A_{3,4}$
@link	2	2	2	0		0	0	0
@img	0	0	1	0		0	0	0
@item	0	0	0	0		0	1	1
@area	2	2	2	1		1	1	1
@icon	1	1	0	0		0	3	3

Step-4: The matrix M_c is generated using the number of pixels, or the colour histogram using RGB value, for rendering each object. For this example, let each object use the colours as follows (in practise, the same type of objects could have different colour attributes):

[@link] black ($=c_1$) : 70px, red ($=c_3$) : 20px, blue ($=c_4$) : 10px
[@img] black: 50px, blue: 30px, green ($=c_5$) : 200px, yellow ($=c_6$) : 150px
[@item] black : 80px, red: 80px
[@area] white ($=c_2$) : 1000px, orange ($=c_7$) : 300px, light blue ($=c_8$) : 300px
[@icon] black : 80px, red: 80px, blue: 80px

M_c:

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	$\cdots$	c_{q-1}	c_q
@link	70	0	20	10	0	0	0	0		0	0
@img	50	0	0	30	200	150	0	0		0	0
@item	80	0	80	0	0	0	0	0		0	0
@area	0	1000	0	0	0	0	300	300		0	0
@icon	0	0	80	80	0	0	0	0		0	0

Step-5: The matrix M_s is created by counting the number of shape features of each object. In this example, let each object use the shape feature as follows (for simplicity, it is assumed that the same type of objects have the same shape features):

[@img] 2 squares ($=s_1$)
[@area] 1 square and 5 rounded-squares ($=s_2$), 1 dotted line ($=s_5$)
[@icon] 3 squares, 3 circles ($=s_3$), and 3 triangles ($=s_4$)
Typically, the hyperlink objects and item objects have no shape feature.

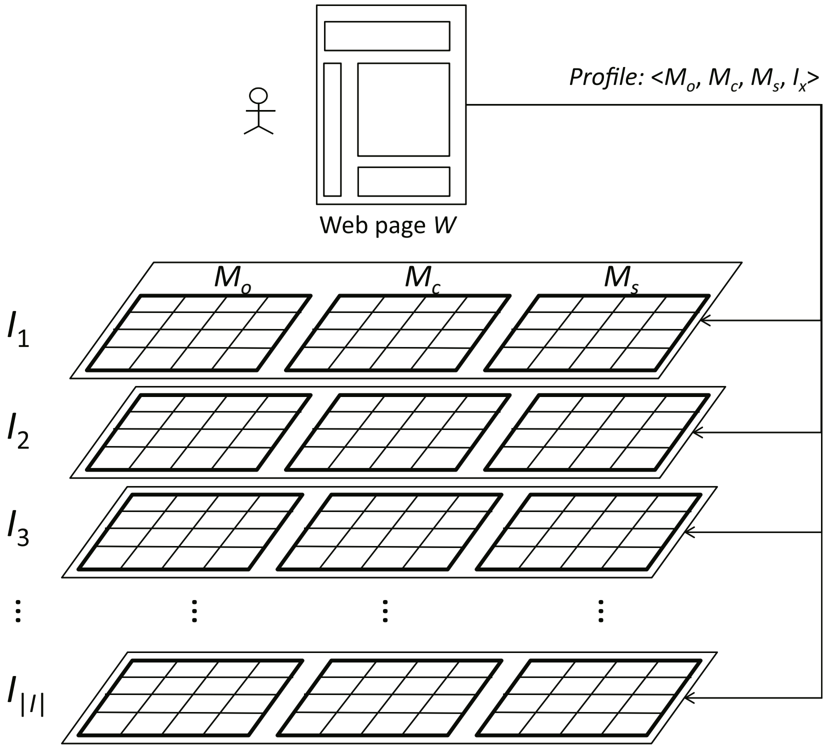

Fig. 6 Learning design and layout associated with browsing behaviour

M_s:

	s_1	s_2	s_3	s_4	s_5	$\cdots$	s_r
@link	0	0	0	0	0		0
@img	2	0	0	0	0		0
@item	0	0	0	0	0		0
@area	1	5	0	0	1		0
@icon	3	0	3	3	0		0

4.3 Layout and Design Learning

When a user exhibits a browsing behaviour I_x on Web page W, the layout and design of Web page W is profiled. The three matrices M_o, M_c, and M_s, representing the visual aspects associated with the browsing behaviour I_x, are created and stored in the preference-thesaurus.

$$profile_{layout\&design}(W, I_x) \longrightarrow \langle M_o, M_c, M_s, I_x \rangle \tag{17}$$

When a user, continuously or in separated sessions, browses Web pages and performs a browsing action or behaviour I_x, the three matrices M_o, M_c, and M_s, associated with the visual aspects of I_x, are updated by summing their current state with the stored state.

$$M_o(I_x) = M_o(I_x) + M_o(I_x)^{cur} \tag{18}$$

$$M_c(I_x) = M_c(I_x) + M_c(I_x)^{cur} \tag{19}$$

$$M_s(I_x) = M_s(I_x) + M_s(I_x)^{cur} \tag{20}$$

where, $M_o(I_x)$, $M_c(I_x)$, and $M_s(I_x)$ are accumulated matrices through profiling, and $M_o(I_x)^{cur}$, $M_c(I_x)^{cur}$, and $M_s(I_x)^{cur}$ are the ones being currently profiled.

