A Semantic Web Personalizing Technique

Dimitris Antoniou, Mersini Paschou, Efrosini Sourla, Athanasios Tsakalidis

Department of Computer Engineering and Informatics, University of Patras

The case of bursts in web visits

Introduction

- The need for predicting the users' preferences in order to expedite and improve the browsing though a site can be achieved through personalizing of the websites.
- A personalization mechanism is based on explicit preference declarations by the user and on an iterative process of monitoring the user navigation.
- The problem that we address is the case where few web pages become very popular for short periods of time and are accessed very frequently in a limited temporal space. Our aim is to deal with these bursts of visits and suggest these highly accessed pages to the future users that have common interests.

Personalizing

- Personalization techniques match an individual, his/her preferences and Web page click stream habits with tailored content based on a user profile. In today's world of information overload many similar technologies are used as a way to filter and organize the data most important to them.

***Benefits from Personalizing*:**

- Correctly executed, personalization of the visitor's experience makes his time on a site, or in an application, more productive and engaging.
- Personalization can also be valuable for an organization, a portal or an e-store, because it drives desired business results such as increasing visitor response or promoting customer retention.

***Our Proposal*:**

- We propose a new web personalization technique, based on advanced data structures.

Bursts- Why are they important?

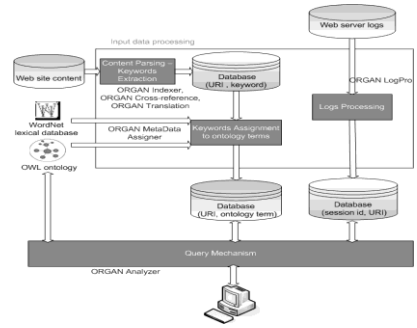
- In this work we try to enhance the case of burst of visits in the personalizing of web sites.
- ***Definition:*** An unexpectedly large number of events occurring within a certain time period is called a burst, suggesting unusual actions or processes.
- Depending on the importance of the phenomenon or the process observed, efficiently detecting bursts is critical.
- Bursts occur in a Website's traffic and affect its functionality in many aspects. A burst may mean:
  1. Expression of interest to specific topic due to recent facts,
  2. Click fraud
  3. Periodical event.

WEBPAGES AND ONTOLOGIES

- Due to the overload of Webpages, dealing with ontologies offer more advantages, since it helps portioning the problem.
- For the ease of the user many websites could organize their Webpages into ontologies in order to aid the search of a WebPage by assigning it to an ontology.
- Our goal is to take advantage of the organization of the WebPages into ontologies and use it in order to deal with the burst of visits to a certain ontology of WebPages.
- **For example,** let us suppose that a user usually visits certain ontology of an e-shop web site and due to a web commercial he visits the ontology of video and sound. At this point, a web personalization technique that deals with bursty visits, would suggest to the user WebPages and ontologies of WebPages that users that prefer the ontology of video and sound visit as well.

Assigning webpages to ontologies

To assign webpages to ontologies, we use the ORGAN Web log analysis tool that offers an integrated solution of building and performing analysis, taking into consideration both the site content semantics and the Web site page visits. *(John Garofalakis, Theodoula Giannakoudi, Evangelos Sakkopoulos, "An Integrated Technique for Web Site Usage Semantic Analysis: the Organ System", Journal of Web Engineering, Vol. 6, No 3, pp. 261-280)*

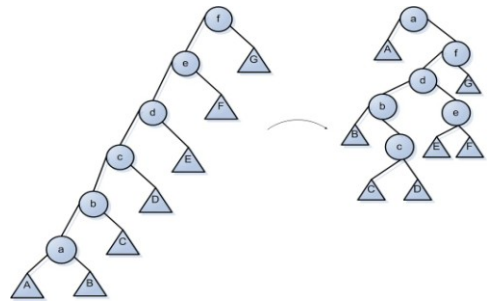


Overview of the algorithm

- We propose a personalization technique, which uses advanced adaptive data structures:
  1. Splay tree and
  2. Binary heaps
- Based on the information gathered from the visits to a website, it suggests to the user other possibly interesting or useful links.
- Our goal is, every time a user clicks on a page that belongs to a certain ontology A, to which we have observed a burst of visits by this user, to suggest a number of most popular ontologies, that most users that visited A prefer to visit as well.

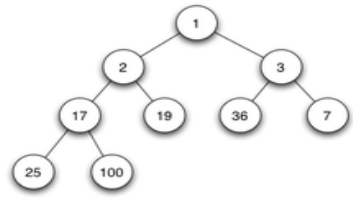
Advanced Data Structures Splay trees

- The splay tree is a self-adjusting form of binary search tree.
- The accessed node is brought to the root of the tree by performing splay operations which consist of rotations bottom up on the access path.



Advanced Data Structures Binary Heaps

- A binary heap is a heap data structure created using a binary tree.
- The binary heap is a special case of the d-ary heap in which each node can have d children (rather than just 2). The property of the binary heap, of which we take advantage of, is the fact that returning a pointer to the node containing the minimum item in the heap, but do not removing it, costs O(1) time.



The problem (1)

- We have a set of P ontologies of WebPages and N users. Each Webpage belongs to certain ontology and each user's profile is kept in a splay tree.
- In the splay tree, we store the ontologies of the Web Pages per user.
- We modify the tree, so that a ontology is splayed (brought to the root) when we observe a burst of visits to it.

