A NEW POLICY TO ENHANCE THE WEB RECOMMENDATION SYSTEMS PRIVACY

http://www.computing-conf.org/confman_ac2008/submitted.asp

http://www.computing-conf.org/confman_ac2008/accepted.asp

Prof. Dr. Alaa Al-Hamami
Amman Arab University for Graduate Studies, Zip Code: 11953, P.O.B. 2234, Amman, Jordan.
alaa_hamami@yahoo.com

Prof. Dr. Hilal Hadi Saleh
Computer Science Department, University of Technology, Baghdad, Iraq.
hhsrq@yahoo.com

Dr. Mazin S. Al-Hakeem
Computer Science Department, University of Technology, Baghdad, Iraq.
mazin_ictc@yahoo.com

ABSTRACT:

The variety web recommendation systems uses several methods to observe the behavior of web users and collect comprehensive information about them, then automatically recommend the interest hyperlinks for each individual or group of web users, in order to help users locate information and services on the large website(s).
The processes of active user's behavior observation and understand the active users are concern the web users privacy, many web users dislike giving any personal information. Therefore the privacy adds a whole new dimension to web personalization.
In this paper, a policy has been proposed to identify the interest of active user, in order to recommend the interest hyperlinks, will not capture any information to enhance the privacy of the recommendation system. This policy will ensure that website(s) truly comply with their published privacy practices.

KEYWORDS:

Web Recommendation System, User-Profile, User Interests, Privacy, Fixed-Size Sliding Window.

1. INTRODUCTION

The web recommendation system is the third mode of the web personalization systems (behind the memorization, customization, and task performance support methods), which automatically recommend hyperlinks that are deemed to be relevant to the web user's interests, in order to help users locate information and services on the large website(s) [7].
Generally, the web recommendation techniques observe the behavior of web users and collect comprehensive information about them (e.g., collect information about website address, phone, direct mail, browsing history, and other channels), then automatically recommend the interest hyperlinks for each individual or group of web users [3], figure (1) shown a general overview about how the web personalization works.
The goal behind collect the relevant data about the current users (or which called also active user) is to understand the users, practically to identify the interest of the user, in order to build user profile for each individual user or for group of users.
A key issue in developing web recommendation applications is constructing accurate and comprehensive user profiles based on the collected data.

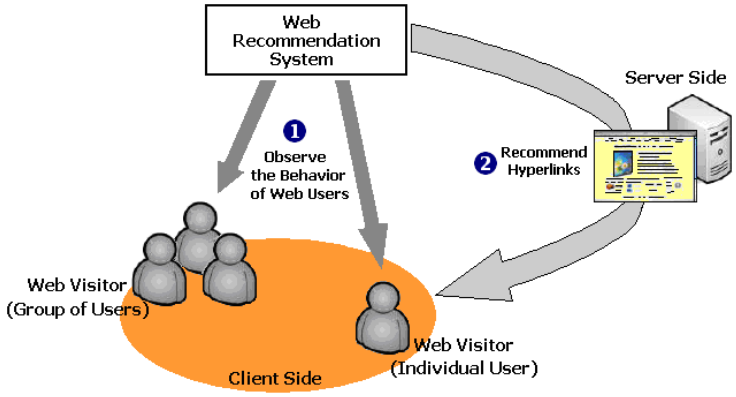

Figure (1): General Overview of Web Recommendation Works [3].

2. WEB PERSONALIZATION PROCESS AND BUILDING USER PROFILE

In general, the Varity web personalization systems (involving web recommendation systems) constitute an iterative process with the following stages:

- Stage One: collecting comprehensive information about the active user via surveys or tracks them.
- Stage Two: Then must be make pre-processing of web data, which includes cleaning data of inconsistencies, filtering out irrelevant information according to the goal of, and completing the missing links.
- Stage Three: Finally converting the pre-processing web data into knowledge stored in user-profiles.

However, the successful web personalization depends on a very large extent on the knowledge about personal preferences and behavior of the users that is usually distilled from the large volumes of granular information about the users and stored in user profiles [4]. However, this stage represents the big technical challenges to the researchers.

There are two types of collection of web data, Explicit and Implicit data :

- Explicit data usually comes from registration forms, rating questionnaires, or cookies.
- While implicit data have several approaches, clickstream data collected approach, past user-activities data collected approach, Data mining approaches, and others.

Table (1) presents the difference between the several implicit data collection approaches.

Table (1): The Difference between the Several Implicit Data Collection Approaches

Clickstream	Past User-Activities	Data Mining (DM)
Exam the user's decisions and length of time spent viewing in each page.	Exam the record in web server logs file or cookies contents.	Mining the web server logs file.

