Enhancing meta-portals using dynamic user context personalization techniques

Christos Bouras *, Vassilis Poulopoulos

Research Academic Computer Technology Institute and Computer Engineer and Informatics Department, University of Patras, Rion GR26504, Greece

ARTICLE INFO

ABSTRACT

The Internet is flooded with information and the last decade its size has grown so many times that information search and presentation have become tedious tasks even for experienced users. Minor changes to existing resources can alter the situation and lead to major changes to the end user experience. In this manuscript we present the dynamic web personalization and document grouping infrastructure for meta-portals and the evaluation of our mechanism on a meta-portal. A meta-portal is an informational node where articles from different sources are collected and presented in a categorized and personalized manner. The web personalization mechanism is based on dynamic creation and update of user profiles according to the users preferences when browsing. In parallel a user's profile is affected by user grouping details, which are constructed by users with similar profiles. Assuming that required information, such as article tagging, keywords to categories matching and articles to categories relation is already part of the meta-portal we present a novel mechanism that can build and maintain a user profile which is formed without disturbing the user. Furthermore, we describe the real-time user-centred document grouping mechanism that is implemented to support the web personalization system and present the experimental evaluation of the whole system.

1. Introduction

The last decade can be inevitably referenced as the decade of dramatic changes to almost every aspect of our everyday life. The advances of technology are huge and the evolution of World Wide Web (Internet) can be recognized as enormous. This weird freedom that the Internet offers, attracts more and more people. More attractive is the fact that people are free to produce on-line content in an extremely easy way making thus the production of web content a trend. The Internet is a vast place of article production and it can be referenced without any doubt as a large newsletter. The problem that arises from the fact that the Internet becomes a place where the sources (media) are more than the consumers (readers) is that the customers are usually unable to locate useful information. By useful information we define the information that an user would like to be presented, without being disturbed by any other means of content.

Searching across the Internet through the wide variety of search engines could be a possible solution to the problem of locating information, but the outrageous number of results is uninviting. The search tools that exist within article's sources and the communication channels provided can be presented as a solution or even the ultimate solution; however, the user must "invent" these places before starting to use these services. Creating customized and personalized sections within web pages is another viable solution but some recent examples seem to become misleading for the plethora of different types of users that exist on the web. User personalization and user profiling seem to be the panacea of the current chaotic web status.

User personalization is usually conflicted with the term customization. The difference is vast as the customization refers to the structure and coloring of the web page, while personalization usually refers to the content itself. What we believe is that the user should be able to adapt not only the structure of a web page, but also the content that is presented. Talking about specific content, somebody can assert that the portals are taking measures towards this problem and the content is enriched with an indication about category and lately with tagging on articles. This is sufficient up to an extent but there is still much to be done in order to extend the portals so as to present user centred information. The solution could be found on user profiling and dynamic changes to the user profile according to his habits.

We present a novel mechanism for user profile construction and maintenance in meta-portals. Many worldwide known meta-portals are Yahoo[1] and Google news.[2] We enhance the operation of our meta-portal peRSSonal by providing dynamically changing

[1] http://news.yahoo.com/—news from Yahoo.
[2] http://news.google.com/—news from Google.

user profiling features fully adapted on the user's needs and without need of any user input.

The rest of the paper is structured as follows. The next section presents the related work while the third section our system's architecture. In the fourth section the algorithmic analysis is presented and the following section includes sets of experiments. The manuscript is finalized with future work and concluding remarks on the implemented system.

2. Related work

Many efforts were presented in the latest years in order to provide a solution to the problem of user profiling within web sites or even across the Internet. There is a slight but enormous difference between user profiling (which leads to personalization) and customization of web sites. Customization is the capability that is provided to the user to alter the layout of a web site; which is the color, the font, the position of the elements, the order of the information and others. In the context of the Internet, personalization implies the delivery of dynamic content, such as textual elements, links, advertisement, product recommendations, and more, that are tailored to needs or interests of a particular user or a segment of users (Baraglia and Silvestri, 2007). Personalization techniques (Bouras et al., 2008) are an alternative, user-centric, approach to addressing the problem of information overload. The ultimate goal of any user-adaptive system is to provide users with what they need without them asking for it explicitly (Mulvenna et al., 2000).

