Hyperlinks Visualization Using Social Bookmarking

Marek Tomša
Institute of Informatics and Software Engineering
Faculty of Informatics and Information
Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
tomsa@fiit.stuba.sk

Mária Bieliková
Institute of Informatics and Software Engineering
Faculty of Informatics and Information
Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
bielik@fiit.stuba.sk

ABSTRACT

We present a method for navigation support by visualization of actual web page context. We browse and incrementally visualize a graph representing an abstraction of web navigation where nodes represent web pages and edges represent relationships between them expressed either by explicit links (one page linking to another through the content) or implied relationships (relevant pages several clicks away). We proposed several metrics for edge relevance evaluation. In the metrics, existing metadata in form of tags associated with bookmarks offered by collaborative social bookmarking sites is employed and user preferences represented by their tag usage are taken into account.

Categories and Subject Descriptors: H.3.3 [Information Systems]: Information Search and Retrieval; H.5.4 [Information Systems]: Hypertext/Hypermedia—*Navigation*

General Terms: Algorithms, Design, Experimentation

Keywords: Navigation, social bookmarking, link relevance

1. INTRODUCTION

The problem of being "lost in hyperspace" is well known for long time. The loss of orientation occurs, in addition to other causes, as a result of inability to capture the link semantics while following the hypertext links. Furthermore, the links as we know them from the Web are single directional, so users lose their context immediately after following the link. The web pages appear from "nowhere" and disappear somewhere again after reading.

Several methods and tools for navigation support have been proposed recently. Most of them are single purpose solutions employable as a navigational support for specific web sites or sitemap generators, providing support in terms of orientation within given web site (e.g. breadcrumbs, navigational bars, link emphasizing and site maps). In [3] authors introduce a recency-ordered list of session history with integrated bookmarking functionality. An adaptive linear visualization of immediate session history is presented in [4]. While existing approaches are facilitate navigation, they do not offer effective support for inter site navigation.

Brusilovsky in [2] proposes an approach to adaptive social navigation. The social enhancement resides in visualization of frequency of visits of all users. Furthermore, number of visits of the particular user is presented. This approach, as

applied in the Knowledge Sea project, is not applicable for inter-site navigation, but offers a good inspiration concerning utilization of usage data.

2. CONTEXT VISUALIZATION

In general, navigation can be understood as a movement in a space. Web navigation can be defined more specifically as "the activity of following links and browsing web pages" [4], where movement is realized by following the links. The Web as a large network of documents related to each other via hyperlinks can be represented as a directed graph. For the web navigation, we specify the graph where nodes represent web pages and edges represent links between them. Links can be explicit (e.g. `href` tags) or implicit (generated or discovered), which denote especially semantic similarity or other relation between web pages.

Our method is based on graph visualization and its incremental browsing [1]. The user specifies the starting page. In every moment, we visualize only specific part of the hyperspace (a window) where the content is selected according relevance evaluations of the presented entities. In every step, the neighborhood of the current page is generated. Metrics for edge relevance are evaluated and the values are aggregated using weighted sum for every edge. Weights for each particular user are stored in the corresponding user model and are adapted based on the edges the user selects during her navigation.

An outline of the proposed method is presented as follows.

Method 1 Web navigation

1. Specify the node v to be presented (manual select or select the node connected with previously displayed node by a link with the highest importance)

2. Generate a vicinity of v: $GenerateNeighborhood(v)$

3. Add generated entities to the current window w

4. For each edge e in w: $Evaluate(e)$

5. Apply filters and sort: $Sort(Filter(w))$

Function $GenerateNeighborhood$ returns a set of nodes that specify a context of current node. Context is based on acquiring the edges of various types: `href` links extracted from current web page (outgoing links); pages having `href` links pointing to current page obtained using Google search (incoming links), links crawled from specified depth by re-
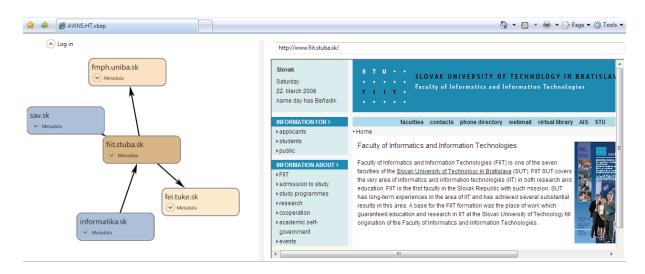
Figure 1: Graphical interface as implemented in the application of adaptive navigation method.

cursively following `href` links. Function *Evaluate* returns an evaluation of every edge of the list provided as an argument by determined context of current node based on weighted sum of defined metrics.

Figure 1 depicts user interface of developed navigator showing the visualization of the context of current web page to the left and a rendered web page on the right. Various types of nodes (href links, similar pages, history pages, etc.) are visualized in different colors.

3. EDGE RELEVANCE METRICS

We proposed several metrics for sorting and filtering the entities in the context of currently displayed node. Some of the metrics employ metadata about web resources expressed in tags stored in collaborative social bookmarking services. We have analyzed several social bookmarking services and their technical limitations concerning offered public APIs. Our current implementation mainly focuses on the largest social bookmarking service – `del.icio.us`.

Relevance based on similarity. This metric is based on an assumption that a tag used by more people is more relevant to determine similarity than tags used by less people. If sets of tags used for two distinct pages are different only in infrequently used tags the similarity remains high. The similarity metric expresses the ratio of common tags to all the tags used to tag both investigated pages.

Relevance based on user preferences. We model user preferences by the tags the users use. We assume that the user uses tags that reflect her preferences. This metric takes into account ratio of intersection of all user's tags and tags used by all users to bookmark the particular page to all tags used by all users to bookmark the particular page.

Relevance based on social trails. We assume that more relevant pages on the Web are those that get more views. The value is determined by the ratio of clicks with the particular page as a target to all clicks made from the source page. Social trails can get various weights depending on the user whose action was logged.

Relevance based on view count. We take into account the overall views of the target page. Pages with more views are rated higher than those viewed fewer times. Compared to previous metric it takes into account sum of all views, whereas the social trail metric evaluates sequences of clicks.

4. CONCLUSIONS

Main contribution of our approach is navigation improvement by the overview of immediate neighborhood of the browsed page based on explicit and implicit relationships. Users are supported in their navigation by evaluating the relevance and recommendations of future browsing trajectory. Compared to existing approaches, the advantage of the proposed method is inter-page navigation support which is not dependent on any navigational aids on the side of browsed hypertext. We consider providing links pointing to the current web page beneficial as these links can be at least as important and relevant as links pointing from the web page.

We performed several experiments with the aim of evaluation of features of the proposed metrics, mainly of those employing tags. Main outcome is showing good applicability of proposed metrics. The advantage of using tags presents at the same time a limitation considering the tags usage on the current Web. However, we can suppose the proportion of tagged web pages will still grow.

5. REFERENCES

[1] M. Bieliková, M. Jemala. Adaptive Incremental Browsing of Ontology Structure. In: Hypertext 2007, Manchester, UK, ACM Press, pp. 143-144.

[2] P. Brusilovsky, R. Rizzo. Using maps and landmarks for navigation between closed and open corpus hyperspace in Web-based education. The New Review of Hypermedia and Multimedia 9, p. 59-82, 2002

[3] S. Kaasten, S. Greenberg. Integrating Back, History and Bookmarks in Web Browsers. In CHI 2001, ACM Press, 2000.

[4] M. Levene, R. Wheeldon. Navigating the World Wide Web. In Web Dynamics - Adapting to Change in Content, Size, Topology and Use, 2004, Springer, pp. 117-151.