Personalized Web - Science, Technologies and Engineering

9th Spring 2011 PeWe Workshop,
Viničné, Galbov Mlyn, Slovakia,
April 2011, Proceedings

**Mária Bieliková, Pavol Návrat,
Michal Barla, Jozef Tvarožek,
Michal Tvarožek (Eds.)**

Proceedings in
Informatics and Information Technologies

Personalized Web – Science,
Technologies and Engineering
9th Spring 2011 PeWe Workshop

Mária Bieliková, Pavol Návrat,
Michal Barla, Jozef Tvarožek,
Michal Tvarožek (Eds.)

Personalized Web – Science, Technologies and Engineering

9th Spring 2011 PeWe Workshop
Viničné, Galbov Mlyn, Slovakia
April 1, 2011
Proceedings

Slovakia Chapter

PeWe Group

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
Faculty of Informatics and Information Technologies

Proceedings in
Informatics and Information Technologies

Personalized Web – Science, Technologies and Engineering
9th Spring 2011 PeWe Workshop

Editors

Mária Bieliková, Pavol Návrat,
Michal Barla, Jozef Tvarožek, Michal Tvarožek

Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 3, 842 16 Bratislava, Slovakia

Production of this publication was partially supported by the Scientific Grant Agency of Slovak Republic, grants No. VG1/0508/09, VG1/0675/11 and by the Cultural and Educational Grant Agency of the Slovak Republic, grants KEGA 028-025STU-4/2010, KEGA 345-032STU-4/2010.

Visit PeWe (Personalized Web Group) on the Web: pewe.fiit.stuba.sk

Executive Editor: Mária Bieliková
Cover Designer: Peter Kaminský

Published by:
Nakladateľstvo STU
Vazovova 5, Bratislava, Slovakia

ISBN 978-80-227-3485-1

Preface

The Web influences our lives for more than 20 years now. During these years, it has continuously been enjoying a growing popularity due to, among other things, its progressive change from passive data storage and presentation vehicle to the infrastructure for software applications and to the place for communication, interaction, discussions and generally collaboration. As the Web has an influence on our work, entertainment, friendships, it attracts more and more researchers who are interested in various aspects of the Web, seeing it from various perspectives – as a science, a place for inventing various technologies or engineering the whole process.

Research in the field of the Web has more than 10 years of tradition at the Institute of Informatics and Software Engineering, Slovak University of Technology in Bratislava. Moreover, topics related to the Web attract many students. This volume is entirely devoted to students and their research. It contains extended abstracts of students' research projects presented at the 9th PeWe (Personalized Web Group) Workshop on the Personalized Web, held on April 1, 2011 in nice environment of Viničné in Galb's Mill near Bratislava. It was organized by the Slovak University of Technology (and, in particular, its Faculty of Informatics and Information Technologies, Institute of Informatics and Software Engineering) in Bratislava. Participants are students of all three levels of the study – bachelor (Bc.), master (Ing.) or doctoral (PhD.), and their supervisors.

The workshop covered several broader topics related to the Web, which served for structuring these proceedings:

- Search, Navigation and Visualization,
- Classification and Recommendation,
- User Modelling, Virtual Communities and Social Networks,
- Domain Modelling, Semantics Discovery and Annotations.

The workshop started by students' presentations of their research. The projects were at different levels according to the study level (bachelor, master or doctoral) and also according the progress stage achieved in each particular project. Moreover, we invited to take part also five of our bachelor study students who take an advantage of our research track offered within their study programme and who just start their bachelor projects – *Ľuboš Demovič, Jakub Kříž, Marek Láni, Martin Lipták* and *Matúš Tomlein.*

Bachelor projects:

- *Pavol Bielik, Peter Krátky, Štefan Mitrík, Michal Tomlein*: Personalized Recommendation and Context-Aware Avatar for Motivating Children to Physical Activity

- *Máté Fejes*: Concept-Cloud Navigation in Educational Web-Based System
- *Róbert Horváth*: Interpretation Support of Terms while Browsing in Slovak Language
- *Peter Macko*: Interactive Browser of Heterogeneous Web Content
- *Balázs Nagy*: Metadata Acquisition via Interactive Games
- *Jakub Ševcech*: Web Content Annotation in Educational Web-Based Application

Master projects (started in the current academic year):

- *Anton Balucha*: Using Social Relationships for Searching Relevant Information
- *Anton Benčič*: Information Recommendation with Use of Context in a Specific Domain
- *Peter Kajan*: Discovering Keyword Relations
- *Milan Lučanský*: Acquiring Metadata from the Web
- *Róbert Móro*: Personalized Text Summarization
- *Ivan Srba*: Encouragement of Collaborative Learning Based on Dynamic Groups
- *Peter Svorada*: Modelling a Tutor for E-Learning Support
- *Márius Šajgalík*: Devising Secure Communication for Decentralized Environment
- *Maroš Unčík*: Combining Different Data-Sources for User Modelling in Personalized Learning

Master projects (started in the previous academic year)*:*

- *Marián Hönsch*: Detecting User Communities Based on Latent and Dynamic Interest on a News Portal
- *Martin Jačala*: Named Entity Disambiguation using Wikipedia
- *Eduard Kuric*: Automatic Photo Annotation Based on Visual Content Analysis
- *Martin Labaj*: Recommendation and Collaboration through Implicit Feedback
- *Michal Lohnický*: Photo Album Visualization as a Collection of Memories and Experiences
- *Tomáš Majer*: Leveraging Microblogs for Resource Ranking
- *Vladimír Mihál*: Relations Discovery in Educational Texts based on User Created Annotations
- *Karol Rástočný*: Semantic Web Navigation Based on Adaptive Views
- *Peter Študent*: Using Implicit Feedback in Recommendation Systems
- *Martin Virik*: Automated Recognition of Writing Style in Blogs

Doctoral projects

- *Michal Holub*: Information Integration on Social Adaptive Web
- *Tomáš Kramár*: Improving the Personalized Search on the Web
- *Michal Kompan*: Group Recommendation
- *Ján Suchal*: New Approaches to Log Mining and Applications to Collaborative Filtering
- *Jakub Šimko*: Harnessing Manpower for Semantics Acquisition
- *Marián Šimko*: Automated Metadata Extraction for Adaptive Social Learning System
- *Dušan Zeleník*: Modelling Context Relations to Discover Hidden Contexts

Considerable part of our research meeting this year was devoted to a *hack-day-like activity* chaired by Michal Barla and Tomáš Kramár. The initial motivation was to allow the students to play and experiment with various available platforms, services, or APIs or data sets and take advantage of innovative potential of PeWe group and unique connection of bright young researchers on various levels of their competences to invent something new and interesting in the field of adaptive social web. We chose our specialized proxy server called PeWeProxy to play the role of a framework upon which workshop participants were asked to build their solutions. PeWeProxy proved as perfect choice. Even within constraints such limited time (4 hours plus instructions given day before the workshop) and teams, which were not used to work together we get really amazing results – interesting ideas and functional prototypes in most cases.

Another workshop activity was aimed at study the real behaviour of humans in order to be able better model users of adaptive systems. To study the emotional contagion, the social status impact, optimal recommendation sequence and aggregation performed by users we designed a *questionnaire* given to workshop participants. Results of this questionnaire were summarized in short note by Michal Kompan.

During almost the whole workshop we conducted the *Seminar Bingo* event chaired by Jakub Šimko. This recessive event was inspired by the popular PHD Comics series, which describes the life of the academic researchers in a humorous, appealing way. Such kind of humour is also appreciated in our community and the event was therefore cheerfully welcomed. An example of the Seminar Bingo sheet used on the workshop is included in short report at the end of these proceedings.

Our workshop hosted for the fourth time recessive activity organized by the *SeBe (Semantic Beer) initiative* aimed at exploration of the Beer Driven Research phenomenon. SeBe workshop was chaired by Marián Šimko, Michal Barla and Jakub Šimko. It included various activities such as traditional LittleBeerGame or newly established collaborative social game named *ya SeBeC* (Yet Another SeBe Competition).

More information on the PeWe workshop including presentations is available in the PeWe group web site at pewe.fiit.stuba.sk. Photo documentation is available at mariabielik.zenfolio.com/ontozur2011-04.

PeWe workshop was the result of considerable effort by our students. It is our pleasure to express our thanks to the *students* – authors of the abstracts, for contributing interesting and inspiring research ideas. Special thanks go to Katka Mršková and Saška Bieleková for their effective support of all activities and in making the workshop happen.

