Search driven Adaptive Web Content Organization

In recent years, a burgeoning growth in the number of web sites has been witnessed. However presenting useful on demand content on particular web site for the web users has become one of the primary challenges since its not a very straightforward process. In order to to establish the fact that each day content organization is becoming a growing challenge, we try to highlight a number of issues and their inter-relationships. What is the general pattern for reading content on web site? How the content be organized on a page, so that a user can quickly reach to the point where he finds the relevant content? What particular set of keywords or information user search from within the local search facility on particular web page of a website? What is the relationship of searched information with certain period of time and its effects on content? How can these search keywords help understand the user's need for certain information on this particular page? What is the information usage of the search results for a user? Is there any visible context or relationship of hyper-link text with the information in the paragraph to highlight the information that a user is looking for? And finally can we adaptively organize the useful content in some order, based on the finding of a relationship between the user's navigation to particular page on a web site and the frequency of certain searched information. According to Jansen and Pooch, in year 2000, 1 in 28 of all web pages that users viewed were search results pages [4]. In this paper we propose an integrated approach that uses, a web user's reading pattern to read the web content [1] in combination with evaluating semantic appropriateness of hyper links in a web site [2] and eye tracking means to find what user looks for in web search [3]. Using these tools and techniques we propose a method to adaptively organize and present the content in the order of importance in particular display area on web page. The prime benefits include: web content is presented in highly readable areas on demand in the source page where search was initiated, user can find useful information quickly without searching; on searching, by means of long snippets information usage in web search results improves, the information task to find useful information becomes easier.
Summary of Papers:

**F-Shaped Pattern For Reading Web Content:** This paper tries to identify a pattern in which a web user reads the content on the web site. This is accomplished by tracking the eye movement across different areas on a web page. The research conductor has summarized his findings of a study conducted on 232 participants who looked at thousand of web pages. It is stated that that users often read web pages in dominantly F-shaped pattern, where users read in a horizontal movement, across the upper part of the content area, next, down the page a bit in second horizontal movement and finally content's left side in a vertical movement. In our paper, this reading pattern helps us organize our most useful information (e.g hyperlinks) as in an F-shaped pattern allowing user to scan information in few seconds.

**Infoscent Eval:A semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site:**

The author in this paper has presented an evaluation tool that automatically evaluates the semantic appropriateness of the descriptions of hyperlinks in web pages. Using latent semantic analysis, tool takes the textual description of the user's search goal and compares with the textual description of each probable hyperlink. The tool helps in predicting the the path user will follow for all links by running LSA algorithm automatically, indicating a user gets frustrated after length of 3 i.e. number of clicks. Therefore, in our paper, this tool can be used both in the early stages of website design and later when modified to establish the relevance of the hyperlinks on the web page.

**What are you looking for? An Eye-tracking Study of Information Usage in Web Search:** In this paper, the effects of changes in the presentation of search results (comprising; document title, snippet-short descriptive text, URL and some meta data-cached pages) by varying the size of the snippet length (short, medium, long) in web search and how people use web search to find information; have been explored. The results of the experiment suggest that for web searches (informational tasks), long snippets are useful, accurate with improved task times and users opt to look at fewer total items. In our paper, using this observation, when a search is carried out within a website, we compute the time user spends while navigating each search-result-link, & build a heuristic for organizing content on demand.
Hypothesis: The content presented on this web page using adaptive content organization was what a web user was looking for and was found quickly.

To help web users to see the useful and adaptive content that they want to see on particular website, we present our prototype as following: I) Building an index of user's 'searched phrase or keywords' for each web page on the entire website: We store this information to analyze what web users did not find on this particular page. II) Next, we track that which links a user visited from the search results which are displayed on rank basis, then as soon as the user navigates to a particular search result link we compute the time s/he spent while reading information on each link. Having collected this information over certain period of time (administrative decision, and depending on nature of web site) for various users, we store and tabulate the data to establish a relationship of 'most frequently searched information' on a particular page, with the most frequently navigated link with highest time spent on that. III) Next, we propose to maintain a four levels of F-shaped pattern index i.e. TH, MV, MH and LV for top-horizontal, middle vertical, middle horizontal and lower vertical arms of F-shape for content representation. This index is assigned to each hyperlink on a particular web page on dynamic basis. IV) Next, using InfoScent evaluation tool we evaluate each web page (URL) for semantic similarity of all the links' descriptions against 'set of most frequently searched keywords for this web page' as a user goal. Based on the predictive path length, we say, hyperlink for the targeted navigated page must be moved to source page and also its f-shape index for representing the content be set higher and similarly this follows for the rest of results as well. V) Next, depending on the nature and business of web site, a portion or entire page can be dedicated to organize and present the content adaptively based on the f-shaped index of each hyperlink. Advantages of adaptive organization of content: presentation of information in highly readable areas by user search patterns, incorporating user's intent for access to information, to cater for sudden user search information patterns, to minimize time in locating the content, to improve information usage in long snippets of search results, reduced effort in updating the content location on the web site.
**Use of Papers for Integration:** 