The collected data must be stored in the user profile; such profiles should provide the most relevant information describing who the users are and how they behave. This profile information is usually defined as a record of values and stored in a relational database, usually user-oriented data warehouse, and one record per user [1].

3. WEB USERS PRIVACY CONCERNS

Traditionally, the user profile consists of simple factual information. For example the user's demographics (such as name, gender, date of birth and address), or be derived from the past transactions of a user (such as the largest purchase value made at a website). But the advance user profile captures more complex online behavioral information of active users. It is usually stored in different forms such as association rules, logic-based descriptions, classification rules, attribute-value pairs, etc [8].
In reality, many web users dislike giving away personal information. Some may also be suspicious of web personalization system that relies on the explicit or the implicit web data. In fact, even if an active user agrees to giving up personal information or accepting cookies or enables other collection methods, there is no guarantee that web personalization system will not exchange this information without the user's consent [8].
Recently, the W3C (World Wide Web Consortium) has proposed recommendations for a standard, called Platform for Privacy Preferences (P3P), that enables websites to express their privacy practices in a format that can be retrieved and interpreted by client browsers. However, legal efforts are still needed to ensure that websites truly comply with their published privacy practices [2].

4. IDENTIFY THE INTEREST OF THE CURRENT USER SESSION

The identification of the user interests (in order to get knowledge about the behavior of the active user) is important to provide accuracy recommendation set to the active user.
The current used policies to identify the interest of the user rely on use one of the following methods; cookies, cache busting, and registration. However, each of these methods has serious of drawbacks [6]:

Cookies can be deleted by the user, or the user prevents cookies to created and run on his client device. ***Cache busting*** is practice of prevents software-browser from using stored local version of a page, forcing a new download of a page from the server every time it is viewed. While with the ***Registration***, many users choose not to browse websites that require registrations and logins, or the users may be provide false information when registering.

All the above methods are not sensitive to the changing of user's interest in the same session. The user interests in the web are time dependent; the users have different needs and goals at different times in the same session.

In addition to the above drawbacks, these methods are again due to privacy concerns.

5. THE PROPOSED USER INTEREST IDENTIFICATION POLICY

5.1 The Proposed Policy

In this paper, a new policy will be forced to identify the user interests. This policy involves keeping track of an *interest of user* in the current session by using a fixed-size sliding window over the active session to capture the current user's interest depth.

For example, if the user visit the following webpage; D*1*, D*2*, and D*3* which have interest *sport*, *news*, *news* respectively (the webpage interest can be captured i.e. from the webpage's meta-tags), so the current session, with a window size of 3, is {*sport*, *news*, *news*}, and when the user references the URL D*4*–with *news* interest then the new active session becomes {*news*, *news*, *news*}.

These given indicate that the current interest of the user is *news*, see figure (2).

This important in the context of personalization since most users go backward and forward while browsing a website to find the desired information. For example a user might brows to a particular page using a path before deciding to go back and follow a different path to another content page representing information. In this case, it is not valuable to use references which made in a previous sub-session to make recommendations during the current sub-session. Thus, the sliding window of size *n* over the active session allows only the last *n* visited pages to influence the recommendation set.

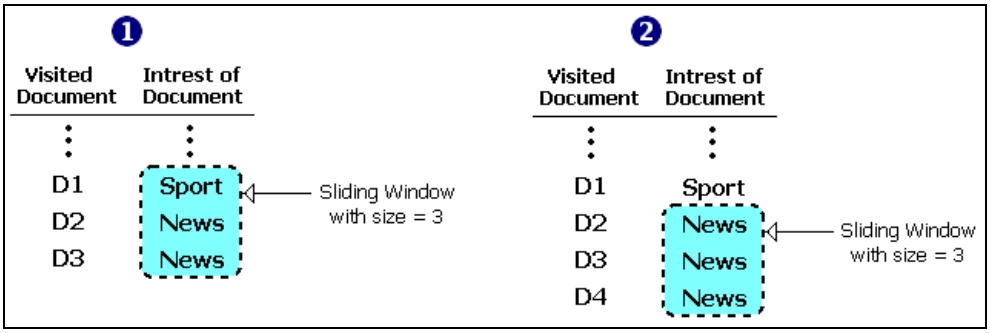

Figure (2): Keeping Track of User-Interest using Fixed-Size Sliding Window.

The window size can be kept as a fixed number (*three* or may be larger than three), figure (3) present the time it takes to make a recommendation set (*which using WWRS recommendation system [5]*) vs. sliding window size.

More than this, since all the interest of visited pages in the current session, with a window size of *n*, are the same; the recommendation set will be generated in the same current user's interest too. This important in the context of user's interest change. For example, if the user visits D5, with different interest, then the new active session will be have different interests, this will be given indicate that the user is began change his interest.