In order to represent the entire profile associated with all browsing behaviours I_x ($x = 1, 2, \cdots, |I|$) defined, as shown in Figure 6, the averages of the matrices are used for matching purpose later and are calculated as follows:

$$M_o^* = \frac{\sum_{x=1}^{|I|} M_o(I_x)}{|I|} \tag{21}$$

$$M_c^* = \frac{\sum_{x=1}^{|I|} M_c(I_x)}{|I|} \tag{22}$$

$$M_s^* = \frac{\sum_{x=1}^{|I|} M_s(I_x)}{|I|} \tag{23}$$

Alternatively, the profiling can be obtained by using moving averages to smooth out irregular fluctuations and highlight consistent behaviour, while at the same time taking into account the possible changes in the visual preference of a user. Using moving averages of previous h number of actions or behaviours provides the flexibility of examining behavioural trends simply by changing h. Also, the total number of actions profiled, may be dependent on the need of specific application; thus, it is advantageous to establish a profile using moving averages on a per-action basis.

For example, let matrices profiled at h previous actions be $M_o^h(I_x)$, $M_c^h(I_x)$, and $M_s^h(I_x)$, moving averages of the current action and h previous actions for H-action matrices are calculated as follows:

$$\overline{M}_o^H(I_x) = \frac{\sum_{h=0}^{H} M_o^h(I_x)}{H} \tag{24}$$

$$\overline{M}_c^H(I_x) = \frac{\sum_{h=0}^{H} M_c^h(I_x)}{H} \tag{25}$$

$$\overline{M}_s^H(I_x) = \frac{\sum_{h=0}^{H} M_s^h(I_x)}{H} \tag{26}$$

The accumulated matrices are then represented as:

$$\overline{M}_o^*(H) = \frac{\sum_{x=1}^{|I|} \overline{M}_o^H(I_x)}{|I|} \tag{27}$$

$$\overline{M}_c^*(H) = \frac{\sum_{x=1}^{|I|} \overline{M}_c^H(I_x)}{|I|} \tag{28}$$

$$\overline{M}_s^*(H) = \frac{\sum_{x=1}^{|I|} \overline{M}_s^H(I_x)}{|I|} \tag{29}$$

4.4 Layout and Design Matching

In order to recommend a Web page W with user's preferable layout and design, the similarity between the current visual context $\langle M_o, M_c, M_s \rangle$ of W and the profile $\langle M_o^*, M_c^*, M_s^* \rangle$ has to be calculated. We have chosen a simple Euclidean distance measure for this purpose,

$$\begin{aligned} &Preference(W) \\ &= Similarity(\langle M_o, M_c, M_s \rangle, \langle M_o^*, M_c^*, M_s^* \rangle) \\ &= w_1 \times MatrixSim(M_o, M_o^*) + w_2 \times MatrixSim(M_c, M_c^*) + w_3 \times MatrixSim(M_s, M_s^*) \end{aligned} \tag{30}$$

where,

$$\begin{aligned} &MaxtrixSim(A,B) \\ &= |a_{11} - b_{11}| + |a_{12} - b_{12}| + \cdots + |a_{21} - b_{21}| + |a_{22} - b_{22}| + \cdots + |a_{mn} - b_{mn}| \end{aligned} \tag{31}$$

and w_1, w_2, and w_3 are the weights assigned to the object, colour, and shape preferences that show a personal order of importance.

When there are multiple pages with similar contents, one can use the layout and design similarity for the purpose of recommendation. The $MaxtrixSim$ values of these pages can be used to rank similar-content pages.

