Towards Social-based User Modeling and Personalization

Michal Barla*
Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 3, 842 16 Bratislava, Slovakia
barla@fiit.stuba.sk

Abstract

As the Web grows and the amount of available information increases, new problems such as information overload or lost-in-hyperspace problem emerge. The solution is to shift from "one-size-fits-all" approach and to provide personalized surfing experience, which would take into account differences among users. These differences are captured in a user model, a structure holding all relevant user-related information. We present contributions in both data collection and processing stages of user modeling process, focusing on open corpus web-based systems, where the content can be dynamically added or changed and we have no content available in the design stage of the web-based system. We introduce two methods falling within the scope of data collection stage, a method for *comprehensive logging of user activity on the Web with preserved semantics*, which combines client side and server side logging into a stream of user events with clearly defined meaning, and a method for *capturing logs of "wild" Web surfing* based on a specialized proxy sever. The second group of methods, devoted to actual user model creation within an open corpus environment consists of a method for *user model inference based on rules expressing navigational patterns*, a method for *term-based open corpus user modeling*, which can be applied to capture user's interest across the third-party web-sites and web-based systems. We proposed also a method for *finding relations between terms based on folksonomies*, which supports our term-based user modeling approach. The proposed methods were evaluated by means of software tools that were incorporated in research projects aimed at intelligent information presentation.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*User issues, navigation*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering, Query*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*User profiles and alert services*; D.2.13 [**Software Engineering**]: Reusable Software—*Reuse models*; I.7.5 [**Document and Text Processing**]: Document Capture—*Document analysis*

Keywords

user modeling, adaptive web-based systems, social web, proxy server

1. Introduction

There is no doubt that the Internet and its most-popular service World Wide Web have changed our everyday lives. Especially the Web has become an ubiquitous source of information, a communication tool, shopping mall etc. Ubiquity of the Web along with the rise of user-generated content of Web 2.0 result in an incredible growth of the amount of available information, which bring new problems such as information overload or lost-in-hyperspace problem. These are addressed by *personalization*, which takes into account individual characteristics of an individual to tailor the presentation, recommend navigation or adjust the functionality of the web-based system.

If we want to be able to provide a user with a personalized experience within our web site, we need to base our personalization on some type of user profile, an instance of a user model. The model can represent various features of the user as an individual such as demographic information (age, education level, interests, preferences, knowledge etc.) or can represent the overall context of user's work, including platform, bandwidth or location. Recently, the Web and its research got fascinated by a social phenomenon. Not only that people are more present on the Web, as they got the tools such as wikis, blogs, social tagging systems or social applications, which allows them to actively participate on the creation of web content, but we discovered the power of masses, wisdom of crowds, which can be used to organize the web content and to provide personalized recommendations using social relationships and group membership. The key are *communities*, either having real-life back-end or pure virtual

*Recommended by thesis supervisor:
Prof. Mária Bieliková
Defended at Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava on February 10, 2011.

Barla, M. Towards Social-based User Modeling and Personalization. Information Sciences and Technologies Bulletin of the ACM Slovakia, Vol. 3, No. 1 (2011) 52-60

ad-hoc communities of people sharing a common property at the given time [8]. While usage driven adaptation, mainly in form or statistics-based recommendations is becoming very popular way of navigation, especially in news portals (e.g., simple recommendation of the most visited content within a time window), other forms of social recommendations, which would take into account various aspects of a *personalized* user community are still rather seldom. The challenge is how to represent, build and use a user model for a dynamically changing, often user-generated content.

In this work, we present a contribution in the field of user modeling for adaptive web-based systems, focusing on a shift from static, closed-corpus application domain into dynamic, vast and thus open-corpus application domains such as the Web is. We also target on enrichment of the user model as well as the user modeling process by social-related information. This can be shrink down into two major goals:

- proposal of novel approaches to collection of data about user, which are independent of the chosen domain, separate thoroughly the data collection from the subsequent inference of higher level characteristics and are at the same time unobtrusive from the user's perspective.
- proposal of novel approaches to user model representation and inference, which respect the openness and dynamics of the information space. We aim at investigating the potential of navigational patterns in the dynamic information spaces as the means of navigation are rather constant even if the content is changing dynamically. We concern to explore such representations of domain and user models, which would allow for automatic and unobtrusive acquisition of these models in the vast and dynamic environment and provide a suitable basis for subsequent personalization of information space exploration.

2. User Modeling

User model represents all kind of information related to user and user's context, which are required in order to provide personalized user experience. It can hold various features of the user (user characteristics) as an individual such as demographic information (age, education level, interests, preferences, knowledge etc.), or can represent the overall context of the user's work, including platform, bandwidth or location.

Current adaptive web-based systems use mostly the following two types of user models [6, 5]:

- *stereotype group model* – is the simplest form of a user model, which maps the individual user onto one of (usually) predefined groups. All members of the group would experience the same "personalized" system. The most simple form of a stereotype model is a *scalar model* [10], expressing only one user characteristic (usually level of knowledge) by a single value on a particular scale (e.g., good, average, none).
- *overlay individual model* – is currently the most popular user model representation, where the model reflects user characteristics related to individual parts of a domain model – it adds a layer with user-related information on the top of the domain model for each user.

The advantage of *stereotype group model* is its ease of use and relatively simple initialization. The user can be assigned a stereotype (group membership) according to answers of few questions or eventually according to the initial behavior in the system. Disadvantage of *stereotype group model* lies in limited personalization possibilities. In fact, there is no personalization at all, as we can only speak about an adaptation when the system does not consider the user as an individual, does not build and maintain a user model and is based purely on rather coarse-grained stereotypes (which is often the case if we define stereotypes manually). This results in "one-size-fits-all" problem of personalization for all users within one stereotype.