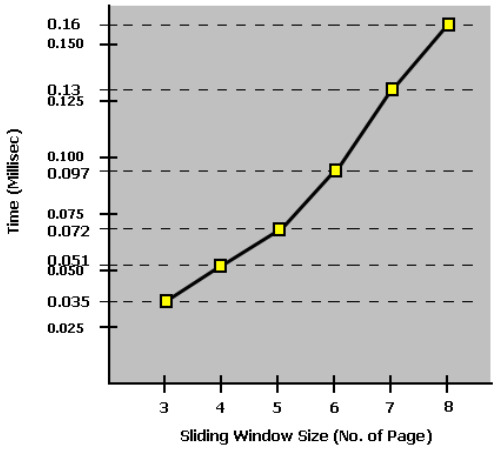

Figure (3): Required Time vs. Sliding Window Size.

The Proposed Policy Algorithm to Identify the User's Interests:

The Input:

User's Request of Webpage.
Empty Session Window with Fixed-Size (equal to 3).

The Output:

User Interest.

The Processing Steps:

Step 1: Get user's request webpage.
Step 2: Get the requested webpage.
Step 3: Extract the meta-tags from requested webpage.
Step 4: Call Identify-Routine (to identify the interest of the requested webpage).
Step 5: Insert the interest of the current webpage into session fixed-size window.
Step 6: If all the interests inside the session window are the same, goto step 8.
Step 7: Goto step 1.
Step 8: Flag the interest (same the interest inside the session window) of the active user at the current time.
Step 9: Call Recommendation-Routine (to generate the recommendation set in the same active user's interest too).
Step 10: If the session is closed, goto 12.
Step 11: Goto step 1.
Step 12: Finish.

5.2 The Proposed Policy Algorithm

The following is the proposed policy algorithm which forced to identify the user interests.

5.3 The Proposed Policy Features

However, we can summaries the proposed policy with the following features:

- The proposed policy is not capturing any information about the users that navigation the website(s), so will not need to build and rely on the user profiles.
- The proposed policy is relying only on the activities of the active user in the current user session.
- The computation of recommendation set is influence only by the user interest at the current time, so the proposed policy is more sensitive to the changing of user's interest.
- The proposed policy is not rely on any of the current used methods, so the proposed policy is not suffer from the drawbacks which associated with using cookies, cache busting, and registration methods.

6. DISCUSSION

Unlike the existing policies, the proposed policy which based on a fixed-size sliding window over the active session, will not record any data about the user, it capture only the interest of the requested webpage(s), in order

to identify the current interest of the active user. So, the recommendation system which based on this policy at his data collection stage will not need to build and rely on the user profile.

There are no user profiles; this means there is no captured information about the users that navigation the website(s). This policy will ensure that website(s) truly comply with their published privacy practices.

The proposed policy is rely only on the activities of the active user in the current user session, so the computation of recommendation set is influence only by the user interest at the current time (which determine according to the current interest inside the fixed-size sliding window), therefore the proposed policy is more sensitive to the changing of user's interest.

The required time to make a recommendation set is increasing along with the slid window size (it could take more time to display a recommendation set to the active user).

7. References:

1. Andrew Rutherford, "Towards Personalized Content", Port Elizabeth Technikon, UK, 2003.

2. Description of the Platform for Privacy Preferences (P3P) Project, Save from support.microsoft.com/kb/290333

3. Formal Web Personalization's website, available at: www.personalization.org.

4. Gediminas Adomavicius and Alexander Tuzhilin, "Personalization Technologies: A Process-Oriented Perspective", Information and Decision Sciences, Carlson School of Management, University of Minnesota, 2005.

5. Mazin S. Al-Hakeem, "A Proposed System to Personalized and Customized Web Content for Mobile Devices", PhD dissertation, Department of Computer Science, University of Technology, Iraq, 2007.

6. Micarelli, A., Gasparetti, F., Sciarrone, F., and Gauch, S.: Personalized Search on the World Wide Web. In: The Adaptive Web: Methods and Strategies of Web Personalization, Brusilovsky, P., Kobsa, A., and Nejdl, W., Eds. Berlin Heidelberg New York: Springer Verlag (2007) this volume.

7. Olfa Nasraoui, "World Wide Web Personalization", Lectures of Department of Computer Engineering and Computer Science, Speed School of Engineering, USA, 2005.

8. Wenning, R., Ed.: The Platform for Privacy Preferences 1.1 (P3P1.1) Specification: W3C Working Draft (2006), http://www.w3.org/TR/P3P11.

IADIS International Conference - APPLIED COMPUTING 2008
Algarve, Portugal, 10-13 April 2008
www.computing-conf.org

http://www.computing-conf.org/confman_ac2008/submitted.asp
http://www.computing-conf.org/confman_ac2008/accepted.asp

Published at: www.uotechnology.edu.iq/dep-cs