If moving averages are employed, the following alternative formula can be used to establish the preference:

$$\begin{aligned} &Preference(W) \\ &= Similarity(\langle M_o, M_c, M_s \rangle, \langle \overline{M}_o^*(H), \overline{M}_c^*(H), \overline{M}_s^*(H) \rangle) \\ &= w_1 \times MatrixSim(M_o, \overline{M}_o^*(H)) \\ &\quad + w_2 \times MatrixSim(M_c, \overline{M}_c^*(H)) \\ &\quad + w_3 \times MatrixSim(M_s, \overline{M}_s^*(H)) \end{aligned} \tag{32}$$

5 Conclusions

In this chapter, we presented a literature survey on Web browsing behaviour, a recommender system based on user browsing behaviour, and the representation and manipulation of attributes associated with the design and layout of Web pages. By modelling and capturing the visual aspects of a Web page, we believe this user preferential information is valuable to complement recommendation utilizing content similarity.

The recommender system that tracks the action aspect of browsing behaviour has been designed and implemented. Preliminary results are positive and warrant further investigation. The design presented here for the visual aspect of a Web page is currently being implemented and incorporated into the recommender.

We plan to have an extensive user study over a long period of time to ascertain the link between information preference and browsing behaviour, and to validate the premise that there is a correlation between user preference and the design and layout of Web pages. This correlation study will be carried out with user experiential interviews and empirical data reviews.

References

1. BBC, `http://www.bbc.co.uk/`
2. Bilenko, M., White, R.W.: Mining the Search Trails of Surfing Crowds: Identifying Relevant Websites from User Activity. In: Proceedings of the 17th International World Wide Web Conference, pp. 51–60 (2008)
3. Brafman, R.I., Domshlak, C., Shimony, S.E.: Qualitative Decision Making in Adaptive Presentation of Structured Information. ACM Transactions on Information Systems TOIS Homepage archive 22(4) (2004)
4. Breese, J.S., Heckerman, D., Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43–52 (1998)
5. Burklen, S., Marron, P.J., Fritsch, S., Rothermel, K.: User Centric Walk: An Integrated Approach for Modeling the Browsing Behavior of Users on the Web. In: Proceedings of the 38th Annual Symposium on Simulation, pp. 149–159 (2005)
6. Cai, D., He, X., Wen, J.-R., Ma, W.-Y.: Block-level Link Analysis. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 440–447 (2004)
7. Chen, Y., Ma, W.-Y., Zhang, H.-J.: Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices. In: Proceedings of the 12th International Conference on World Wide Web, pp. 225–233 (2003)
8. Chirita, P.-A., Firan, C.S., Nejdl, W.: Personalized Query Expansion for the Web. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 7–14 (2007)
9. Claypool, M., Brown, D., Le, P., Waseda, M.: Inferring User Interest. IEEE Internet Computing, 32–39 (November/December 2001)
10. Dumais, S., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R., Robbins, D.C.: Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In: Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 72–79 (2003)

11. Dupret, G., Piwowarski, B.: A User Browsing Model to Predict Search Engine Click Data from Past Observations. In: Proceedings of the 31st annual international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 331–338 (2008)
12. Fiala, Z., Hinz, M., Houben, G.-J., Frasincar, F.: Design and Implementation of Componentbased Adaptive Web Presentations. In: Proceedings of the ACM Symposium on Applied Computing, pp. 1698–1704 (2004)
13. Francisco-Revilla, L., Crow, J.: Interpreting the Layout of Web Pages. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, pp. 157–166 (2009)
14. He, X., Cai, D., Wen, J.-R., Ma, W.-Y., Zhang, H.-J.: Clustering and Searching WWW Images Using Link and Page Layout Analysis. Proceedings of the ACM Transactions on Multimedia Computing, Communications, and Applications 3(2) (2007)
15. Huang, J., White, R.W.: Parallel Browsing Behavior on the Web. In: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, pp. 13–18 (2010)
16. iGoogle, http://www.google.com/ig
17. Karreman, J., Loorbach, N.: Paragraphs or Lists? The Effects of Text Structure on Web Sites. In: Proceedings of the IEEE International Professional Communication Conference IPCC, pp. 1–5 (2007)
18. Kawai, D., Kanjo, D., Tanaka, K.: My Portal Viewer for Content Fusion Based on User's Preferences. In: Proceedings of the IEEE International Conference on Multimedia and Expo., pp. 2163–2166 (2004)
19. Kelly, D., Teevan, J.: Implicit Feedback for Inferring User Preference: A Bibliography. ACM SIGIR Forem 37(2), 18–28 (2003)
20. Kumar, R., Tomkins, A.: Characterization of Online Browsing Behavior. In: Proceedings of the 19th International Conference on World Wide Web, pp. 561–570 (2010)
21. Lam, W.W.M., Chan, K.: Analyzing Web Layout Structures Using Graph Mining. In: IEEE International Conference on Granular Computing, pp. 361–366 (2008)
22. Lee, J., Choi, G., Jang, J., Nang, J.: An Effective Keyword Extraction Method for Videos in Web pages by Analyzing their Layout Structures. In: Proceedings of the IEEE Region 10 Conference TENCON, pp. 1–4 (2007)
23. Lerman, K., Getoor, L., Minton, S., Knoblock, C.: Using the Structure of Web Sites for Automatic Segmentation of Tables. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 119–130 (2004)
24. Li, Y., Feng, B.-Q.: Page Interest Estimation Model Considering User Interest Drift. In: Proceedings of the 4th International Conference on Computer Science & Education, pp. 1893–1896 (2009)
25. Liang, T.-P., Lai, H.-J.: Discovering User Interests from Web Browsing Behavior: An Application to Internet News Services. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences, pp. 2718–2727 (2002)
26. Lim, S.H., Zheng, L., Jin, J., Hou, H., Fan, J., Liu, J.: Automatic Selection of Printworthy Content for Enhanced Web Page Printing Experience. In: Proceedings of the 10th ACM Symposium on Document Engineering, pp. 165–168 (2010)
27. Liu, C., White, R.W., Dumais, S.: Understanding Web Browsing Behaviors Through Weibull Analysis of Dwell Time. In: Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 379–386 (2010)
28. Marcialis, I., Vita, E.D.: SEARCHY: An Agent to Personalize Search Results. In: Proceedings of the 3rd International Conference on Internet and Web Applications and Services, pp. 512–517 (2008)
29. Morita, T., Hidaka, T., Tanaka, A., Kato, Y.: System for Reminding a User of Information Obtained Through a Web Browsing Experience. In: Proceedings of the 16th International World Wide Web Conference, pp. 1327–1328 (2007)