An *overlay individual model* deals with this kind of a problem, when a personalized system creates a separate user model for each user. The challenge of overlay user modeling is in the complicated acquisition of user model. While in case of a model based on stereotypes, the only required task is to assign a stereotype to the user, overlay user modeling must discover and assign a user-related information to each element of a domain model. When considering maintenance of user models, it is easier to detect and react to a change in one stereotype compared to a change in one of many user characteristics related to particular domain elements. Another problem of *overlay model* lies, which we need to deal with, is its increased demand for system resources as the number of user increase.

The solution of the aforementioned issues lies in employment of specialized approaches, which are a must in case of a large or even open information spaces. One way is to create a more loose coupling of user and domain models, where characteristics relate to types and/or attributes of concepts, not the concepts themselves. Another solution could be the use of different strategies for user model acquisition, which would use distinct layers within the overlay user model. Like that, we can create one overlay only in such parts of the information space which were already visited by the user and use information from this layer to set or update information in another layer covering the rest of the information space [23].

The latter mentioned approach shows us a very important aspect of layering furthermore the overlay user model. Each user model consists of at least an *evidence layer* (called also usage data layer [7]) holding facts about user interactions with the system. Even if this layer does not hold any user characteristics, it can be used as a basis for personalization. Next, we can add additional layers representing *higher level characteristics* coming from different sources or inferred by employing different strategies with various reliability and accuracy.

2.1 User Modeling Process

User Model life cycle is defined by a sequence of actions performed by an adaptive web-based system during user's interaction. The process is depicted in the Fig. 1.

We separate the user modeling process into three distinct stages, as defined in [6]:

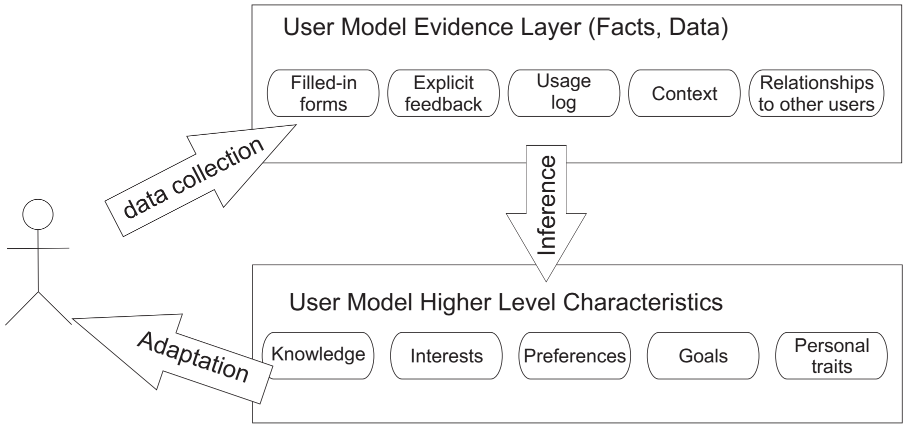

Figure 1: User model life cycle derived from [6].

1. *data collection*, when an adaptive system collects various user-related data such as forms and questionnaires filled-in by the user, logs of user activity within the system, feedback given by the user to a particular domain item, information about user's context (device, browser, bandwidth, etc.). All these information represent observable facts and are stored in an evidence layer of the user model. The facts can be supplied also from external sources, for instance we can acquire information about user relationships with other users from an external application for social networking.

2. *user model inference*, when an adaptive system processes the data from the evidence layer into higher level user characteristics such as interests, preferences or personal traits. It is important to mention that these are only estimates of the real user characteristics.

3. *adaptation and personalization* is the actual use of a user model (both evidence and higher level characteristics layer) to provide a user with personalized experience when using an adaptive system. The success rate of the supplied adaptation effect gives us a feedback about an accurateness of the underlying user model.

The process has a cyclic character, as an adaptive system continuously acquires new data about the user (influenced by already performed adaptation and thus by a previous version of the user model) and continuously refines the user model to better reflect reality and thus to serve as a better base for personalization.

An adaptive system must handle specific problems related to each step of the process. In *data collection* we must find a balance between user privacy and amount and nature of data which is required in order to deliver a successful personalization. We must find such sources of data, which do not pose additional extra burden on the user and design data collection process in such a way, which is unobtrusive and runs automatically in background.

The main challenge of the actual *user model inference* is (apart from actual transformation of user action history into user characteristics) maintenance of user characteristics, which should keep pace with user personal development, changing interests and knowledge.

3. Sources for User Modeling

Whole user modeling process stands and falls by the quality of data (facts) related to the users that an adaptive system is able to gather. If we aim at automated user modeling process, which poses no extra burden on user herself, Data collection stage of the user modeling process is obviously based on user activity logging where two main approaches exist: the use of standard web server logs or the deployment of proprietary logging/reporting solutions often tailored to the needs of specific systems. User model creation is based either on explicit user feedback or on automatic discovering user characteristics based on collected data.

Standard web server logs are commonly used as input for various data mining techniques in *Web Usage Mining* [13] whose results are mostly common sequential patterns or clusters of users and pages. These techniques match the active user session (or her previously stored profiles) to usage patterns of user groups. Because of their social aspect, they do not reveal characteristics of an *individual* user, which is of our prime interest. Web server logs cannot be used directly and require very complex preprocessing, trying to identify all accesses of an individual user within a particular session [12]. Use of proxy servers and tunnels makes this task non-deterministic, producing only estimations of log interpretation. More, web server logs are often incomplete due to the caching mechanism of web browsers, which usually do not re-request a resource which was accessed earlier during previous sessions. As a result, server side logs cannot serve as a satisfactory standalone source of all user behaviour evidence.