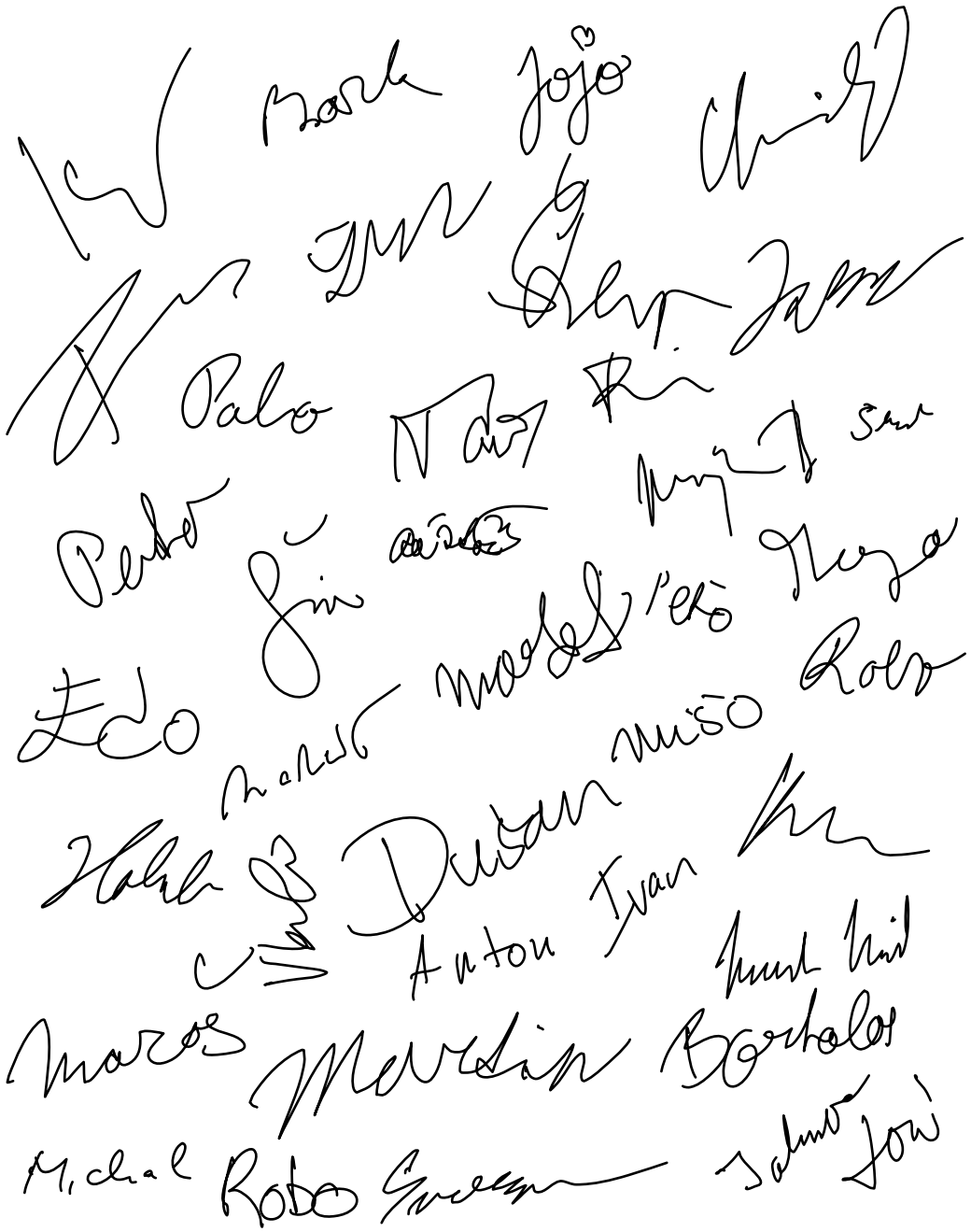
April 2011

Mária Bieliková, Pavol Návrat,
Michal Barla, Jozef Tvarožek, Michal Tvarožek

Table of Contents

Workshop Events Reports

Workshop participants
Viničné, Galb’s Mill



PeWe Ontožúr Participants in Action

Search, Navigation and Visualization

Using Social Relationships for Searching Relevant Information

Anton BALUCHA*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
a.balucha@gmail.sk

Web is experiencing huge growth in recent years. What was formerly accessible only to few people and universities, nowadays is now available to common people and constitutes the most extensive communication tool available in the world. It is the largest, most widely known and most accessible source of information.

Internet enabled communication and connections between people from its beginning. In the beginning there was e-mail, later there were IRC channels and portals that enabled communication between people. Presently, social networks contain a significant part of communication between people in contrast to other communication tools. People exchange messages, photos, videos and look for new friends.

When users search for new friends, they proceed by their private preferences. In present days, there is ever more personal information available about people. Somebody wrote this information about some other person or, especially in recent years, some people wrote this information about themselves. They use various portals for presenting their knowledge, discussion forums for presenting their opinions, blogs for expressing some experience, social networks for communication with fiends and photos for saving their memories. We can thus conclude that if users actively use all aforementioned communication and presentation tools, we can get complex view about their personality. On many sites there is information about the same people, and at the same time, many times these are only namesakes with nothing in common.

If we use search engines to search for information about people, their knowledge, opinions, photos or contacts we receive scattered information and must browse many pages to get a decent overview. Many times, the same information is found in different locations and different formats. These various pages with much information about people can be integrated and presented in a coherent form to users or used for further information retrieval tasks.

* Supervisor: Anna Bou Ezzeddine, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 3–4.

In the huge amount of data stored in the Web, it is hard to find information with proper quality and relevance. Using social relationships can help us to search for information, choosing and better specification of the desired information.

Since we often have friends who are similar to us, we can exploit this fact to choose better, more relevant information. These friends have their own interests, they browse their favourite pages, they publish information about themselves and all this can help us to find better and more relevant information. As we pass through the process of getting information about people, then about their friends and then we find the information, we also pass through some possibilities of searching and storing information in these phases.

We search for information about people via regular expression and by means of predefined rules. We deal with using different formats for storing information about people - such as vCard or microformats, and also the PIMO format, which is Personal Information Model for description of personal data and can be used in the Semantic Web. Another interesting format is IMS Enterprise 1.1 XML Binding, which can, beside personal information, also store information about processes. Of course, beside text formats for storing information we examine possibilities to store this data into a properly created database structure.

The next phase is searching and storing information about social connections between people. We examine possibilities of getting information about people from pages and also their search dynamics. The last phase is searching for information via profiles of people exploiting the idea that we search for friends who are similar to us.

In this process of searching for relationships and information, we use algorithms for searching information inspired by behaviour of social insects and work with search inspired by algorithms of fuzzy ants.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Lefever, E., Fayruzov, T., Hoste, V., De Cock, M.: Clustering web people search results using fuzzy ants. *Information Sciences: an International Journal archive*, Vol. 180, No. 17, 2010.

[2] Kanade, P.M., Hall, L.O.: *Fuzzy Ants and Clustering*. Tech. Report, Dept. of Computer Science and Engineering, ENB118, University of South Florida, 2006.

[3] *Personal Information Model*, Available at http://www.semanticdesktop.org/ontologies/pimo/ (17-03-2011)

Concept-Cloud Navigation in Educational Web-Based System

Máté FEJES*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
mfejes@azet.sk

In this project we focus on improving the navigation within the web-based educational system ALEF. The goal is to propose a method of content recommendation with the help of displaying keywords representing content elements. The method is focused on effective detection and elimination of misconceptions in student knowledge. The ALEF system consists of so-called *learning objects* of three types: *text-explanations, questions* and *exercises*. *Metadata* is used to define learning objects. The learning objects are linked to *concepts*, which are represented as keywords. In our method, we are particularly interested in these concepts.

In our approach, the keywords (concepts) are displayed in a form of tag clouds. Two types of concepts are displayed in the concept cloud in separate boxes. The first type is *Related concepts* – all the concepts linked to the currently opened learning object. They serve the purpose of fast switching among similar learning objects that belong here. The second part of the cloud – *We recommend* – is a personalized set of recommended concepts devised from the user's activity history.

During navigating in the system, we can move to learning objects using the main menu. By selecting a learning object in the menu, we get to a new state, and the particular information is displayed. On the other hand, the tag cloud is used differently. Instead of navigating the user to an actual learning object, it recommends topics he/she should deal with. By clicking on a concept in the tag cloud, the related learning objects are consequently highlighted in the main menu. In this way we indicate a suitable direction for the user to navigate but the final decision to select one of the recommended objects remains to be made by the user.

Recommendation methods are processes, which have certain inputs and outputs. The inputs of these processes are learning objects of a given type. For each input object a single concept is recommended on output. Each output concept has an associated priority, which represents the strength of the recommendation. Our method then selects the concepts with the highest priorities and displays them in the tag cloud part "We recommend".

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 5–6.

One of the recommendation methods is *Method based on questions and examples solutions* (see Figure 1). The input of this method is a question or example, which was incorrectly solved by the user. In this case the priority of the concepts linked to the input learning object is given by quotient of the representation relation weight and the knowledge of the concept. If there are more concepts linked to one input learning object, we choose the one with the highest priority. The outputs of the recommendation are concepts of learning objects, which should be studied again over by the student in order to improve proficiency in the particular topic.

Another method is a *Method based on reading of learning materials* (see Figure 1). The input is a text-explanation that was read (visited) by the user, but it does not have any linked concepts, which have a knowledge relation with the user. It is about learning materials students have already read, but have not practiced yet. The system recommends questions and examples which are related to the previously read text.