30. Morita, M., Shinoda, Y.: Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In: Proceeding of the 17th ACM SIGIR, pp. 272–281 (1994)
31. MyYahoo!, `http://my.yahoo.com/`
32. Oard, D.W., Kim, J.: Modeling Information Content Using Observable Behavior. In: Proceedings of the ASIST Annual Meeting, vol. 38, pp. 481–488 (2001)
33. Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., Reidl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work Conference, pp. 175–186 (1994)
34. Sah, M., Hall, W., De Roure, D.C.: Dynamic Linking and Personalization on Web. In: Proceedings of the ACM Symposium on Applied Computing, pp. 1404–1410 (2010)
35. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.: Analysis of Recommendation Algorithms for E-Commerce. In: Proceedings of ACM Conference on Electronic Commerce, pp. 158–167 (2000)
36. Seo, Y.-W., Zhang, B.-T.: Learning User's Preferences by Analyzing Web-Browsing Behaviors. In: Proceedings of the 4th International Conference on Autonomous Agents, pp. 381–387 (2000)
37. Song, R., Liu, H., Wen, J.-R., Ma, W.-Y.: Learning Important Models for Web Page Blocks Based on Layout and Content Analysis. Proceedings of the SIGKDD Explorations Newsletter 6(2), 14–23 (2004)
38. Spalteholz, L., Li, K.F., Livingston, N.: KeySurf: A Character Controlled Browser for People with Physical Disabilities. In: Proceedings of the 17th International World Wide Web Conference, pp. 31–39 (2008)
39. Takano, K., Li, K.F.: An Adaptive Personalized Recommender Based on Web-Browsing Behaviour Learning. In: Proceedings of the 2009 IEEE International Symposium on Mining and Web, pp. 654–660 (2009)
40. Teevan, J., Dumais, S.T., Horvitz, E.: Potential for Personalization. Proceeding of the ACM Transactions on Computer-Human Interaction 17(1) (2010)
41. Tso-Sutter, K.H.L., Marinho, L.B., Schmidt-Thieme, L.: Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1995–1999 (2008)
42. Weinreich, H., Obendorf, H., Herder, E., Mayer, M.: Not Quite the Average: An Empirical Study of Web Use. Proceedings of the ACM Transactions on the Web 2(1) (2008)
43. Viermetz, M., Stolz, C., Gedov, V., Skubacz, M.: Relevance and Impact of Tabbed Browsing Behavior on Web Usage Mining. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, pp. 262–269 (2006)
44. Yu, X., Liu, F.: A Short-term User Interest Model for Personalized Recommendation. In: Proceedings of the 2nd IEEE International Conference on Information Management and Engineering, pp. 219–222 (2010)
45. Zheng, L., Cui, S., Yue, D., Zhao, X.: User Interest Modeling based on Browsing Behavior. In: Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering, pp. V5455–V5458 (2010)

Glossary of Terms and Acronyms

CSS (Cascading Style Sheets)
DOM (Document Object Model)
HTML (HyperText Markup Language)
Q & A (Questions and Answers)
RGB (Red Green Blue Colour Model)
URL (Uniform Resource Locator)
XML (Extensible Markup Language)