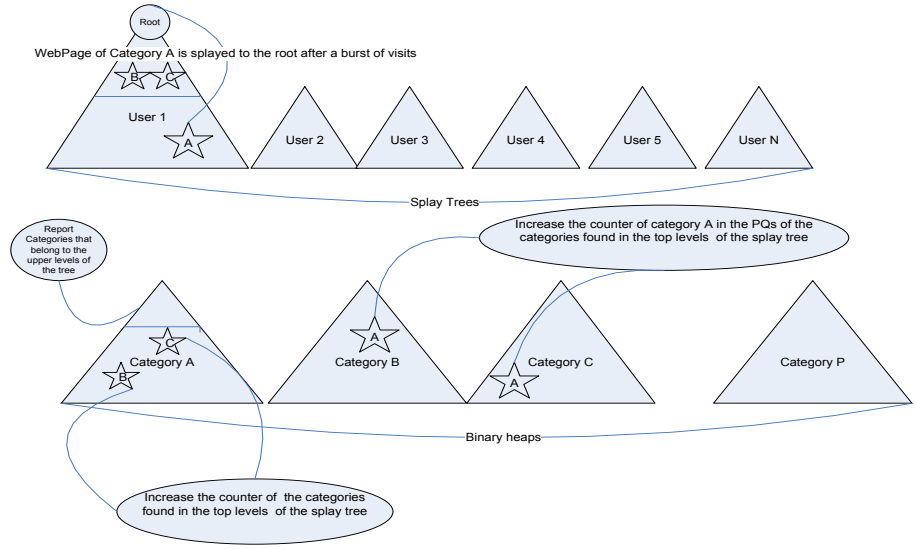
The problem (2)

- For each ontology we build a binary heap where we store the rest of the ontologies.
- Every time we observe a burst of visits to a ontology A:
  1. We increase by one the counter of the popularity of this ontology in all the priority queues of the ontologies found in the top levels of the splay tree of the particular user.
  2. We increase the counters of these ontologies in the priority queue of the ontology A.
- Hence, we are able to extract the most popular ontology to people that visited ontology A in constant time.

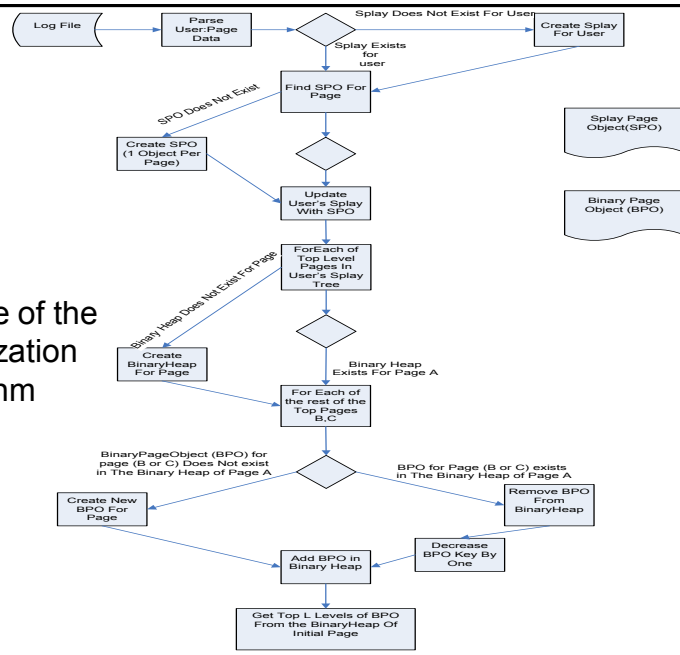
Algorithm in pseudocode

- 1. While (A webpage of ontology W is accessed by a user)
  *\/*Retrieve information from the Webpage's logfile*\/*
- 2. if (this access has created a burst of visits to the ontology W)
- 3. then
  / Rearrange the splay tree of the user, so that the ontology with the latest bursty access pattern is brought to the root *\/*
- 4. Splay the ontology W to the root of the user's splay tree
  / Update the priority queues of the ontologies in order to promote the popular ontologies of the user *\/*
- 5. Define the set of ontologies ,TOP, that already exist in the top levels of the splay tree
- 6. Increase the counter of W in all the priority queues of the ontologies that belong to TOP
- 7. Increase the counter of the ontologies that belong to TOP in the priority queue of W
- 8. Return as suggested the root of the priority queue of W
- 9. endif
- 10. else
- 11. continue

Execution of the algorithm



Architecture of the Personalization Algorithm



ANALYSIS
Space requirements

- As far as the space complexity is concerned, the space, as expected, is mainly consumed by the two data structures.
- Splay trees: We need 1 splay tree for each user. In worst case, in each splay tree we store W WebPages. Therefore, taking under consideration the extra fields needed for each node of the splay tree, the space required is 5·N·W.
- Priority queues: For each ontology we keep a priority queue. Hence, for P ontologies, we need $O(P^2)$ space.

ANALYSIS
Time requirements

- As far as time complexity is concerned, per access we need:

1. $\tau \cdot \log(\frac{\tilde{V}}{\tilde{1}})$ in order to splay τ nodes. That is at most τ·log(#pages).
2. We need O(1) time to return the root from each priority queue. Hence, N· O(1) time is required to suggest a ontology to N users.
3. We need to update the priority queues. In other words, before suggesting the root of a ontology's priority queue, we have to increase the keys of the ontologies that we find in the user's favorites if they are already in the priority queue of the splayed ontology.
4. Finally, we have to increase the key of the ontology splayed, in all the priority queues of the top levels of the splay tree of the user.

- So, in total, k·logP time is needed.

Future work

- Future steps include the evolution of the algorithms to take into account additional implicit user feedback of the final products chosen and not only the e-shops and services.
- This is particularly efficient case for e-businesses implementations based on lightweight RESTful mobile Web Services.

A Semantic Web Personalizing Technique

- Thank you. Questions?"
- Contact: antonid@ceid.upatras.gr