Concepts for faster exploratory navigation

Laurian Gridinoc

Grapefruit srl, Romania
laur@grapefruit.ro
http://www.grapefruit.ro/

Abstract. A navigation session across the web is a mix of forward linear traversals, cyclic paths and backtracks —which reflects the user's exploratory navigation in finding relevant information for its particular interest.
The user encounters two situations, both are about assessing the relevance of a linked document —before and after traversal —both involving some degree of cognitive overhead.
This paper puts forward a concept which using semantic web technologies, tries to minimise not only this overhead, but also the overhead in assessing the necessity of backtracking after traversal.

1 Introduction

Finding the right information on the web is a challenging task, as hyperlinks and search engines often do not lead us to the expected information. Navigation involves a series of decisions that are associated to a cognitive overhead characterised by Conklin[1] as "the additional effort and concentration necessary to maintain several tasks or trails at one time".

A widely employed solution to reduce cognitive overhead before the traversal is to improve the substance of the links.

Furthermore, the cognitive overhead in assessing the quality of the document after traversal may be addressed by helping the user identify if and which parts of the document contain relevant information.

The semantic web may make available information about relationships between documents, relationships which may be used in enriching the link information further.

2 Concept

Users may employ arbitrary criteria in assessing the quality of a particular destination, it is possible that the additional link information which is delivered to the user is not discriminatory in the selection of a candidate from the link set, and may induce a degree of information overload.

Considering that the semantic web provides users with information about the relationships between two documents, it is possible that this information

that may be showed to the user may be discriminatory in selecting the link to one document in the context of the other, but not in the full navigation session context, which defines better what the user is looking for.

Moreover, when it comes to the way the web works, it is known that, when users browse, they usually cross the boundaries of multiple authorities —currently, this navigation session context is available only to the user agent and may be to its proxies. This means that any link information setup and delivered from the originating server may not be able to convey the discriminatory information that the user needs, because the originating server cannot be aware of the user's full navigation history.

In order to minimise the cognitive overhead in link selection, it would be better to show to the user if a particular link is in the scope of his navigation session, and if traversing it, it will preserve, narrow or widen the navigation context.

Furthermore, in order to minimize the overhead in assessing the quality of the target document in the current navigational context —upon traversing —the relevant information should be marked somehow, while eliminating the information which is redundant within the navigation session. Otherwise, the user must actively recognize and ignore this type of information.

3 Possible Implementation

From the link decorations employed on the web, the ones that will have minimum cognitive overhead are the ones that do not involve decoding information, like text clues and icons do. If one is restricted to link colourisation (and pitch change in voice browsers), it is not hard to setup a set that will express the relevance of the targeted document in the current browsing context, the minimal set would be: relevant/irrelevant and the desired one would also indicate if the current navigation context will be preserved, narrowed or widened.

Upon traversal, the user tries to assess the quality of the page by scanning —instead of reading as studied by Nielsen[2] —through the page and trying to find relevant information. The signal to noise ratio may be improved in a graphical browser by highlighting information blocks which are not redundant in the current navigational context, and for a voice browser by inserting links that will help the user skip over redundant blocks.

In order to have such a navigational system independent of the user agent capabilities, it should be implemented as a proxy, and should not employ client-side scripting, nor rely only on visual clues.

The context represented by a trail of documents may be computed as a set of document features, each new traversal adding new features or increasing the weights of existing features, while each backtrack will decrease the weights of the features expressed by the rejected document.

The navigation session may have a long stack of documents in its history, but the navigation context should be computed for a smaller window situated at the top of the stack, as users may shift the focus of their exploration.

Features may be directly extracted from documents —from the RDF data the documents may embed or refer directly to. Alternatively, the same features may be extracted indirectly via GRDDL, or by using external data stores that can be queried.

A promising data source is represented by social bookmarking systems which can provide a series of labels, called "tags" for each document, if such information exists. Also, retrieving the set of tags —that is, the features that label the target from such a system —does not involve prefetching the target for analysis by the navigation system.

The redundancy of information should be computed on block level elements, such as paragraphs, by computing a fingerprint, which may be a simple checksum upon the normalized contained text, a signature made by terms relative frequencies, or may employ natural language processing and information extraction techniques.

Redundant areas should be wrapped by a skip navigation helper, such as a skip link that anchors to the end of the redundant area, thus providing a way to scan using a voice browser through possible relevant information on the page. The simplest implementation would then rely only on weights, as well as information from a social bookmarking system, and will act as a proxy which will keep the context feature set. Furthermore, it will also mark each link within the transfered documents accordingly. Finally, the context will store the checksums of each transited block of information in order to determine redundancy and mark it accordingly.

4 Conclusions

The semantic web may provide information that may enrich the navigational experience on the web, but this must be done in a way that will meet the user's needs —expressed by his navigational trail and behaviour —without flooding the user with information specific to particular relationships, but irrelevant to the user's quest.

It is to be determined when and to what extent enriching link information may cause cognitive overhead by information overload and when backtracking is better and faster in reaching the desired information since the user may use arbitrary criteria in assessing the quality of the target document.

Enhancing web interaction with semantic data calls for smarter user agents and smarter personalised proxies that will act as personal agents for the user.

References

1. Conklin, J.: Hypertext: An Introduction and Survey. IEEE Computer (1987), 20(9):17-41
2. Jakob Nielsen: Jakob Nielsen's Alertbox for October 1, 1997. How Users Read on the Web. Available at `http://www.useit.com/alertbox/9710a.html`