**F-Shaped Pattern For Reading Web Content [1]:** The finding of the F-shaped pattern from this paper, helps address what are key areas where highly useful content should be presented on the web page. Since, our prototype addresses the adaptive content organization and presentation on a particular web page, therefore, this idea strongly fits to present the user with most useful information that he's looking for at the top of the screen or distributing the same as per different F-shaped levels. We introduce an F-shape index ranking for the most searched information on particular web page and then identify the key hyperlinks which if not present, should be introduced or alternatively promoted to the higher level in the F-pattern. This proves to be highly useful, because it allows web user to step through the key information presented on top of page very quickly, and in fact this information also reflects the user search intent. **Infoscent Eval:A semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site[2]:** In our prototype, InfoScent tool is used to take the list of 'searched information' one by one as a user's search goal and evaluate if the most frequently target website navigated to a particular user goal (searched information) has a lower LSA index, indicating that the path to navigate to target link for searched information is quite distant (may require multiple clicks). Moreover, we extend this evaluation functionality of a tool by incorporating additional functionality of the F-shape pattern index assignment. This tool based on administrative settings assigns the f-shaped pattern index to each hyperlink, by promoting up or demoting down the hierarchy. **What are you looking for? An Eye-tracking Study of Information Usage in Web Search[3]:** We have integrated search facility local to a website(uses contemporary search engine google with modifications in search result display settings to long snippets) , to facilitate a user and see the 6 to 8 lines of descriptive text under the document title for each search result. Though this takes a few additional seconds to skim the text, but it improves the user's ability to decide for the best link to navigate to. Hence, user would spend more time on relevant link, and chances of information access directly on the page where user initiated search would increase. Hence it reduces the user's valuable time to navigate to multiple web pages around the web site to find that some of the pages were not relevant.
Evaluation: The hypothesis would state that 'the content presented on web page using adaptive content organization influencing search results is what a web user was looking for and was found quickly.' The study design focus was to investigate the organization of the web content adaptively based on the user searched information queries on each web page and evaluating the results and their impact on whether user could find the information he needed. The study is to be conducted across subjects, to make same group perform both tasks on two different data sets in random manner. The variables for this study are, interaction as independent and ease of use as dependent. The metrics for this study include: appropriateness of location of content, time to locate information on web page, access to relevant information in search results, level of frustration in finding information. The study population comprises of 20 participants both male and females who are generally familiar with browsing web sites and they generally have basic idea of searching over internet. For the study, two running websites (simulating for e.g UF Registrar's website containing different services) would be created which uses local search (google) to locate information, where one would use our prototype design to consider the effect of searched information. The process would involve participants to sign IRB and consent form before they undertake the tasks in experiment. A questionnaire would be handed over to participants to answer questions for their feedback. Results would be compiled and based on the findings the null hypothesis would be accepted or rejected and the hypothesis vice versa.

Conclusion: A prototype was presented to understand the effects of searched information on a particular web page to help better organize and present information adaptively in the highly readable areas of the web page from where search was initiated. The proposed approach can have very significant impact for various different websites, especially for commercial, academic and research purposes where the retrieval of the useful web content, organized and neatly presented in the right location is highly mandatory. Imagine thousands of students looking for particular hyperlinks during different time periods (commencement deadline, course registration deadline) and then to some other; in this case, using adaptive content organization can greatly help bring related hypertext links at top.
References:

[1]. Jakob Nielsen, F-Shaped Pattern For Reading Web Content