To address some shortcomings of this approach and to capture information about purely client-side interactions (e.g., hovering on tooltips) client-side logging methods were proposed. Most are either pure client-side only systems [19] or standalone desktop applications communicating with a server [15]. Standalone client-side applications can provide a high level of detail but may present serious threats to user privacy. The use of standard client web technologies such as JavaScript or Java applets [14] appears to be more suitable due to its restricted scope of operation and higher user acceptance though still without event semantics.

Pure client-side only solutions suffer from the loss of control over the logging process as users can disable their execution. Therefore, it is suitable to combine both client side logging with server side logging.

3.1 User Activity Logging with Semantics

We proposed a method of server-side event logging that allows for the logging of events with defined semantics by both server-side and client-side tools and for the integration of these events into continuous streams of events for a particular user and user session. Moreover, since only individual presentation/interaction tools "understand" the semantics of events that they process we propose them to be responsible for the logging of their own events by means of a specialized logging interface (Figure 2, bottom right).

We believe that the combination of the two approaches is necessary, since we have no direct control over the execution of client-side logging mechanisms (no data can be gathered if the user disables monitoring on the client side). The combination with the server side logging mechanisms guarantees that at least some data is always available.

3.1.1 Client-side Logging

Client-side logging captures events that occur entirely on the client-side (such as *Load*, *Click* or *Mouseover*) during user interaction with elements on the displayed pages and also captures precise time-related data about individual actions which might be missed by server-side logging. For each captured event, we collect the following data: *type of event*, *time* when it occurred and the *context* of the captured event, e.g. what link was followed.

Furthermore, additional events invisible to the server-side might be captured such as *Change*, invoked when the content of a form control changes. The sequence of such events indicates the order in which a user fills a form. Another example is the *Scroll* event, invoked when the user scrolls the content of a page.

3.1.2 Server-side Logging

We decided not to rely on standard web server logs, which lead to an application-dependent user modeling solution, but propose a proprietary server-side logging approach, similar to user modeling servers described in [17, 11].

Server-side logging aggregates events from various sources and stores their semantics which mostly include references to concepts from a domain model. To further separate log analysis from log creation we devised a common *event ontology* that defines the semantics of individual events and their attributes. Furthermore, to evaluate the reasons behind user actions we also log the description of display state of the user interface at the time when an event took place (what was displayed). This allows us to analyze user decisions (i.e., occurred events) based on what the user saw in the web browser interface (i.e., what her reasons were).

3.2 User Data Collection for Ubiquitous User Modeling on the Web

If we want to shift the personalized browsing experience from few isolated web applications, each performing its own user modeling process and thus building its own proprietary user model, we need to find a way how to do user modeling and adaptation independently of the web-sites a user is currently visiting. In other words, we need to shift the whole personalization layer towards the client side. However, at the same time we want to maintain a centralized logs of user activities, which serves as a basis for collective wisdom inference when analyzing surfing behavior of groups of users.

We evaluated a specialized HTTP-based proxy service as a most suitable host of a personalization layer [2]. As it has control over user requests and responses from the Web, it can perform all needed operations. As we mentioned earlier, a comprehensive log of user activity should result from a combination of server-side and client-side logging. This is also the case of our proxy-based solution. When proxy server is sending a web page to the client, it injects a tracking javascript into it, which updates the server with information about *time* user spent actively *on the page*, amount of *scrolling*, which user did on the page or amount of *copy into clipboard* actions user performed. All this client side information are added to a record with `url`, `ip address`, `timestamp` and `referrer`.

4. Open Corpus User Modeling

Traditional methods and techniques for user modeling and personalization require a limited set of documents, which need to be described by respective metadata at design time (and often manually) [9]. Such *closed corpus* approaches can not effectively keep pace with dynamics of our world, not to mention the Web, where information are constantly being updated as well as outdated. Closed corpus adaptive systems, which can provide *personalized* access to only limited set of resources are competing against all dynamic and virtually unlimited resources, which can be however accessed only by *one-size-fits-all* paradigm.

Open corpus adaptive web-based systems tries to solve the problem by incorporating metadata and external models, which exist independently of the information space of hyperlinked documents. If concepts from these models can be linked to the document space at run-time, we can model user characteristics related to those concepts instead of documents themselves.

4.1 User Model Inference based on Navigational Patterns

Our approach to data collection described in section 3.1 allows us to employ multiple reasoning agents to process the records of user interactions and update the user model. We proposed a method built around rule-based approach to analysis of user logs [4], i.e. a specialized rule formalism for capturing interesting patterns of user-system interaction and a user modeling inference agent driven by these rules. Externally defined rules represent heuristics, which link interesting navigational patterns with sets of changes to be applied on the user model if an occurrence of the pattern is detected. The inference process overview is depicted in Fig. 3.