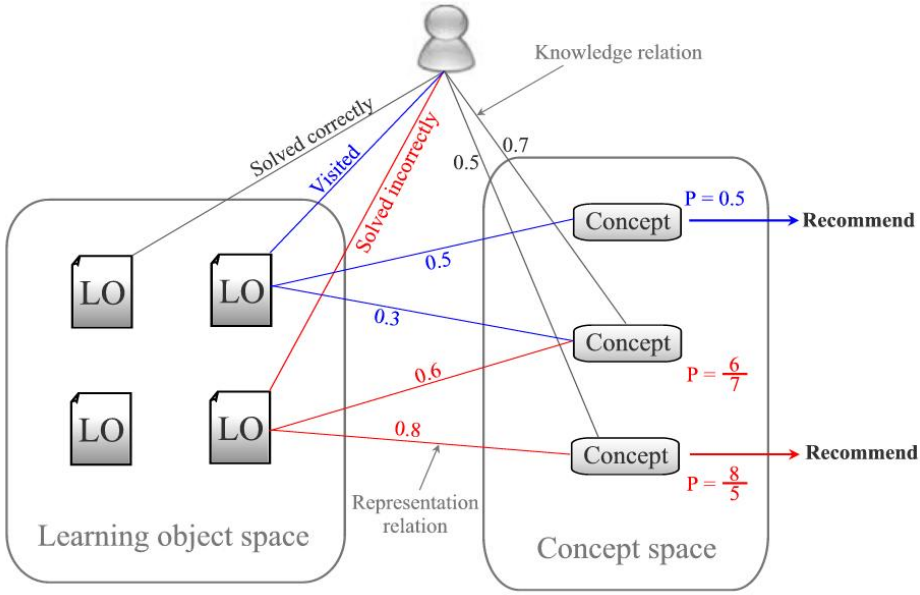


Figure 1. Selecting concepts for recommendation utilizing user's activity history.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Fejes, M.: Concept-Cloud Navigation in Web-Based Educational System. In *Proceedings of Student Research Conference, IIT.SRC 2011*, Bratislava, 2011.

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

Interpretation Support of Terms while Browsing in Slovak Language

Róbert HORVÁTH*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
roberthorvath89@gmail.com

Today, Web is an inseparable part of our everyday life. When accessing web pages, it is necessary to fully understand their meaning in order to take full advantage of information contained within. In many technical articles there are words and phrases, whose meaning is unknown to many users. Most users look for the explanations of phrases in online dictionaries – they have to open a new window with the dictionary and manually enter the corresponding phrases. Furthermore, when considering polysemy, they need to identify the correct meaning. This is not very fast neither comfortable. Another possibility for users is to use web browser extensions that can automate the whole process. Typically, extensions are made to show the main explanation of words in tooltips (e.g., Google Dictionary, Dictionary Lookup). Their drawback being that they are not created to support Slovak language.

The goal of our work is to create a tool that will provide explanations of Slovak words during web browsing by automatically looking up their definitions in available online dictionaries. If a word has multiple meanings we need to perform word sense disambiguation [1]. In order to keep the process automatic, we aim only for unsupervised methods without the need for human help. We combine existing methods of text processing and text comparison to acquire correct term definitions. In order to evaluate our approach to word definition acquisition, we have decided to create a web browser extension. The extension utilizes a web service designed for lemmatization and text processing. When a user finds an unknown word while browsing the Web, he simply selects the word and the extension displays the most probable meaning of the word by showing a word definition.

Our approach to acquire term definition consists of the three main steps (see Figure 1):

1. web page part (a selected term and its neighborhood) text pre-processing,
2. potential definition lookup,
3. correct definition selection.

* Supervisor: Marián Šimko, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 7–8.

Input data gathered from a web page contains some information not relevant for finding of meaning or similarity measurement like HTML tags, numbers, symbols, etc. Such data (e.g., HTML tags, stop words) are removed and each word is transformed into its basic form in order to enable further comparison.

One of the most important parts of our approach is word definition lookup. There are many types of online services able to provide such functionality. We consider online dictionaries, but search engines can be utilized as well. Online dictionaries typically need words in basic form – *lemma* as an input. Due to the polysemy issue present in every natural language, input often matches more than one definition. Thus, the output of this step is typically a list of potential definitions. We utilize existing online dictionaries that represent extendable and configurable options for users.

Correct word definition selection considering its context (as means to resolve polysemy) is done via similarity calculation. In this step we calculate cosine similarity [2] between the selected word context (i.e., the textual neighbourhood of the word) and potential definitions. We expect words describing the correct meaning to be located close to the unknown word. Experiments we had conducted have shown that *paragraph* was the best neighbourhood option (80% success rate in comparison with sentence – 60% and the whole page text – 70%, tested for 20 homonyms).

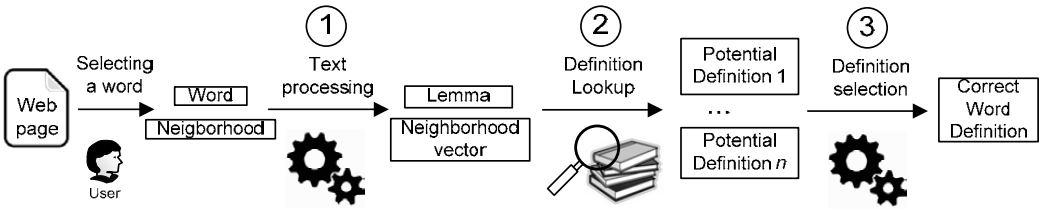


Figure 1. Process of acquiring term definition.

In the current stage of our work, we have already have created a browser extension, which provides users with context based term definitions. We plan to conduct small experiments with a select group of users to gather feedback and evaluate the extension.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Paralič, J., Furdík, K., Tutoky, G., Bednár, P., Sarnovský, M., Butka, P., Babič, F. Preprocessing text data. In *Mining knowledge from texts*, Technical University Košice, pp. 15-65, 2010 (in Slovak).

[2] Krajči, S., Novotný, R. Finding common base of Slovak words employing common suffix. In *Proc. of 1st Workshop on Intelligent and Knowledge oriented Technologies*, pp. 99-101, 2007 (in Slovak).

Improving the Personalized Search on the Web

Tomáš KRAMÁR*

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

In our daily contact with the Web, we are faced with a vast amount of documents and their number increases every moment. The situation is in large part caused by the rise of the Social Web, where users are encouraged to not only passively consume, but also actively create the content. The amount of new content created every day is tremendous and it is obvious, that to access an information base this large, powerful methods for content analysis and retrieval are required.

To facilitate easy access to this knowledge base, search engines were created. In a highly oversimplified point of view, search engines enable finding the document by entering a set of keywords which describe user's intent – if these keywords are found in the text of the document, it is retrieved and presented to the user. However, the fulltext-only search is often incapable of handling user queries satisfactorily and has several known disadvantages.

To mitigate this problem, several approaches to search personalization have been researched, each with the ultimate aim to help the user find the relevant content, without trying to change how humans think, or work. There is relevance feedback, query expansion, search intent detection, alternative ranking schemes and many others.

In the domain of user modeling, the term context traditionally refers to the attributes of the environment (i.e., user's location, time, her mood, etc.). In the domain of personalized search, the term context is commonly used to describe user's needs, goals and intent [2]. But there is a difference between a personalized search and a search within a specific context. The personalized search deals with adapting the search to a user's "personality", while the contextual search is concerned with adapting the results to a particular context. The difference is in the scale – the personalized search deals with long-term user's preferences, while the contextual search focuses more on the intermediate needs. While many personalization approaches exist, the contextual search is still largely unexplored area. In this work we research various possibilities to make the search more contextual.

One of the biggest shortcomings of current personalized search systems is that they usually treat user as one big monolithic entity. The user model is built as rich as

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 9–10.

possible, usually from all actions the user has ever made and that are available to the system in some form. This is partly correct, because the search system must have enough data about the user in order to make confident decisions, but this may also sometimes cause problems. For illustration, a biologist may be interested in the wild jaguar animals and the personalized search system can leverage this kind of information to favor animal-related documents in her day-to-day searching. But if the user is temporarily interested to buy a car, the search for jaguar is suddenly not so useful. We can alleviate this problem by distinguishing between user's long-term interests and her short-term goals, and combining both to fetch the most helpful documents that focus on the intermediate needs. That does not mean that the long-term interests should be abandoned, we just need to find a better way to distinguish when is the right time to use them [1]. It is also important that we are able to tell the exact moment when the user changes his current interests, so that the system can start to build a new short-term profile.

Another problem is the lack of user model specialization. This problem is again caused by the monolithic way the user models are built. No user has a single area of interest; usually, a user has multiple personas – the social roles, the facets of life. One might be a researcher in her day-to-day job, but after work changes to a mother of a family or perhaps to a passionate traveler through weekends. The personalized search system should be able to distinguish these roles in order to achieve better personalization. The challenge in this research area is in a way to detect and model the user's current role, both at the user model creation time and at the query time.

