Browsing and Keyword-Based Profiles: A Cautionary Tale

Michael Shepherd, Carolyn Watters, Jack Duffy and Raj Kaushik
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia, Canada B3H 1W5
{shepherd | watters | kaushik} @cs.dal.ca
jack.duffy@dal.ca

Abstract

In this research, adaptive user profiles are used to rate Web pages with respect to possible user interest. The user profiles consist of weighted keywords and the adaptation is based on the Hebbian Learning Model with direct user feedback. A user study was conducted to determine if the system would “learn” over a series of five session when there was no explicit task other than to browse. The results are analogous to reading the news, i.e., it is not possible to predict what pages a user will read if there is no explicit task. We suggest that a shift away from content to document style (genre) and other user characteristics may be more effective.

1. Introduction

In the context of finding information on the Web, users typically follow a strategy of browsing or searching, alone or in combination. Browsing has been defined as “an exploratory, information seeking strategy that depends on serendipity” [13] and can be characterized by an absence of prior strategic planning. Browsing is often used in new or unfamiliar information spaces, as is typically the case on the Web. Web users rely on mental browsing models based on structures commonly available there: hypertext links and hierarchical directories [17].

Searching and browsing strategies on the Web fit two behavioural theories that have previously been applied to news reading activities: utilitarian or uses and gratification and play or ludenic [22]. The uses and gratification theoretical perspective is based on the assumption that the reader has some underlying goal, outside the Web activity, that the information sought will satisfy [25], i.e., “serves some ulterior purpose external to the communication behaviour itself” [6]. Browsing, on the other hand, is a good example of ludenic or play behaviour as introduced by Stephenson [22]. This behavioural theory asserts that, “The process of news reading is intrinsically pleasurable, and that the intrinsic pleasure is at the root of a mature, orderly, and highly ritualized form of news reading as well as a more casual, spontaneous, and unstructured form of news reading.”

Two characteristics of ludenic reading behaviour are particularly relevant to Web browsing: individual path selection (convergent selectivity) and “knowing when they see it” (apperception). Convergent selectivity refers to the idiosyncrasy of path selection, so that even among the same document set users choose different articles in different orders, based on individual preference. Apperception is a characteristic of human learning wherein people perceive only those aspects of a complex situation that fit within their current interest and/or their current understanding.

Much of the personal use of the Web is based on browsing or “surfing” activities as a strategy to find information “of interest”. In this research we ask whether the findings on user behaviour reading the news is generalizable to behaviour on the Web. In particular, whether the contents of pages previously visited and rated by the user can be used as a predictor of future choices by that same user. That is, if we analyze the contents of the pages the user picked in one browsing session, can we use this information to predict which pages the user is likely to choose in the next session. The ludenic theory predicts that this will not be effective. In this study, we use a Hebbian learning model [3] for continuous revisions of the user profile. In our case, input is based on feedback from the user rating of the retrieved pages. The occurrence of keywords in those documents is used to update a profile of user interest that can be used by the system to rate new pages in terms of probable interest to the user.

---

1 Although ludic is probably a more correct term, the term ludenic was used in the original reference.