We evaluated the combination of described data collection and user model inference in different application domains – online job offers (project NAZOU, `nazou.fiit.stuba.sk`) and digital libraries and scientific publications (project MAPEKUS, `mapekus.fiit.stuba.sk`). We performed an experiment in the job offers domain with hypothesis that the effort required to retrieve desired job

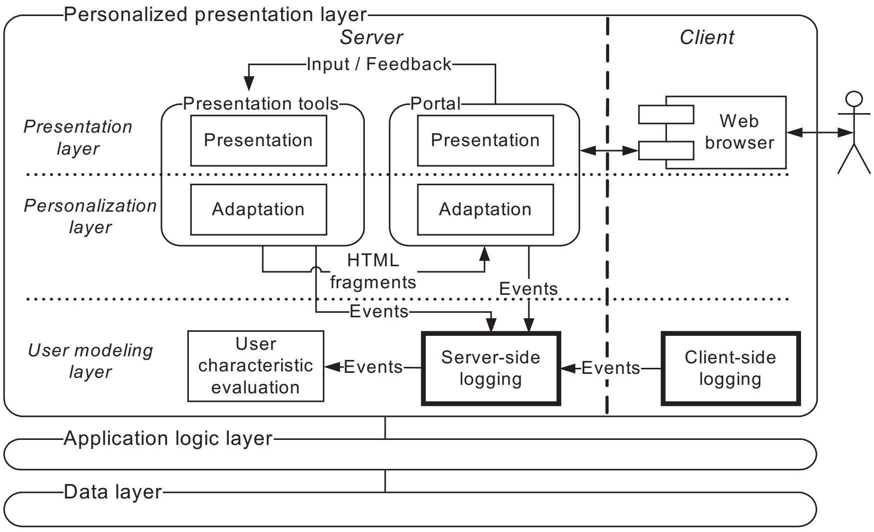

Figure 2: Web-based IS architecture of the personalized presentation layer.

offer will decrease if the user uses the system for a longer period time or revisit the system. The hypotheses was proved as we indeed observed lower click count in the revisiting user session compared to the first one indicating successful adaptation via navigational shortcuts based on the user model acquired during the first session (which was still updated in the second session). Moreover, the initially identified characteristics were reinforced in the second session while only few new ones were found indicating consistent user behaviour (i.e., good estimation of characteristics on the same search goals).

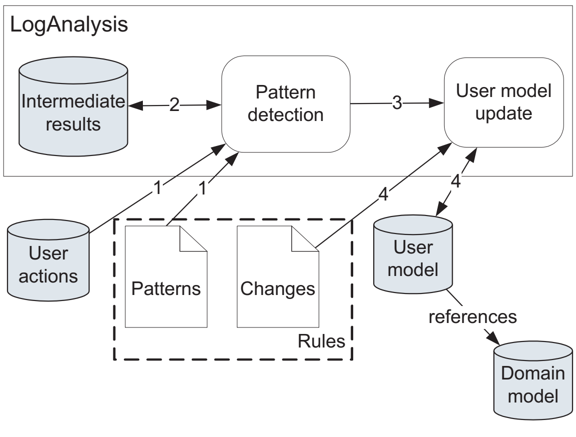

Figure 3: Overview of user characteristics acquisition process. Data from presentation tools and client-side logging are stored in a database of user actions. LogAnalyzer tries to detect (1) occurrences of predefined patterns and optionally stores intermediate results (2). The heuristics associated with the detected pattern (3) predict the update of characteristics stored in a user model (4).

4.2 Term-based User Modeling for "wild" Web adaptation

If we want to employ a model for the purpose of personalization in such a vast and dynamic environment as the whole Web is, we need to create a universal and flexible representation of the domain (which is the Web) – we need to acquire metadata about any web document visited by the users. The Semantic Web, which would be a natural candidate for such metadata, cannot be used for this purpose as there is still not (and probably never will be) enough of heavyweight semantics present on the Semantic Web when compared to the normal Web [22, 16]. Therefore, we need to rely on a lightweight conceptualization, which can be discovered automatically by machines and which can benefit also from knowledge emerging in online social systems (e.g., knowledge encoded within folksonomies).

Our user model is defined as a hypergraph $H := \langle V,\ E\rangle$, with a set of vertices $V = A \cup P \cup T$ and $P \cap T = \emptyset,\ A \cap P = \emptyset,\ A \cap T = \emptyset$ and a set of binary edges $E = \{(a, p, t) \mid a \in A,\ p \in P,\ t \in T\}$, where A represents a set of user accesses to pages $A = \{a_1, ...a_k\}$, P represents a set of pages $P = \{p_1, ..., p_l\}$ and T represents a set of terms $T = \{t_1, ..., t_m\}$. Pages and terms can be considered as a lightweight domain model, while user accesses add user model overlay on the top of it.

Such a term-based user model holds mainly facts, evidence of user accesses to web pages along with lightweight conceptualization of those web pages and implicit feedback indicators informing about user's interests. The main purpose of this model is to serve as a basis for virtual communities detection, which would allow for employment of community-based personalization to enhance the browsing experience. However, terms stored within the model can serve also as an indicator of user interests, supposing that as the time flows, the terms which represent user interests occur more often than the others.

4.2.1 Terms Acquisition

Our approach to terms acquisition follows the standard process for web document modeling [20] divided into two stages:

1. document pre-processing,
2. actual term extraction

In document-preprocessing stage, first an HTML page is processed in order to strip-off all the HTML markup and leave only main textual content, omitting navigational parts, banners etc. Second, the text is translated into English (if it is not already in this language) as majority of available extractors work only with English texts.

The second step, actual term extraction, is performed by invoking and combining results of different term extractors providing relevant terms for given text or URL. These extractors could be either online web services available through different kinds of API (usually REST-based) such as OpenCalais (`opencalais.com`) or Alchemy (`alchemyapi.com`) or locally invoked algorithms and libraries.

We realized the method within our enhanced proxy platform, already mentioned in section 3.2. Figure 4 depicts the complete term-based user modeling process with client-side logging and on-the-fly metadata (term) extraction.

We conducted an experiment on a small dataset of English and other language web-pages (having a representative of technical and non-technical texts and different writing styles) to evaluate efficiency of different available extractors. There is no clear winner among the evaluated extractors and it seems that they are, in effect, mutually eliminating their weakness, which means that they could be effectively used in a combination. For instance, when pure NLP oriented approaches fails, the semantically-enhanced services such as OpenCalais or tagthe.net are able to achieve fair results and vice versa. More details can be found in [3].