Creating large user models is impractical; such user models are difficult to work with, contain a lot of noise and may consume lot of resources. User models for personalized search systems are commonly built from the user-created metadata trails that the users leave on the Social Web. These metadata have the advantage, that they are accurate (as they are created by the users themselves) and they provide the gist of the document they were extracted from. They are good replacement for an overlay model, but they are sparse. The metadata may also be extracted automatically, by applying natural language processing methods, but the automatically extracted metadata is inaccurate. When this type of metadata-based user model is used to link similar users, the inaccuracies present a problem. There is a need for methods that can either provide clear metadata, or that are able to reliably link users based on the noisy metadata.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] White, R. W., Bennett, P. N. and Dumais, S. T.: Predicting short-term interests using activity-based search context. In *Proc. of the 19th ACM int. conf. on Information and knowledge management CIKM '10*, pp. 1009–1018, 2010.

[2] Lawrence, S.: Context in web search. *IEEE Data Engineering Bulletin*, Vol. 23, No.3, 25–32, 2000.

Photo Album Visualization as a Collection of Memories and Experiences

Michal LOHNICKÝ*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
lohnicky.michal@gmail.com

Digital photography has existed for nearly 20 years (since 1991). About ten years later the first mobile phone with an integrated camera was manufactured and nowadays more than 90% of mobile phones have a built-in camera. More than 36 billion photographs were uploaded to the biggest social network Facebook during the year 2010 to share users' social activities.

The character of mobile photography or the way how Facebook users share their photos proves that for most people a photo gallery is a way how to present experiences from their holidays, trips and events. In most cases, taking pictures is a process of saving memories via a camera and therefore the visualization of photo albums should communicate these memories to make events unforgettable [1]. Thus it is important to create visualizations providing a global overview of the experienced events, which helps users to

- *recall memories* – for those who were attending the events,
- *recognize memories* – for those who were not attending the events.

As times go by, recall is becoming recognition because users are forgetting their memories. This is the reason why recognition is also important for those who were attending the events.

Our visualization method provides users with a global overview of photo albums, shows them what and when happened, what were the most important events, reproduces storytelling etc. For this purpose we have proposed the *Chart of Interests* as an abstract of a photo album (see Figure 1). *The Chart of Interests* is a waveform placed in a chart where the x-axis denotes time and the y-axis denotes the level of interest of events in the specific point in time.

The method is based on the idea that the more pictures a user takes the more interesting an event was. The interest function in a given time is computed as the density of taking pictures during a time period. This is naturally true because when

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 11–12.

there is no interesting event happening, people take no or only few pictures in a long period of time. On the other hand, when there is a really interesting event there is often a need to document every second of the event.

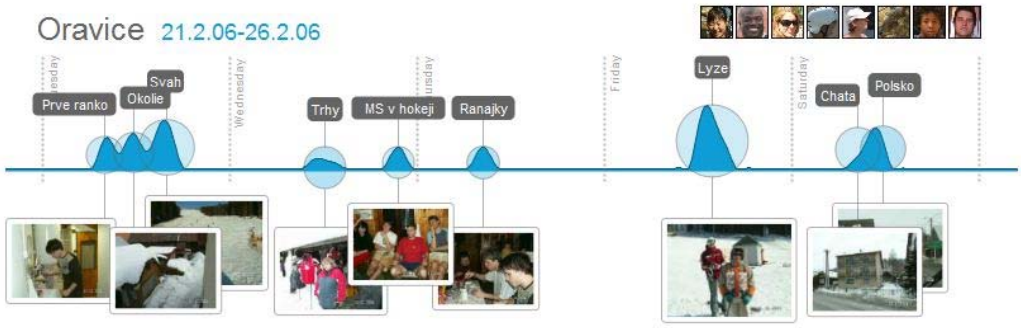


Figure 1. The Chart of Interests as an abstraction of a photo album.

The Chart of Interests is a graphical overview of a photo album where users can see a photo album in context. This visualization method utilizes the following features:

- *identification of events across a photo album*
- *chronological order of events*
- *identification of most important events of time periods in a photo album*
- *photo album placement in a time context*
- *storytelling*
- *navigation element which can be used supplement single photo browsing, photo album on a map presentation, events commenting etc.*
- *identification of participants of events*

The main goal of our work is to supply innovative navigation and browsing in photo albums which supports recall mostly via storytelling and also recognition of digitally saved experience. This style of browsing photos can be used by users for sharing photos more efficiently, for finding photos which they miss in their photo collections, to view places where they intend to go in various time periods etc. It is also usable in commercial sphere - in travel agencies or botanic monitoring.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Chen, J. and Hibino, S. Reminiscing. View: Event-Based Browsing of Consumer's Photo and Video-Clip Collections. In: *Proc. of the 10th IEEE Int. Symp. on Multimedia*, ISM. IEEE Comp. Soc., Washington, DC, pp.23-30, 2008.

Acquiring Metadata from the Web

Milan LUČANSKÝ*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
lucansky06@student.fiit.stuba.sk

Our work focuses on mining relevant information from the web pages. Unlike plain text documents, web pages contain another source of potentially relevant information – easily processable mark-up. We propose a method to keyword extraction that enhances Automatic Term Recognition (ATR) algorithms intended for processing plain text documents with analysis of HTML tags present in web documents.

The keyword extraction using an ATR algorithm is domain specific, because with different collections it yields different results. Different ATR algorithms use different measures, which are based either on statistical or probabilistic approach while some algorithms combine both approaches [2]. The extraction of keywords from web pages is even more specific, because they typically cover topics from various domains and a variable number of pages relate to every topic. Such diversity of topics usually reflects into extraction of less descriptive keywords. An approach to keyword extraction from web pages could possibly benefit from other sources of information that the Web offers. The challenge is to consider web mark-up and to make use of emergent semantics that HTML tags represent [1].

Advantages of ATR algorithms, such as the ability to extract most relevant single and multi-word terms or processing plain text, seem to be appropriate for textual content in web environment. But none of them (to our best knowledge) considers the structure of document as potential source of additional information to find the best candidates for keywords. In our approach we combine the ATR algorithms with the processing of HTML tags.

We aim to enhance a way how a final weight of the candidate term is computed. We introduce a *TagRel* coefficient that modifies weight of a term obtained by an ATR algorithm according to the relevance of HTML tag enclosing the term. Our method for keyword extraction consists of the following steps:

1. Web structure preprocessing.
2. Term extraction.
3. Keyword selection.

* Supervisor: Marián Šimko, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 13–14.

In the first step we analyze the link structure of examined web pages to obtain information about other pages present "outside" the pages. When examining a particular page, we focus on the anchors of links pointing to that page from the rest of pages in order to extract terms describing the page. We can either crawl pages (in case of closed corpus such as a website) or leverage already existing indices of web search engines (in the case of the open web, e.g., by using '*link:*' operator provided by Google). From the crawled page we also obtain HTML mark-up which emphasizes some words. The most interesting words are visually distinguished from the rest of textual content, such words are enclosed in tags: **, *<b>*, *<i>*, **, *<h1-6>*, *<li>*, etc. Widely used and popular in styling web page content are Cascade Style Sheets (CSS), therefore we analyze the linked .css file, in order to obtain additional information from examined web page. Another possible option is to analyze targeted advertisements placed on some web pages.

Having web structure analyzed, in the second step we extract terms (i.e., candidate keywords) and compute weights reflecting their significance for a document. We modify the weight obtained by an ATR algorithm by *TagRel* coefficient bound to a tag enclosing the term:

$$w_i' = w_i \times TagRel_T \tag{1}$$

where w_i' is improved weight of a term *i*, w_i is weight of a term *i* obtained by a ATR algorithm and $TagRel_T$ represents relevance of a tag *T* that encloses term *i*.

TagRel varies among different HTML tags. We consider tag importance as an indicator of how much a tag is important for a page as some HTML tags are more likely to contain keywords (terms) than others, therefore their value assigned to *TagRel* will be greater. We have an ambition to parameterize value of *TagRel* according to factors by which an HTML tag or CSS element occurs on web page. The value of *TagRel* will be dependent on the count of occurrences of an HTML tag or the length of phrase emphasized by a tag/css element.

Evaluation of our method will be performed on set of web pages from different domains.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Hodgson, J.: Do HTML Tags Flag Semantic Content? *IEEE Internet Computing*, Vol. 5, No.1, 20-25, 2001.

[2] Knoth, P., Schmidt, M., Smrž, P., Zdráhal Z.: Towards a Framework for Comparing Automatic Term Recognition Methods, In: *Znalosti 2009*, pp. 83-94, 2009.