Coming to our first statement about the difficulty of search engine usage we investigated research work that is done to the past and it is a great proof that the situation remains almost unchanged through the years. In the majority of the currently existing search engines, when different users submit the same query, the same results are returned in the same order, regardless of who submitted the query. A recent change to Google's search engine result, seems to be misleading as the same user, submitting the same query from different machines is getting different results. Obviously, it is unlikely that all the users of a search engine are so similar in their demands that a sole approach to searching fits all needs. Indeed, in terms of searching, one-half of all retrieved documents have been reported to be irrelevant compared to what the user expected (Casaola, 1998). Additionally, a number of studies have shown that a vast majority of queries to search engines are short and underspecified (Jansen et al., 2000) and different users may have completely different intentions for the same query Lawrence (2000) and Krovetz and Croft (1992).

Some important efforts towards personalization can be found in Zaiane et al. (1998) and Mobasher (2007) where it is obvious that for more than one decade the research community is trying to apply web personalization through data mining activities and generally heuristics while (Anand and Mombasher, 2005) present some of the first more "advanced" techniques of web personalization for the web2.0 that was born back in 2005. The approaches described in Huang (2001) and Srivastava et al. (2000) are of high importance in the research literature on the issue as the first one introduces a cube model for knowledge extraction about the user's behaviour and the second deals with usage patterns from web extracted data.

Kim and Chan (2008) present a robust context for personalization based on UIH which is the user's interest hierarchy that is constructed with the usage of a tree model of the user profile. Other approaches like the ones presented in Sieg et al. (2007) and Garofalakis et al. (2008) that are applying personalized features either on portals or on search procedures by utilizing semantic

information of the user are also interesting as they gather information from meta-data and not only direct information from the user. Evaluation of the user models learned from the data involves the estimation of the accuracy of the models for predicting content that may be interesting to an user as well as other aspects such as explain ability of the recommendations, diversity of the recommendation set, serendipity of the recommendations, and user satisfaction (Herlocker et al., 2004). Finally, it is important to have a reference on the ongoing discussion that is focused on the part of privacy and web personalization. It is a fact that some of the constructed mechanisms are utilizing private information which is obtained without the user's consent. Extended information about the ease of use of privacy and web personalization can be found in Wang and Kobsa (2007) where the formula for reconciling both is presented and analysed.

3. Architecture

The architecture of the system relies on distributed components which form the dynamic web user profiling system. We are putting the focus on the personalized profiling subsystem. We are also doing brief analysis of the other modules in order to cross-connect the features of our complete system, peRSSonal.[3]

The architectural schema consists of a series of subsystems, as depicted in Fig. 1. The collaboration between the distributed parts is based on the open standards (for input data and output data) and on the communication with a centralized database. The general procedure is as follows: at first, web pages are captured and only the useful text is extracted from them. Then, the extracted text is parsed in order to extract keywords and metrics while this procedure is followed by summarization and categorization. Finally we have the presentation of the personalized results to the end user.

3.1. Flow of information

In Fig. 2, we can see the general schema and flow of the advanced and personalized profiling system.

The personalization procedure of the portal that is supported as a medium of communication between all the procedures and the users can be used in order to personalize the summarization on each user. According to the algorithmic procedures of the personalized portal, the system creates a vector that represents the user's profile. To be more precise, each user has two vectors for his profile: a "positive" vector and a "negative" one. The positive vector represents semantically the interests of the user on the article content and the negative represents what is out of user's interest. The vectors are constructed from tables with keyword/value pairs. According to the user's behaviour when browsing the meta-portal the vectors are dynamically altered. The main factors that affect the user's behaviour are depicted in Table 1.

3.2. Document grouping

The system that we are presenting is utilizing the user's behaviour in order to achieve enhanced document grouping. The document grouping procedure of the system leads to creating sets of articles that are identical. By identical, we define the articles that refer to exactly the same fact but have different sources. The document grouping procedure is a never ending procedure because articles occur every 5 min (execution time of

---