Decentralised User Modelling and Personalisation

Márius ŠAJGALÍK*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
sajgalik@stuba.sk

In our work we focus on a Web browser as a basic tool of a user. The goal is to get closer to the user and enable personalization of the Web content directly in the browser to improve and simplify users' work. That is the main difference compared to other solutions and also an advantage, because a user has all the data and she decides whom to trust and allow an access to it. It is not only the decentralized and distributed nature of the solution that differs, but the direct integration into the browser in the form of an extension. In our solution we have access to such key information, which remote servers have not (e.g., the user's detailed Web browsing history, user activity itself like working with browser tabs or mouse movements).

Our work can be divided into two basic logical units. The first is the modelling of user interests based on the browsing history and user's activities within a browser. This part can also be seen as a kind of intermediate stage of pre-computation, which is later used in the second part –personalisation. As a distributed and decentralised solution, it enables personalisation directly on the client device with the possibility of communication with other users, making use of knowledge and experience of more users to achieve even better personalisation.

Proof that modelling the user and carrying out the personalisation directly in the browser is not just possible, but even richer on captured data can be induced from [1]. The purpose of user modelling is to capture such user characteristics, which can be used further on to customize the browsed Web content. In the core of our implementation and design we focus on modelling the user interests based on the current web browsing history, which can be later (within personalisation subsystem) even more extended and customised.

The process of interests modelling is partly similar to indexing pages on the Web. Indexing involves the creation of an index of all pages that a crawler has found on the Web. In our case we have in hands a kind of an intelligent crawler, which does not visit all pages on the Web, but only those that are of interest to the user. In fact, it is the user

* Supervisor: Dr. Michal Barla, Institute of Informatics and Software Engineering

Spring 2012 PeWe Workshop, April 1, 2012, pp. 41–42.

herself who is in the role of crawler, who visits these pages and chooses them according to her own personal interests. Since we extract weighted terms representing user interests from the content of visited pages (evaluation in [2] shows it can be done satisfyingly accurately), we can say that our indexer produces an index of user interests instead of a regular index of any keywords and terms. Interesting part in our model of user interests is the ability of identifying also the local interests of user within some specific domain.

The second part of our solution deals with the personalization of web pages in the browser. Based on the user interest model we personalise the web content to the user so that her work in the browser is easier, more intuitive and efficient. The solution is not limited to one user, but uses the entire network of users and their relationships, thereby increasing the quality of the personalisation. Using communication, users can help each other, share their experiences and collaborate to recommend the common interests of what they are most interested in.

The realisation of our work represents a decentralised distributed collaborative platform for personalisation of user interests in the Web browser. Physically, it is composed of multiple instances of the browser extension of individual users. These extensions communicate with each other, allowing multiple users to collaborate. Since the main goal is to enable personalisation of content on the Web, this platform provides an interface for access to user interest model, which can be further extended.

The real personalisation is done via personalisation extensions, which are basically executed when a web page is loaded and which can modify its content and thus personalise it to the user. These extensions are basically pieces of JavaScript code where besides the basic possibilities of this language also other functionality is available, like built-in support of jQuery framework and external JavaScript files inclusion, support of cross-origin requests, access to the user's browsing history and persistent database API, personalisation API and communication API. Database API provides a unified interface for database access and personalisation API provides access to information about user stored in user model. Communication API enables communication among users and has a form of channelled multicast in which users do not communicate directly with each other, but they use the communication channel to send the messages. Its purpose is to bring together a group of similar users based on their common interests.

Acknowledgement. This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

References

[1] Ohmura, H., Kitasuka, T., Aritsugi, M.: A Web Browsing Behavior Recording System. In: Lecture Notes in Computer Science: Knowledge-Based and Intelligent Information and Engineering Systems, Springer, Berlin, (2011), vol. 6884, pp. 53-62.

[2] Matthijs, N., Radlinski, F.: Personalizing web search using long term browsing history. In: WSDM '11: Proc. of the fourth ACM international conference on Web search and data mining, ACM Press, (2011), pp. 25-34.