Interactive Browser of Heterogeneous Web Content

Peter MACKO*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
mr.petermacko@gmail.com

Multimedia content is very important for many users since it gives them fun and relaxation, but may also serve to improve education and skills. Just a few years ago it was extremely difficult to transmit large amounts of data, required by multimedia content, through the Internet. The arrival of new revolutionary technologies with higher speed of data transmission has changed this situation significantly in recent years. Therefore multimedia content is ever more available on the Web. Its main advantage is that the quality of the transmitted content, thanks to the constant acceleration of transmission speed, may still improve.

The main topic of my bachelor thesis is focused on displaying and viewing multimedia content and on the improvement of existing photo browsers and viewers. The browsers nowadays include the possibility of browsing and searching images based on specified criteria. The first theme of my project concerns the integration of a video player into an existing multimedia exploration solution since videos will become an integral part of it. Besides that, these videos will also include metadata which will be useful for easy searching and navigation in video content. Videos displayed by the system will be transmitted through the Internet in the highest possible quality.

These videos will be used mainly for study purposes. There are already some doctoral thesis defences, which have been documented in the form of video recordings. We would like to include these videos into the system and use the options they offer. Our system allows users watching these videos to save their notes at specific time frames, and thus interact also with each other by sharing annotations while learning or commenting events in the videos.

We also give users the opportunity to add chapters to videos. Thanks to this feature user navigation in video content is much easier. Via previewing the timeline users can immediately see major topics and parts of video presentations. We also allow users add questions to topics covered in video. These questions are divided into two categories:

* Supervisor: Michal Tvarožek, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp.15–16.

- Open questions where users write answers into prepared text fields. All replies are stored so the question sponsor will know who wrote each answer. This way the author of a video can check for the understanding of the information presented in the video.
- Questions with default answers, with several predefined options, so the question answering will be quick and easy thus not unnecessarily disturbing users.

We realize the question answering interface in a user-friendly way so that the announcement of questions or comment and their addition/answering do not significantly disturb user experience.

If users want to answer a question, we either automatically pause the video or present the question in background such as not to disturb watching. We expect our videos to be much more documentable and searchable, and thus more useful both entertainment and educational purpose s via comments, questions and chapters.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Sood, V.: *Building your first Smooth Streaming Player using SSME*. 2009. [online]. Available at: http://blogs.iis.net/vsood/archive/2009/10/14/building-your-first-smooth-streaming-player.aspx

[2] Lacko, L.: *Silverlight – introduction to interactive application development*. Brno, C-Press, 2010. 464 p. (in Czech).

Semantic Web Navigation Based on Adaptive Views

Karol RÁSTOČNÝ*

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

The amount of information in web repositories is growing exponentially. This natural evolution affects the number of identified results for user queries, which can decrease the relevance of results. Users can increase the success rate of their query for conventional search engines (like Google or Bing), when they can exactly describe required result with keywords. This problem is quite successfully solved with keyword-based query expansion [3] and with approaches based on exploratory search [1]. But it will often happen that a user has already found the required result and wants to find similar and/or related results. One type of this service is provided by Google Similar, but results offered by this service are only displayed in list view without any information, why they are evaluated as similar (e.g. why the second result for a similar image to cotton bolls photo is the photo of a dog). In our approach we address this problem via view-based search using zoom-based graph navigation.

Our graph navigation approach is based on bachelor work of Adrian Rakovský [2], where the concept of web browsing based on graph visualization was proposed. The main benefit of this approach was that users could see dependencies between the resource (original result) and new results visualized in a graph. Despite the advantages, this approach also introduced new challenges: the visualized graph can quickly grow to enormous size and become unclear and unusable for conventional users. To avoid this we extend this approach with result clustering, facet marking, and adaptation to user interests, next action recommendation and zoom-based navigation. The first three extensions aim to reduce the number of nodes displayed in the graph. The main task of next action recommendation is to help users with orientation in graph with highlighting of nodes, with which other users interactive under similar conditions. Our zoom-based navigation approach offers four types of graph visualizations: literal attributes graph view, object attributes graph view, restricted RDF graph view and full RDF graph view. These visualizations gradually increase the detail of displayed information (see Figure 1).

* Supervisor: Michal Tvarožek, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 17–18.

Users employ different strategies to find and explore information on the Web such as lookup, learning and investigation [1]. These kinds of user search activities cannot be easily separated as they are often performed in succession. When users start searching for information they do not know the information space very well. Users then often redefine queries while learning about the identified results. Once users have found the desired results and learned basic information about them, they often switch to browsing of related resources (e.g., when a user finds a painting, he often explores its author), so they need support for navigation in them. We address this problem via navigation based on adaptive views that helps users navigate from searching to browsing among related and similar results. This navigation starts with hierarchical cluster-based result browsing, where we display results of a faceted browser classified in hierarchical clusters. After identification of the desired results, details about them are displayed in adaptive item details view and lastly discoverable via zoom-based graph navigation.

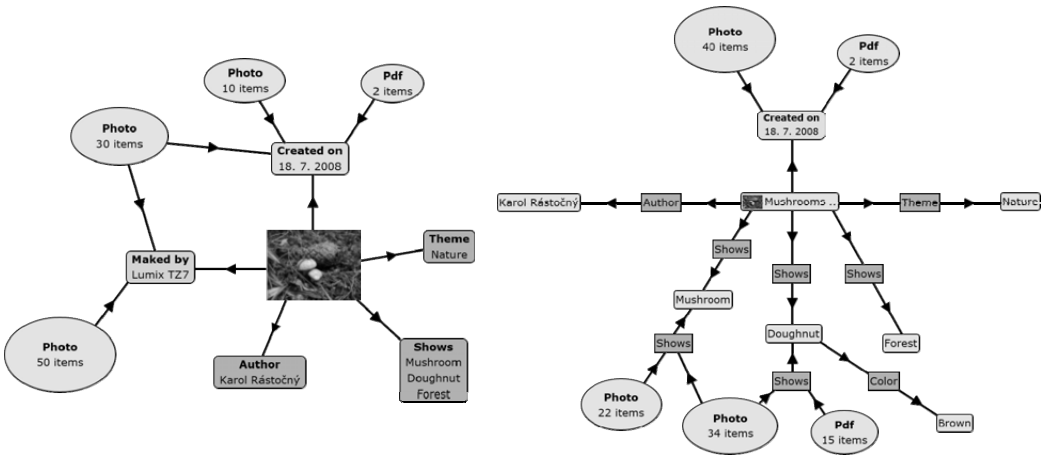


Figure 1. Visualization of object attributes graph (on the left) and RDF graph (on the right). Object attributes graph contains less nodes and is better readable, but RDF graph allows users to change context and shows attributes of objects in values of the original object's attributes.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Marchionini, G.: Exploratory search: from finding to understanding. *Communications of the ACM - Supporting exploratory search*, Vol. 49, No. 4, 41-46, 2006.

[2] Rakovský, A.: *Web browsing based on graph visualization.* Bc. thesis, Slovak University of Technology in Bratislava, 2009 (in Slovak).

[3] White, R.W., Marchionini, G.: Examining the effectiveness of real-time query expansion. *Information Processing and Management: an International Journal*, Vol. 43, No. 3, 685-704, 2007.

Classification and Recommendation

Information Recommendation with Use of Context in a Specific Domain

Anton BENČIČ*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
bencican@live.com

Our every action, wish or preference is shaped by our context, be it short-term or long-term. Studies about mobile devices, which have long passed on being just mobile phones, show that people use them and the connected services they offer even when they are in a presence of a computer [1]. Besides the basic information or context like location, orientation, browsing history or installed applications that can tell a lot about the user, we can access the user's data from social networks like Facebook or Twitter to get even more of the context.

The vast amount of information about user's environment and current whereabouts provide foundation for a number of applications that make use of it. The problem however with most of them is that they only choose a small subset of this context and act upon it. This subset includes most often only the location of the user, which while being useful in certain scenarios, cannot fill up for the whole picture and for example a pattern matching is often infeasible or annoying for the user when put in practice because of its inaccuracy [2]. Few methods [2, 3] today are attempting to gather context information from multiple sources such as mobile context, social context and smart space sensor context, and combine them together. By leveraging more than a location or a time by itself, these contexts are much more capable of recognizing patterns in the user's behaviour [2].

In addition to combining context from multiple sources to get a better picture of the user's current situation, selection of relevant context information is also important, because using (regardless of current situation) every piece of the context information introduces the same problems. The context-aware recommenders must strive to make use of the entire relevant pieces of the context information in order to perform effectively and efficiently. On the other hand, choosing the relevant context is not a straightforward problem as it varies among users, making room for group formation and collaborative methods.

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 21–22.

