

# “Common” Web Paths in a Group Adaptive System

Maria Barra

Delfina Malandrino

Vittorio Scarano

Dipartimento di Informatica ed Applicazioni “R.M. Capocelli”

Università di Salerno

84081 Baronissi (Salerno), Italy

{marbar, delmal, vitsca}@dia.unisa.it

## ABSTRACT

In this paper we describe how we use a group of users’ accesses and interactions with web pages to discover and recommend relevant common navigation paths to other users. We collect data using a social navigation environment called GAS (Group Adaptive System) that we developed [1] and are currently integrating the common path navigation tool into the system. The goal is to use the common path of a subset of users in the system as a recommendation for other users.

## Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia – navigation. H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – Collaborative Computing.

## General Terms

Algorithms, Design.

## Keywords

Adaptive Hypermedia, Recommendation Systems.

## 1. INTRODUCTION

Social navigation in information space examines users behavior and interactions while they are information-seeking. The objective of social navigation is to support the people of a community by taking advantage of the navigation of other users who have similar tastes or interests. Social navigation uses information about other people’s behavior to give suggestions on what to do.

Our view of a social environment is a collaborative and adaptive system on the web named GAS [1] where a group’s interactions are the basis of the adaptation. In GAS, the user’s choices are driven by group adaptation as visualized through an adaptive “common information space” that enhances the awareness of the community’s activities.

In this paper we describe how we use the users’ accesses and interactions to visualize and recommend relevant navigational paths. Our goal is to improve the suggestions provided by the GAS system that are currently based on the behavior of the whole group, by adding information on the common path followed by a subset of users.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT’03, AUGUST 26–30, 2003, NOTTINGHAM, UNITED KINGDOM. Copyright ACM 1-58113-704-4/03/0008.

The comparison of user navigation paths on the web or fragments of visits to websites constitutes an important source of information that can be used for analytical tasks and applications

like user classification and recommendation systems. In [1] the authors explore how comparing web navigation histories may help a person in getting a rough idea of other people’s interests. The interview data showed that people are especially interested in contextual information about colleagues. Of course, this is a motivating factor to our study: people that use to cooperate (on their job) have still much to learn from a comparison of their navigation paths. As the authors observe, viewing a history comparison can produce interesting discussion topics between two people that are familiar, while the web history alone could not provide a sufficient level of detail.

Web servers’ log files have been analyzed and studied to predict surfing patterns. In [3] the authors analyze how to use past surfing behavior to predict future surfing choices in order to solve problems mainly related to the perceived efficiency of the server, such as websites design and usability, as recommend and visualize content or reduce user perceived network latency (content pre-fetching).

## 2. GAS: GROUP ADAPTIVE SYSTEM

GAS offers collaborative and adaptive navigation through (a) a collaborative interface that enables information exchange in the form of documents and comments, (b) a common information space (adapted on participants’ interactions with pages and with others) and the underlying structure of the portions of the web that users interact with (c) a “communication channel” obtained by annotating the pages users navigate. The common information space (showing new resources, new users, new annotations) is the source of information based on group interests and, it is a place from which participants can gather others’ ideas and knowledge. These kinds of tools have been shown to be most useful within a small close group of users, such as a work group, where people already know of each other’s interests. See for example in Swiki [4], a collaborative web server in which every user can easily modify and extend pages, while footprints give the feeling of the activity of a virtual community.

The model we use to adapt web resources with respect to group interests is based on Kleinberg algorithm [5] for discovering authoritative sources of information (authorities and hubs) leveraging on the linked structure of the web and based on the key concept that placing a link from a page toward another page infers authority on a resource that is believed (by the author) to be relevant. GAS is a proxy server infrastructure that lies between a client and the web servers developed using WBI framework [6].

### 3. THE COMMON PATH

In a collaborative and adaptive system like GAS the choices of each user are explicitly “guided” by all the group members since they share and access pages using the common information space computed on the group behavior [1]. It is therefore reasonable to assume that useful information is hidden in the navigation of a subgroup of users that choose to navigate that particular space of information. Of course, the value of information provided to user *Alice* is proportional to her knowledge of the group of users that the Common Path was evaluated from.

For example, to discover that all members’ navigation follow significant common navigational paths, or that relevant paths are characterized only by a subgroup of the participants is useful information that can be applied by the system to enhance awareness of group, and subgroup, common interests. Relevant Common Paths can be recommended by the system to provide knowledge and awareness of others’ activity and can be used by the participants to infer relationships with others. Interviews with users [1] suggest that common interests and preferences can be inferred from navigational path comparisons. Discovering common relevant paths can provide authoritative guides or recommendations based on group choices, but can also be used as a metric for measuring users’ similarities.