In order to evaluate the user modeling process in real world conditions, we developed an enhanced proxy server, which allows for realization of advanced operations on the top of requests flowing from a user and responses coming back from the web servers all over the Internet and a set of personalization tools – proxy plugins, which take advantage of term-based user model to deliver personalized surfing experience.

One of the plugins was enhancing the google search engine service by disambiguating search queries according to user context [18]. Apart from ordinary search results, it was also providing results coming from optimized and disambiguated queries. This time, we used term-based user models as a basis for construction of social networks and virtual communities of proxy users. We observed that users clicked and stayed on the results coming out from the extended queries in 54.7% of cases, which is a significant improvement against normal googling pattern of our users without any recommendations, where they stayed only on 27.4% of all clicked results, which mean that our recommendations were meaningful. As the whole search optimization was driven by automatically constructed open corpus user model using terms extracted from the visited web-pages, we can conclude that our approach is able to extract useful metadata from the web pages, produce a good enough user models which provide good basis for ad-hoc virtual communities creation and subsequent community-based personalization.

4.2.2 Term Hierarchy Creation

When considering term-based user models in a (term- or tag-based) Web 2.0 environment, we are facing the need to be able to compare various terms, tags and keywords. We might need to compare user characteristics or even whole user models, represented by terms, to find similarities between users, which could serve for user model maintenance as well as for more complex tasks such as online community creation or recommending. We might also need to compare the domain items represented by terms (e.g., a web page) in order to evaluate the explicit or implicit feedback and update the user model appropriately.

The raise of the tagging systems naturally provoked increase of interest in analysis of folksonomy data among researchers. We are aware of various approaches in deriving additional knowledge from folksonomies for different purposes. We believe that relationships between keyword and tags coming out of a folksonomy have a great potential for user modeling purposes, as the relationships are based on users' point of view – on the real usage of these tags and not on an artificially created linguistic knowledge.

We proposed a method for finding relationships between tags [1], which combines three distinct approaches:

1. Deriving of parent-child relationships between tags from a given folksonomy.

2. Determining similarity between tags by applying spreading activation on the top of the folksonomy graph.

3. Interconnecting tags by additional semantic relationships as well as enriching the whole tag corpus by adding external keywords; both coming from the Wordnet lexical database.

The proposed approach to creation of hierarchy from a folksonomy hypergraph is based on a rather simple assumptions of set theory, similarly to [21]. However, Mika does not take into account contextual conditions nor popularity of the tags. It seems that it was performing acceptably only on a small chosen domain (such as keywords related to semantic web).

We evaluated the approach on datasets of two popular online folksonomies – one obtained from delicious bookmarking site (`http://delicious.com`) and the other from a system for tagging and searching for scholarly papers CiteULike (`http://citeulike.org`).

We executed the algorithm for deriving parent-child relationships between tages several times on both folksonomies with different thresholds for establishing or discarding a relationship and were observing the resulting trees. The manual inspection of resulting hierarchies proved the viability of our approach, where meaningful relations between keywords were created, which would definitely provide a good basis for a tag-based user modeling. The quality of the result depends highly on the configuration of the algorithm.

The proposed method has even broader potential, as our representation of user model, holding all evidence of user

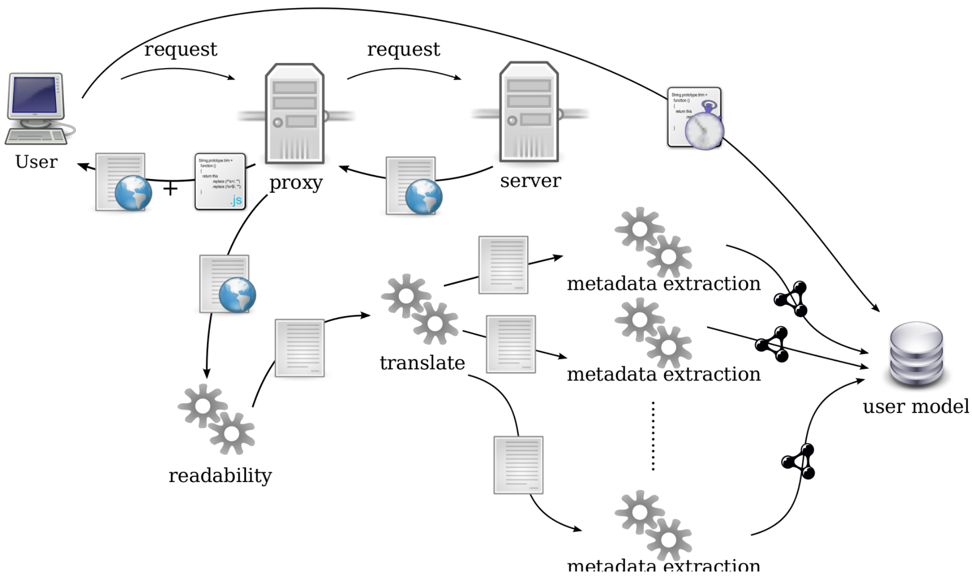

Figure 4: User Modeling process based on an enhanced proxy platform.

actions while surfing on the Web, is similar to folksonomy hypergraphs coming from social bookmarking systems. By applying the method on the logs of user actions, we can directly reveal the hidden relationships between different users as well as pages and their terms.

5. Conclusions and Contributions

We need to get used to a fact that the we need a machine assistance in order to work with large amount of dynamically changing information efficiently. Only machine-performed personalization allows us to keep pace with increasing amount of information. The majority of the current Web is however built using "one-size-fits-all" principle without any kind of personalization. Even more, those web-based systems, which are in effect personalized are often based on their proper models without any real possibilities to share and reuse user-related information across the Web.