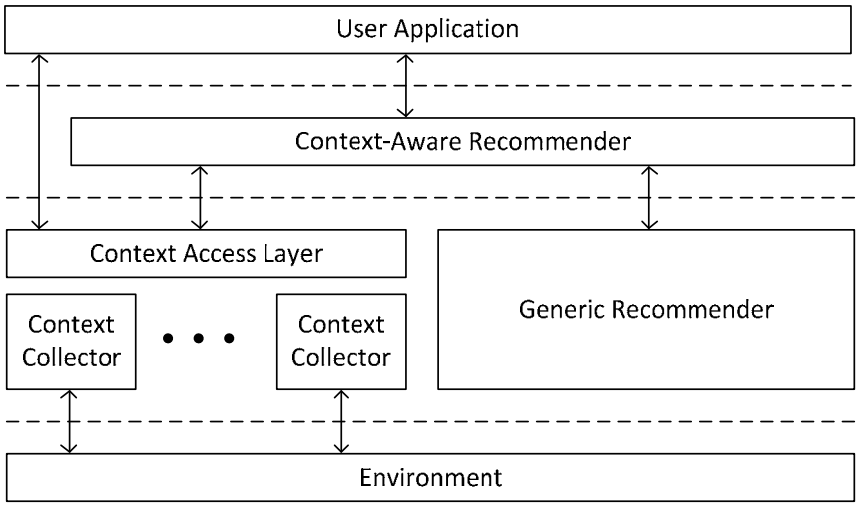


Figure 1. Context Fusion and Context-Aware Recommendation.

Our aim is to create a context framework that will provide applications with access to user's context through a unified and transparent interface, allowing the applications to choose the recommendation method as well as the context subset they want to use. The overall architecture is shown in Figure 1 where the aforementioned context framework consists of an extensible set of raw context collectors, context access layer and a context-aware recommender.

With the context framework at hand we will perform a number of experiments with one or more specific domains, like an internet newspaper or a digital library in order to see what context subsets are relevant in those particular domains and with what combination of recommendation methods we can achieve the best results.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Fry, M. et al.: PersonisJ: Mobile, Client-Side User Modelling. In *Proc. of 18th Int. Conf. on User Modelling, Adaptation and Personalization*, Springer-Verlag Berlin, Big Island of Hawaii, HI, USA, pp. 111-122, 2010.

[2] Beach, A. et al.: Fusing mobile, sensor, and social data to fully enable context-aware computing. In *Proc. of 11th Workshop on Mobile Computing Systems & Applications*, ACM, Annapolis, MD, USA, pp. 60-65, 2010.

[3] Dongmin S. et al.: Context-Aware Recommendation by Aggregating User Context. In *Proc. of Commerce and Enterprise Computing*, IEEE, Vienna, Austria, pp. 423-430, 2009.

Personalized Recommendation and Context-Aware Avatar for Motivating Children to Physical Activity

Pavol BIELIK, Peter KRÁTKY, Štefan MITRÍK, Michal TOMLEIN*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
icup_fiit2011@googlegroups.com

Over the last few decades, overweight and obesity became a problem of global proportions. Insufficient amount of physical activity and unhealthy dietary habits have been cited as its primary causes. While measures that need to be taken have been known for a long time, their execution has been hindered by a lack of motivation present in all age groups.

Tracking activity is an important step towards improvement. Most current activity tracking solutions rely on stand-alone single-purpose measurement and tracking devices, which are not very popular among users [2]. The current generation of smartphones has made it possible to track movement using built-in sensors such as GPS receivers, accelerometers and gyroscopes. We use smartphones of this class to provide an activity tracking solution.

Using a device people already carry, we are now able to collect data and provide the user with statistics and recommendations, which are a powerful source of motivation. However, a different approach to motivation is necessary for children, who are not easily motivated by figures and charts. Our solution is based on rewarding, a natural and effective way to motivate. We provide parents by means to specify short-term and long-term, non-repeating or recurring rewards for their children. Children exchange points they have earned by being physically active for these rewards.

When performing a physical activity, one requires a certain extent of knowledge on the right kind and amount of activity. Not enough activity may result in no improvement while too much activity may be harmful. It is therefore vital that users are advised on these matters. Our solution provides the user with advices and recommendations. We use the concept of a daily plan to give recommendation on the type and amount of activity for a given day. Daily plans are generated automatically

* Supervisors: Mária Bieliková, and Michal Barla, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 23–24.

based on several factors, such as the current weather conditions or previous activity, but also take into account fulfilment of the previous plans.

Much like motivation, advising children requires some additional consideration. Animated pedagogical agents have proven to be effective in capturing children's attention. We propose an animated character (avatar), which is fully customizable in terms of appearance. The main role of the pedagogical agent is to provide a better communication interface between the system and the child, which is necessary in order to provide the child with educational content, such as advice on healthy lifestyle and exercise. We also use a social networking model to incite competition among friends in order to provide an even stronger source of motivation [1]. In conjunction with competition, motivation is also achieved through engaging data visualization, which provides an overview of the accomplished progress.

While the general concept of our solution is not limited to a specific age group, we have chosen to target children as our primary focus group for a number of reasons. Most importantly, the differences in the approach that needs to be taken with younger users are too numerous and require a fundamentally different design of the overall solution. In particular, the described means of motivation we employ could not have been implemented as an afterthought. Moreover, experience with children suggests that if any measurable progress is to be achieved, the involvement of parents in the process is necessary. In addition to being the source of rewards for children, we provide means for parents to oversee their children's progress and interfere when and where appropriate.

To achieve the above goals, we have created a Move2Play system consisting of a mobile phone application for children and a web application for parents. The mobile application serves to track activity and provides recommendation and advice through the avatar. It is a social hub designed to connect friends and encourage competition. It also gives children access to rewards provided by their parents. Parent application, provides parents with information on the progress of their children and activity recommendation. It is also the entry point to specify rewards for children.

Move2Play takes a new, unique approach to solving the problem of sedentary lifestyle. It brings together several ideas and concepts such as activity tracking, motivation through rewarding, motivation through competition, activity recommendation and, specifically for young users, an animated pedagogical agent and the involvement of parental guidance.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Campbell, T. et al.: Game Design Principles in Everyday Fitness Applications. In *Proc. of CSCW* 2008, San Diego, CA, 2008.

[2] Consolvo et al.: Design Requirements for Technologies that Encourage Physical Activity. *In Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, 2006.

Information Integration on Social Adaptive Web

Michal HOLUB*

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

Today almost everything can be found on the Web. The problem is that information is scattered across many sources. No web page contains everything; therefore there will always be a demand for more information in addition to what is presented to the user at a time.

Around 66 % of searches are triggered after visiting some web page [2]. This means that a user requires additional information to supplement what he/she reads on a single web page. We can anticipate this behavior and identify important entities in the text using methods for named entity recognition. The recognized entities are candidates that might need further explanation. We can search for more information about them in advance, analyse the results and integrate interesting facts with the current web page. This is a form of automatic text annotation. The user then sees an enriched page with more information.

A use case scenario involving integration of information from various sources is proposed in [1]. The core is formed by a blog with movie reviews. Each review is automatically enhanced with information from other sources like pictures from movie's official website, biographies of the actors and director from an encyclopaedia or schedules of local cinemas playing the movie. We could also employ the user's social network to recommend the movie to people who might like it. The challenge the authors present is to persuade the owners of websites to include more semantics in it. They also call for a wider adoption of standards like RSS.

Search engines are effective in locating the documents which contain the keywords from a query. It is then up to the user to go through these documents, find relevant pieces of information and join them together in order to get a broader view of the problem. However, search engines fail in presenting relationships among information contained in returned documents. We try to accomplish this by integrating information across many websites and creating mashups of them. We are working on a method for combination of web objects extracted from different websites. The

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 25–26.

combination may answer user's search query and/or it may provide user with further information related to the current web page.

We have done work in the domain of integration of events and grouping of news articles. In the first case we analyzed the web portal of our faculty and extracted information about upcoming events. Using implicit feedback we determined the interest of users in those events. Then we created a personalized calendar from events, in which the user was interested, and displayed this calendar on the website.

In the second case, we grouped news articles from different news providers. Our goal was to group articles informing about the same occasion together. This is useful when we want to have an overview of events which happened during the day. Each event has a list of articles published by various providers and the user can select which source he/she wants to read about the event. This is similar to NewsBrief project [3], which we used as a baseline.

NewsBrief uses k-means clustering to group articles. However, there are often errors in the results, where they put articles about different events to the same group. We downloaded 106 articles which were divided into 24 groups. We then manually evaluated the results; one group contained articles about 8 different events.

In our method we use keywords to compute the similarity between two articles. The similarity expresses how probable it is that two articles inform about the same event. We used Metall web service developed at our faculty for keywords extraction. It uses third party services to translate the article to English and to find relevant named entities and keywords in the text. We compared keywords of an article A with keywords of already grouped articles. If the number of identical keywords between two articles with the maximum keywords intersection was higher than a threshold, we put these articles in the same group. If there was no article with sufficient similarity to the article A, we put the article A into a new group.

We experimented with various values of threshold. We achieved the best result with the value of 40 % for the threshold. 4 articles (out of 106) were put into wrong groups. However, using our method we split the group with articles about different events into correct groups. In this way we outperformed the baseline.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Ankolekar, A. et al.: The two cultures: mashing up web 2.0 and the semantic web. In *Proc. of the 16th Int. Conf. on World Wide Web*, ACM Press New York, NY, USA, pp. 825-834, 2007.