#### The Common Path

The definition of a Common Path is not an easy task. The definition as the longest common subsequence (LCS) (as in [7]) looks too restrictive: if Alice navigates pages ( $xxABCDxx$ ) (here  $x$  means any page but  $A, B, C$  or  $D$ ) and Bob navigates pages ( $xAxBxC$ ) their navigation should be considered similar, even though the longest common subsequence only has length 1. The authors of [7] are conscious of this limitation and show that LCS is not particularly useful for similar context, such as predicting web browsing. Moreover, they also suggest that a relaxed version of the LCSs can be useful, by allowing users to take side-trails and coming back to the well-traveled path. Therefore, our definition of Common Path relaxes the LCS definition by using a definition similar to the Levenshtein [8] distance (also called *edit distance*) used (differently) in [3].

In our model the pages navigated by each user over an interval of time (for example a week) is represented as a sequence of nodes indexed by URL. Users’ sequences are gathered by GAS usage; for each user the system keeps his navigation history as nodes on a sequence representing the pages accessed by the user and annotated by the user or by others.

Our parametric algorithm is a well-known example of dynamic programming; it takes as input  $n$  sequences of pages navigated by GAS users over a time interval  $t$  and computes the *longest common subsequence with distance  $d$* . The distance  $d$  is the maximum number of mismatches in URLs allowed between two matches computed on the  $n$  sequences. For example, if  $d = 2$ , the 3 sequences of pages ( $AB$ ) ( $ACB$ ) and ( $AEFB$ ) all contribute to the Common Path ( $AB$ ) while the sequence ( $AEFD$ ) is not. Then a  $d$ -distance Common Path is a path of URLs such that two consecutive web pages are separated, in each single user navigation path, by at most  $d$  pages.

The algorithm first creates an ID table where a unique ID number corresponds to each web page URL. Then, each user is associated with his/her path, as a sequence of integers. The

algorithm randomly picks a pair of users (Alice and Bob) and evaluates all the common subsequences (with distance  $d$ , i.e. allowing at most  $d$ -length side trails) between Alice and Bob. Finally, all the sequences are compared with the navigation paths of each user and, at the end, the longest is chosen.

Since the algorithm just described is computationally expensive, it is not possible to have a real-time evaluation nor evaluate the common paths on all the subsets of users. In the current prototype, we evaluate CP-ALL, the common path that is obtained for each user  $U$ , from all the users but  $U$ , therefore indicating to the user if he was left “behind” by the others. The distance parameter is  $d = 2$ .

### 4. CONCLUSIONS

In this paper we presented the state of an ongoing project, GAS. The steps described here consists in providing recommendations and adapting the navigation links based on the evaluation of the common path of a subset of users. Of course, the choice of users is crucial. As reported, we have already implemented a statically defined subgroup, i.e., “all the users but one”. Our future plans are to give the user the capability to define the subgroup. In this way, each user can discover similarities among the participants in the subgroup, and follow some of the indications. Then, we plan to test the mechanism in a study case in order to prove (a) the efficacy of the Common Paths on real users’ data, (b) the validity of the suggestions for the users.

**Acknowledgments:** the authors thank Elio Diamante Brogna for his help during the implementation.

### 5. REFERENCES

- [1] Barra M., Negro A., Maglio P., Scarano V.: “GAS: Group Adaptive System”. Proc. of AH 2002 Intern. Conf. on Adaptive Hypermedia and Adaptive Web-based Systems. Spain. 2002. <http://isis.dia.unisa.it/projects/GAS>
- [2] Bilezikjian M., Tang J., Begole J., Yankelovich N. “Exploring Web Browser History Comparisons,” Proc. of 2002 ACM Conf. on Human Factors in Computing Systems (CHI 2002), Minneapolis, April 20-25, 2002.
- [3] Pitkow J., Pirolli P. “Distribution of servers’ paths through the WWW: Empirical Characterizations”, World Wide Web Journal, vol. 2 (1999) pag. 29-45.
- [4] Guzdial M. “Swiki” Georgia Tech. <http://www.cc.gatech.edu/mark.guzdial/squeak/>.
- [5] Kleinberg J. “Authoritative sources in a hyperlinked environment”. Journal of the ACM 46(1999).
- [6] Maglio P., Barrett R.: “WebPlaces: Adding people to the Web”. Poster Proc. of 8th Intern. WWW Conference. Canada 1999.
- [7] Tauscher L., Greenberg S. “How people revisit web pages: empirical findings and implications for the design of history systems”. Int. J. Human-Computer Studies (1997), 47, pag. 97-137.
- [8] Levenshtein V.I. “Binary codes capable of correcting deletions, insertions and reversals”. Soviet Phys Dokl. 10,8, 707-710, 1966.