In this thesis, we proposed methods which bring novel approaches to user modeling process. We focused on an open corpus user modeling, an approach which handles vast and dynamic information spaces such as the web is. Our main goals were to alleviate incorporation of the user modeling process into legacy web-based systems by extending a standard presentation layer by a user modeling sub-layer. Next, we were looking for solutions, which would bring the personalized experience to the ordinary everyday browsing on the Web by ubiquitous user and community modeling.

More specifically, we proposed and evaluated

- a method for comprehensive logging of user activity on the Web with preserved semantics. The method was published in *Computing and Informatics* journal [4], where we defined principles of turning the presentation layer of a web-based system into an adaptive one and presented a use case on the application of job offer portal.
- a method for capturing logs of "wild" Web surfing based on a specialized proxy sever. The method was presented at international *WWW/Internet 2010 conference* [3] where we showed how we can achieve a valid term-based model by combining various approaches to terms extraction.
- a method for open corpus user model inference based on rules expressing navigational patterns. The method was published in *Computing and Informatics* journal [4], where we showed how user model can be build and maintained regardless of the chosen application domain, considering only user's navigation.
- a method for term-based open corpus user modeling, which can be applied to capture user's interest across the third-party web-sites and web-based systems. The method was presented at international *User Modeling, Adaptation and Personalization conference, UMAP 2010* [18], where we showed how a term-based user model with implicit interest indicators can server as basis for virtual communities detection. We showed a potential of such virtual communities to deliver efficient personalization while surfing the "wild" Web on a use case of a personalized search engine.
- a method for finding relations between terms, which supports our term-based user modeling approach, based on folksonomies. The method was presented at *International Conference on Collective Computational Intelligence, ICCCI 2009* [1], where we showed how a set theory approach applied on a folksonomy can reveal hierarchical relationships between terms.

The proposed methods represent a contribution in the field of user modeling for adaptive web-based systems, more specifically in all parts of the user modeling process. New method for logging of user activity with preserved semantics combined with proposed method for rule-based user model inference allows for straightforward transformation of user's clickstream within the web-based system into a set of user model updates. More, the method can be setup to make very quick initial estimates of the user model according to the defined navigational patterns and is thus useful for both user model *maintenance* and *initialization*, to reduce cold-start problem.

We believe that social-based personalization is the key to finally deliver ubiquitous personalized experience to every website. We do not need to completely "understand" the content, but rely on behavior of visitors to distinguish what is interesting and what is not. However, we need to "understand" the content to the extent that allows us to cluster visitors accessing the content into virtual communities, to find users with similar interests so that they can mutually benefit from their community wisdom. This is exactly what we were aiming at by proposing the methods grouped around term-based user modeling, plugged in our specialized proxy server. They contribute to ubiquitous user modeling and open up various possibilities for community-based personalization.

The importance of new methods for open corpus personalization is even more visible, if we realize that traditional closed corpus web-based systems are being replaced by the systems driven by Web 2.0 principles. The most important one from our point of view is the support for user-generated content, which eliminates the difference between the roles of a *Web producer* and a *Web consumer*. Majority of web-based systems are thus becoming open-corpus and cannot directly apply traditional methods and techniques for closed corpus user modeling.

Acknowledgements. This work was partially supported by the State programme of research and development "Establishing of Information Society" under the contract No. 1025/04, the Slovak Research and Development Agency under the contract No. APVT-20-007104, the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010, the Scientific Grant Agency of SR, grant No. VG1/0508/09 and it is a partial result of the Research & Development Operational Programme for the project Support of Center of Excellence for Smart Technologies, Systems and Services II, ITMS 26240120029, co-funded by ERDF.

References

[1] M. Barla and M. Bieliková. On Deriving Tagsonomies: Keyword Relations Coming from Crowd. In N. Nguyen et al., editors, *ICCCI 2009*, LNAI 5796, pages 309–320. Springer, 2009.

[2] M. Barla and M. Bieliková. "Wild" Web Personalization: Adaptive Proxy Server. In *Workshop on Intelligent and Knowledge oriented Technologies, WIKT 2009*, pages 48–51. Centre for Information Technologies, FEI TU KE, 2009.

[3] M. Barla and M. Bieliková. Ordinary web pages as a source for metadata acquisition for open corpus user modeling. In B. White, P. Isaías, and D. Andone, editors, *WWW/Internet 2010*, pages 227–233, 2010.

[4] M. Barla, M. Tvarožek, and M. Bieliková. Rule-Based User Characteristics Acquisition from Logs With Semantics for Personalized Web-based Systems. *Computing and Informatics*, 28(4):399–427, 2009.

[5] M. Bieliková. Presentation of adaptive hypermedia on the web. In L. Popelínský, editor, *DATAKON 2003*, pages 72–91, Brno, ČR, 2003.

[6] P. Brusilovsky. Methods and Techniques of Adaptive Hypermedia. *User Model. User-Adapt. Interact.*, 6(2-3):87–129, 1996.

[7] P. Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1/2):87–110, 2001.

[8] P. Brusilovsky. Social information access: The other side of the social web. In V. Geffert et al., editors, *SOFSEM 2008: Theory and Practice of Computer Science*, LNCS 4910, pages 5–22. Springer, 2008.

[9] P. Brusilovsky and N. Henze. *The Adaptive Web*, chapter Open Corpus Adaptive Educational Hypermedia, pages 671–696. LNCS 4321. Springer Berlin/Heidelberg, 2007.