[2] Cheng, Z., Gao, B., Liu, T.: Actively predicting diverse search intent from user browsing behaviors. In *Proc. of the 19th Int. Conf. on World Wide Web*, ACM Press New York, NY, USA, pp. 221-230, 2010.

[3] Steinberger, R., Pouliquen, B., Goot, E.: An introduction to the Europe Media Monitor family of applications. In *Information Access in a Multilingual World – Proc. of the SIGIR 2009 Workshop*. Boston, MA, USA, pp. 1-8, 2009.

Group Recommendation for Multiple Users

Michal KOMPAN*

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

Group recommendation is presently an interesting research area. There are several activities, which we are doing in a social rather than an individual manner. In this situation individual recommender system cannot be applied. TV watching, going to a cinema, a restaurant or a pub are only some examples. These activities are usually attended after some agreement in a group. We also distinguish situations, where we cannot choose such as music played in a gym or in public transport vehicles.

The user model or some user preference representation is an inherent part of recommendation. In the case of the group recommendation three types of group modeling are used – individual user's preference aggregation, single recommendation aggregation and group profile construction. The individual user's preference aggregation is the most used and studied approach, because of several advantages (e.g., the group can change during the recommendation process)

Various approaches for the data acquisition have been used. Standard implicit and explicit feedback is often used in single recommendation approaches. Also new approaches as sharing preferences and negative feedback were proposed for group preference modeling.

When recommending to a group as a whole, several aspects should be considered. We distinguish between active and passive groups. When recommending to active groups, we can delegate some tasks to users themselves as they can achieve some level of consensus [1]. Group homogeneity and size are also important factors, which have great influence on the recommendation process and user satisfaction. Group homogeneity refers to the group structure (students of informatics in PhD. degree, or visitors of a gym) [3].

The group recommendation task can be extended for not only one item recommendation, but the recommendation sequences (ordered sets) of items [2]. This is very useful in order to satisfy more users and there is no content which is suitable for the whole group. Then we can partially satisfy every user separately (other users on a lower level).

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 27–28.

There is a difference in how recommendations are presented. A recommender system can generate only some kind of "suggestions" - the final decision is still on the group. On the other hand, the recommended items can be experienced by users immediately (TV program) [2].

Sequence recommendation is an interesting task, especially in group recommendation, where we have to consider not only the order of the sequence for individual users but also its influence on other group members. This is strongly connected to designing satisfaction functions, which should model the satisfaction level of the group.

As we try to model real life group characteristics, it is important to incorporate user personality since it was shown that a user's mood and personality could have a significant influence on other group members' feelings. In other words, when a respected extrovert is unsatisfied, other members will probably share her feelings, even if they were partially satisfied before.

Group recommendation can also be used for solving problems of standard single recommendation. Using multiple criteria usually complicates recommendation since several attributes must be evaluated. Merging strategies can be used for overcome the multiple criteria problem, while only some modifications are needed (e.g., not considering fairness) [1]. Most of the today's recommenders suffer from the cold start problem - the state when new users come and want to interact with a system and there is not yet enough information about them. In this case, group recommenders (the group consisting of new users and all or several representative old system users) to solve this problem.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Ricci F., Rokach L., Shapira B., Kantor P. B., editors. *Recommender Systems Handbook*. Springer, 2011.

[2] Masthoff J. Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Modeling and User-Adapted Interaction,* Vol. 64, No.14, Springer, 37-85, 2004.

[3] Masthoff J, Gatt A. In pursuit of satisfaction and the prevention of Embarrassment: affective state in group recommender systems. *User Modeling and User-Adapted Interaction*. Vol. 16, No. 3-4, 281-319, 2006.

Recommendation and Collaboration through Implicit Feedback

Martin LABAJ*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
martin.labaj@computer.org

In the field of e-learning, the identification of difficult and/or interesting parts of learning text can be a useful feature for tasks like rewriting learning text, displaying focus or providing adaptive help to student. In our work, we track implicit feedback/interest indicators including scrolling (read wear [1]). Using data collected from many users, we can determine which fragment of the document is the most time-consuming and therefore interesting and/or difficult. As in any method dealing with time-based user tracking, there is a possibility that the user is pursuing different activities during assessed time periods. We try to avoid this by using low-cost webcam and employing physical user tracking – gaze tracking. This way we can exclude time periods when the user is not directly using computer or when he/she is at the computer but working with different application. The gaze detection also increases precision of fragment identification as an additional implicit interest indicator [2].

Collected interest data combined into an attention index can be used in various scenarios:

- *Interesting fragments visualization and summarization*, only fragments with highest attention index are highlighted or selected. Users can quickly scan through document either on first read or on revisit.
- *Adaptive guide to (learning) application*, user's work with web system is evaluated and adaptive hints are provided. If user notices recommended items with his gaze, but does not use them, different advice is provided than when he did not notice recommendations at all. Also explicit feedback questions can be asked the same way. Hints or questions adapted to current situation should make the user respond more easily.
- *Augmented instant message communication*, users are provided with reading positions of their peers in the same document. Students observe who is possibly stuck on the same fragments and cooperation can be encouraged.

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 29–30.

We implement gaze tracking using open source gaze tracker OpenGazer as a standalone C++ and .NET desktop application due to webcam access and required processing power. We collect interest indicators via custom extension of Firefox web browser (Fig. 1). This extension is connected to the gaze tracking via local client-server communication using sockets. In order to save web application's resources, we filter and process collected feedback in the extension, but unprocessed feedback is also stored for offline analysis and review on an independent server.

We have already partially evaluated the gaze tracking alone and incremental parts of the implementation. Currently we are working towards the evaluation of a complete solution via ALEF Adaptive LEarning Framework and possibly on the open Web.

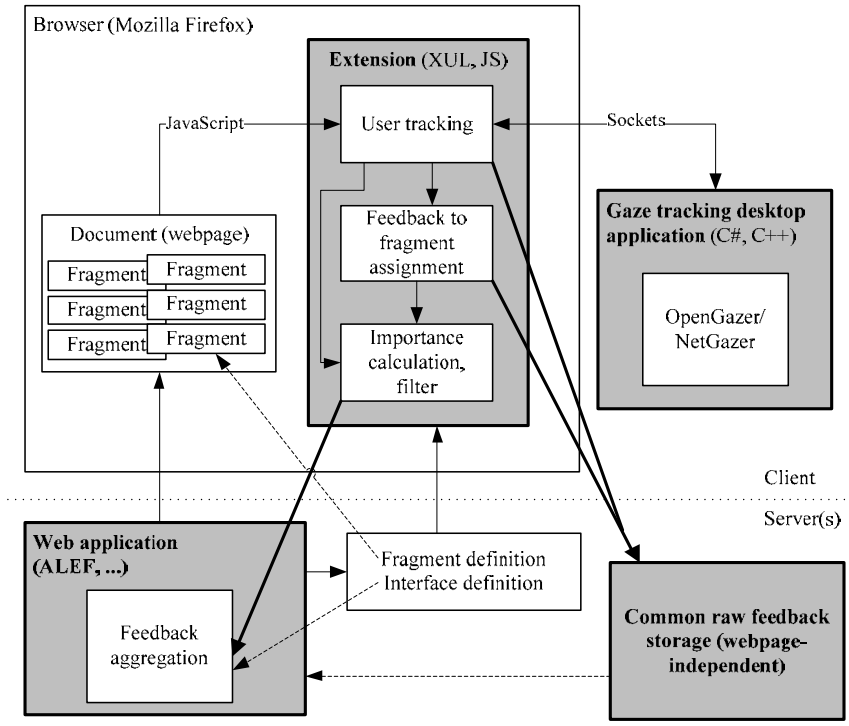


Figure 1. Overview of feedback collection architecture and technologies used.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T.: Edit wear and read wear. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. ACM Press, 3-9, 1992.

[2] Xu, S., Jiang, H., & Lau, F.: User-oriented document summarization through vision-based eye-tracking. In *Proc. of the 13th Int. Conf. on Intelligent User Interfaces*, ACM Press, 7-16, 2009.

Leveraging Microblogs for Resource Ranking

Tomáš MAJER*

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

Nowadays, the thinking of modern people and basically of the whole society is affected by the Web and many people benefit from the Internet. Its most important service – the Web – no longer consists of only static pages that people browse and search information in. User generated content has become more important than ever, which caused that a new medium for publishing brief posts has emerged – microblogging.

Microblogs are a lightweight form of traditional blogs, where users publish only brief reports called tweets. This phenomenon has become popular mainly thanks to Twitter, which enabled users to communicate with each other in an entirely new way through microblogs. They can share their actual feelings and experiences as well as opinions. The opinions are often linked to various entities, which typically represent personalities, products or events. By aggregating different views of many users on a topic, microblogs are considered a very valuable source of information.