[10] P. Brusilovsky and E. Millán. *The Adaptive Web*, chapter User Models for Adaptive Hypermedia and Adaptive Educational Systems, pages 3–53. LNCS 4321. Springer Berlin/Heidelberg, 2007.

[11] P. Brusilovsky, S. Sosnovsky, and O. Shcherbinina. User Modeling in a Distributed E-Learning Architecture. In L. Ardissono, P. Brna, and A. Mitrovic, editors, *User Modeling 2005*, LNCS 3538, pages 387–391, Edinburgh, Scotland, UK, July 2005.

[12] Z. Chen, A. Fu, and F. Tong. Optimal Algorithms for Finding User Access Sessions from Very Large Web Logs. *World Wide Web*, 6(3):259–279, 2003.

[13] M. Eirinaki and M. Vazirgiannis. Web mining for web personalization. *ACM Trans. Internet Techn.*, 3(1):1–27, 2003.

[14] M. Etgen and J. Cantor. What does getting WET (Web Event-logging Tool) Mean for Web Usability? In *Human Factors & The Web: The Future of Web Applications*, NIST, Gaithersburg, Maryland, USA, 1999.

[15] K. Fenstermacher and M. Ginsburg. Mining Client-Side Activity for Personalization. In *Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems, WECWIS '02*, pages 205–212, Newport Beach, California, USA, 2002.

[16] M. Fernandez et al. Semantic search meets the web. In *ICSC '08: International Conference on Semantic Computing*, pages 253–260, Washington, DC, USA, 2008. IEEE Computer Society.

[17] J. Kay, B. Kummerfeld, and P. Lauder. Personis: A Server for User Models. In P. De Bra, P. Brusilovsky, and R. Conejo, editors, *Adaptive Hypermedia and Adaptive Web-Based Systems, AH 2002*, LNCS 2347, pages 203–212, Malaga, Spain, 2002. Springer.

[18] T. Kramár, M. Barla, and M. Bieliková. Disambiguating Search by Leveraging the Social Network Context. In *User Modeling, Adaptation and Personalization, UMAP 2010*, LNCS 6075, pages 387–392. Springer, 2010.

[19] H. Lu, Q. Luo, and Y. Shun. Extending a Web Browser with Client-Side Mining. In X. Zhou, Y. Zhang, and M. Orlowska, editors, *Web Technologies and Applications, 5th Asian-Pacific Web Conference, APWeb 2003*, LNCS 2642, pages 166–177, Xian, China, 2003. Springer.

[20] A. Micarelli, F. Sciarrone, and M. Marinilli. Web document modeling. In *The Adaptive Web*, LNCS 4321, pages 155–192. Springer Berlin/Heidelberg, 2007.

[21] P. Mika. Ontologies are us: A Unified Model of Social Networks and Semantics. *J. Web Sem.*, 5(1):5–15, 2007.

[22] M. Sabou et al. Evaluating the semantic web: a task-based approach. In *ISWC '07: International Semantic Web Conference*, LNCS 4825, pages 423–437, Berlin, Heidelberg, 2007. Springer.

[23] M. Šimún, A. Andrejko, and M. Bieliková. Maintenance of Learner's Characteristics by Spreading a Change. In M. Kendall and B. Samways, editors, *IFIP Int. Federation for Information Processing*, volume 281 of *Learning to Live in the Knowledge Society*, page 223226. Boston: Springer, 2008.

Selected Papers by the Author

Barla, M., Tvarožek, M., and Bieliková, M. (2009). Rule-Based User Characteristics Acquisition from Logs With Semantics for Personalized Web-based Systems. *Computing and Informatics*, 28(4):399–427.

Barla, M., Bieliková M., Bou Ezzeddine, A., Kramár, T., Šimko, M., Vozár, O. (2010). On the Impact of Adaptive Test Question Selection for Leatning Efficiency *Computers & Education*, 55(2):846-857.

Kostková, K., Barla, M., Bieliková M. (2010) Social Relationships as a Means for Identifying an Individual in Large Information Spaces In *Artificial Intelligence in Theory and Practice III*, IFIP Advances in Information and Communication Technology, Vol. 331, Springer, pp.35-44.

Šimko, M. and Barla and Bieliková, M. (2010) ALEF: A Framework for Adaptive Web-based Learning 2.0 In *Proc. of IFIP Advances in Information and Communication Technology, 2010*, Springer, Vol. 324/2010, pp. 367-378.

Barla, M. (2006) Interception of User's Interests on the Web. In Wade, V., Ashman, H., Smyth, B. (Eds.): *Adaptive Hypermedia and Adaptive Web-Based Systems, AH'06*, LNCS 4018, Springer, pp. 435–439.

Bartalos, P., Barla, M., Frivolt, G., Tvarožek, M., Andrejko, A., Bieliková, M., Návrat, P. (2007) Building Ontological Base for Experimental Evaluation of Semantic Web Applications. In J, van Leeuwen, et al. (Eds.): *SOFSEM 2007*, LNCS 4362, Springer, pp. 682–692.

Tvarožek, M., Barla, M., Bieliková, M. (2007) Personalized Presentation in Web-Based Information Systems. In J, van Leeuwen, et al. (Eds.): *SOFSEM 2007*, LNCS 4362, Springer, pp. 796–807.

Tvarožek, M., Barla, M., Frivolt, G., Tomša, M., Bieliková, M. (2008) Improving Search in the Semantic Web via Integrated Personalized Faceted and Visual Navigation. In: Geffert, V. et al. (Eds.): *SOFSEM 2008*, LNCS 4910, Springer, pp. 778-789.

Andrejko, A., Barla, M., Bieliková, M. (2007) Chap. Ontology-based User Modeling for Web-based Information Systems. In: Magyar, G., et al. (Eds.): *Advances in Information Systems Development, Vol. 2.* Springer, pp. 457–468.

Barla, M., Bartalos, P., Sivák, P., Szobi, K., Tvarožek, M., Filkorn, R. (2007) Chap. Ontology as an Information Base for a Family of Domain Oriented Portal Solutions. In: Magyar, G., et al. (Eds.): *Advances in Information Systems Development, Vol. 1.* Springer pp. 423–433.

Barla, M., Bieliková, M (2009). On Deriving Tagsonomies: Keyword Relations Coming from Crowd. In Nguyen, N.T., Kowalczyk, R. (Eds.): *Computational Collective Intelligence: Semantic Web, Social Networks and Multiagent Systems, ICCCI 2009*, LNCS 5796, Springer, pp. 309-320.

Barla, M., Bieliková, M (2009). An Extensible Open-Source Framework for Social Network Analysis In *Advances in Intelligent Web Mastering - 2: Proc. of the 6th Atlantic Web Inteligence Conference – AWIC 2009*, Springer, pp. 63-60.

Kramár, T., Barla, M., Bieliková, M. (2010) Disambiguating Search by Leveraging a Social Context Based on the Stream of User's Activity. In De Bra, P., Kobsa, A., Chin, D. (Eds.):*User Modeling, Adaptation, and Personalization, UMAP 2010*, LNCS 6075, Springer, pp. 387-392.

Bieliková, M., Šimko, M., Barla, M. (2010) Personalized Web-Based Learning 2.0 In: *ICETA 2010*, Elfa

Barla, M.,Bieliková, M. (2010) Ordinary Web Pages as a Source for Metadata Acquisition for Open Corpus User Modeling In White,B., Isaías, P., Andone, D., (Eds.): *WWW/Internet 2010*, IADIS Press, pp. 227-233

Bieliková, M., Šimko, M., Barla, M., Chudá, D., Michlík, P., Labaj, M., Mihál, V., Unčík, M. (2010) ALEF: Web 2.0 Principles in Learning and Collaboration In: *e-Learning10*, Riga Technical University, pp. 54-59

Barla, M., Andrejko, A., Bieliková, M., Tvarožek, M. (2007) User Characteristics Acquisition from Logs with Semantics. In Kelemenová, A., Kolář, D., Meduna, A., Zendulka, J., (Eds.): *ISIM '07: Information Systems and Formal Models*, Slezská universita v Opavě, pp. 103–110.

Barla, M., Bartalos, P., Bieliková, M., Filkorn, R., Tvarožek, M. (2007) Adaptive Portal Framework for Semantic Web Applications. In Brambilla, M., Mendes, E., eds.: *AEWSE'07: Workshop on Adaptation and Evolution in Web Systems Engineering held at ICWE 2007*, Politecnico di Milano, Milano, Italy, pp. 87-93.

Barla, M., Bieliková, M. (2007) Estimation of User Characteristics using Rule-based Analysis of User Logs. In: *Data Mining for User Modeling, Proc. of Workshop held at UM2007*, pp. 5-14.

Tvarožek, M., Adam, M., Barla, M., Sivák, P., Bieliková, M. (2006) Spot-it: Going Beyond the Vision Loss Boundaries In *ISIE2006*, Microsoft Research Ltd, pp. 67-76.

Barla, M., Bieliková (2007) What characteristics of an individual are we able to capture automatically when searching in large information space? How much of them do we need? In Laclavík, M., Budinská, I., Hluchý, L., (Eds.), *Workshop on Intelligent and Knowledge oriented Technologies WIKT 2006*, II SAS, pp. 34-36 (in Slovak).

Barla, M., Bieliková, M. (2008). Leveraging Semantic Web technologies for personalized e-Learning. In F. Babič, J. Paralič (Eds.), *Workshop on Intelligent and Knowledge oriented Technologies WIKT 2007*, Centre for Information Technologies, FEI TU KE, pp. 36-40, (in Slovak).

Barla, M., Bieliková, M (2009) Leveraging social networks to initialize user model for personalized web-based systems. In Návrat, P., Vranić, V. (Eds.), *Workshop on Intelligent and Knowledge oriented Technologies WIKT 2008*, STU, pp. 5-8 (in Slovak).

Barla, M., Bieliková, M (2009) "Wild" Web personalization: Adaptive proxy server In F. Babič, J. Paralič (Eds.) *Workshop on Intelligent and Knowledge oriented Technologies WIKT 2009*, Centre for Information Technologies, FEI TU KE, pp. 48-51 (in Slovak).

Barla, M., Kramár, T., Bieliková, M (2010) Adaptive Proxy Server: Operation and Experiences After a Year In Laclavík, M., Hluchý, L., (Eds.), *Workshop on Intelligent and Knowledge oriented Technologies WIKT 2010*, II SAS, pp. 28-31 (in Slovak).

Andrejko, A., Barla, M., Bieliková, M., Tvarožek, M. (2006) Software tools for user characteristics acquisition. In Vojtáš, P., Skopal, T., eds.: *Datakon 2006*, pp. 139-148 (in Slovak).

Barla, M., Bieliková, M (2009). Working teams creation supported by a social network. In *Datakon 2009*, pp. 153-159 (in Slovak).

Kramár, T., Barla, M. (2009) Web site usage pattern mining. In *Znalosti 2009*, pp. 309-312 (in Slovak).

Barla M., Kompan, M., Suchal, J., Vojtek, P., Zeleník, D., Bieliková, M. (2010) Project on news articles recommendation. In *Znalosti 2010*, pp. 171-174 (in Slovak).