Twitter represents a potential source of valuable user data. It is possible to monitor current users' opinions all around the world. Different world events can be monitored through microblogs [2] such as elections, putting a new product on market, or revolutions in what is very current and extensive user content. This content can be processed and utilized in order to improve access to information, e.g. by improving search. Currently search takes into consideration several criteria, where the most important are the content and structure, format, links with other sites or domain name credibility. By utilizing microblogs and by mining user opinions, it would be possible to obtain data coming directly from the users. Similarly, there are other areas where it is possible to use ranking according to user reviews. For example, product reviews and ratings can be modified by owners or manufactures in e-shops. Thinking of Twitter as of a separate service, it contains un-moderated user data as the user posts are not as biased and more objective.

Twitter is a service combining elements of both social networking and microblogging with certain specifics [1]. While user profiles in social networks are connected bi-directionally, connections on microblogs have only one-way orientation. Users may follow other users in order to track their posts, however, it does not

* Supervisor: Marián Šimko, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 31–32.

necessarily mean that also reverse links exist, i.e., a followed user may choose not to follow ones that follow him. In contrast with traditional blogs, post comments cannot be assigned directly to a post. However, a new post linking to another could be created (referred to as *retweets*). User profiles on Twitter are usually simple and public.

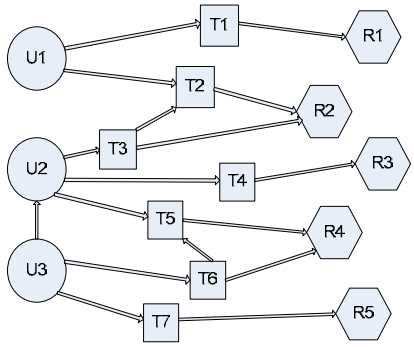


Figure 1: Basic Twitter graph representation.

The Twitter graph contains three types of entities (see Figure 1). The first type denotes users and they are represented by circles on the left side (U1-U3). Squares in the middle represent tweets (T1-T7). Hexagon represents resources (R1-R5) and they are located on the right side of graph. Otherwise there are also more tweets on Twitter which are not connected to pages.

Relations between users mean that a user follows another user. Tweets can be linked to users, another tweets and pages. Relations between tweets mean the tweets are re-tweets of parent tweets. Tweets with included text that contains links to external sites on the Web, create relation between tweets and pages.

We believe the structure of microblog posts represented by a graph (Figure 1) can be leveraged to determine ranking of resources referenced in microblog tweets. We propose a method for computing a resource rank, which we call TweetRank. TweetRank is derived from tweets topology by our method. The principle of the method lies in user ranking estimation and propagating such rank via graph relationships into web resources. By leveraging microblog specifics, we have presented how micro-blogging data can be used to improve web resource ranking traditionally based on resource content or resource links.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 345-032STU-4/2010

References

[1] Huberman, B.A., a D.M. Romero: *Social networks that matter: Twitter under the microscope*. First Monday, 2009: 8.

[2] Dong, A.: *Time is of the essence: improving recency ranking using Twitter data.* In *Proc. of the 19th Int. Conf. on World Wide Web.* ACM Press, 331–340, 2010.

New Approaches to Log Mining and Applications to Collaborative Filtering

Ján SUCHAL*

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

In era of information overload we do seek help in recommendation systems aiding us to focus our attention to items like products, articles [1] or websites, that might be relevant to our needs based on prior knowledge. Typically this is done by mining knowledge about our interests from logs of our previous activity.

Our work focuses on two main goals strongly related to recommendation engines. Novel approaches to log mining and the potential usage of these implicit data in recommendation systems, especially collaborative filtering exploiting implicit negative feedback data. In general, while interpreting implicit feedback from logs can be a challenging task [2], mining and interpreting negative implicit feedback from positive visit logs is much more challenging.

First, we present a method for mining sources and cascading graphs of viral visits from raw logs. Such information can be useful for evaluation of marketing targeting to detect influencers and potential sources of viral traffic. We detect users that visit pages via viral recommendations (instant messaging, email...) and look for users that visited such pages before and estimate probability of viral recommendation source. A probabilistic viral cascade graph can be reconstructed from such data. We also categorize types of sites for which our method can be used and experiment on real world dataset containing the massively viral start of foaf.sk service.

Second approach focuses on mining negative interests of users from basic server logs in the domain of news articles. We propose two different methods for mining negative feedback, the first is based on time-based identification of articles which users do not read, second is based on detecting such articles, whose title (or even an abstract) was probably seen by the user, but did not raise enough of interest to actually visit and read it. Such data can be used in addition to positive interest that are normally used for generating recommendations. Experiments on live traffic on largest Slovak news portal www.sme.sk show that incorporating such feedback into collaborative filtering

* Supervisor: Pavol Návrat, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. pp. 33–34.

recommender gains 8.5% higher click-through rates and lowers recommendation rejection rate by 5% when compared to baseline collaborative filtering algorithm [3].

Finally we present a novel method for linearly scalable nearest-neighborhood based collaborative recommender system using specially prepared fulltext indices [3]. Evaluation is done on datasets from largest Slovak news portal sme.sk and github.com recommendation contest. Comparison with graph-based spreading activation recommendation method shows comparable results in means of relevance (20% precision on top 10 list) and with superior scalability characteristics.

Future work focuses on combining content-based and collaborative recommendation approaches into a superior hybrid approach exploiting positive and negative implicit feedback.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Das, A.S., Datar, M., Garg, A., Rajaram, S.: Google news personalization: scalable online collaborative filtering. In *Proc. of the 16th Int. Conf. on World Wide Web*. New York, NY, USA, ACM, pp. 271–280, 2007.

[2] Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting clickthrough data as implicit feedback. In *Proc. of the 28th Int. ACM SIGIR Conf. on research and development in information retrieval*. New York, NY, USA, ACM, pp. 154–161, 2005.

[3] Suchal, J., Navrat, P.: Full Text Search Engine as Scalable k-Nearest Neighbor Recommendation System. In Bramer, M., ed.: Artificial Intelligence in Theory and Practice III. Volume 331 of IFIP Advances in Information and Communication Technology. Springer Boston, pp. 165-173, 2010.

Using Implicit Feedback in Recommendation Systems

Peter ŠTUDENT*

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

The amount of data on the Web is growing day by day and therefore it is still more and more difficult to find content that is interesting for the specific user. This problem has a major impact especially on the news web portals, with a highly dynamic content, with new articles coming very often, which increase a chance that a user will miss an article, which she would considered as important or be simply interested in.

The aim of our work is to analyze the contribution of negative implicit feedback observed from user behaviour to the process of searching for interesting content on the Web and generating automatic recommendations for that content. Main advantage of using implicit feedback in comparison with explicit feedback is not only that there is no unnecessary burden on users to perform redundant operations but implicit feedback brings also a possibility to capture short-term interest [1][2]. One of the indicators falling under negative feedback, which we mainly focus on is information about situations when user immediately leaves a page containing uninteresting content, without reading it.

The aim of introducing negative form of implicit feedback in the process of generating recommendations is to reduce the amount of data processing and increase the speed of generating recommendations compared to traditional systems without this kind of feedback. We have analyzed different forms of adaptation of negative feedback in the process of generating recommendations:

- Elimination of uninteresting items based on the common negative feedback from the group of users that have interest in a group of similar items.
- Elimination of uninteresting items with negative feedback from the group of similar users, where similarity is based on the fact that they have given negative feedback for group of similar items.

* Supervisor: Ján Suchal, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 35–36.

– Identification of groups of similar users, where similarity is based only on common negative feedback; recommendation of unseen items, which were positively evaluated by group members to other users of the group (Figure 1.).

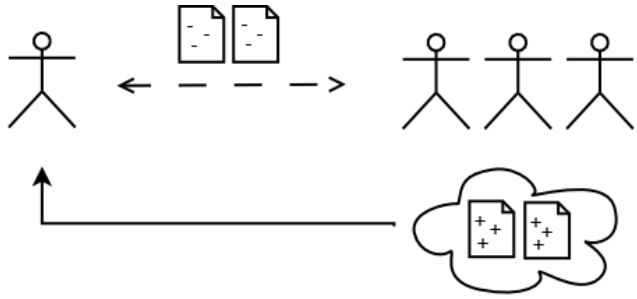


Figure 1. The method for creating recommendations using negative implicit feedback.

Proposed solutions negative implicit feedback utilization in process of creating recommendations are subject to experimental verification on dataset from the existing web news portal. We made a prototype of recommender system based on full text search engine k-nearest neighbour recommendation system [3] with incorporated use of described negative implicit feedback method.

In the current stage of the research we had done synthetic tests which have shown that the last of proposed solution can increase speed of generating recommendations by 5% with almost same quality of generated recommendations compared to traditional recommendation systems.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Claypool, M., Le, P., Waseda, M., Brown D.: Implicit interest indicators. In *Proc. of ACM Intelligent User Interfaces Conference*, 2001.

[2] Brusilovsky, P., Lee D., Reinforcing Recommendation using Implicit Negative Feedback. *User Modelling, Adaptation, and Personalization*, Vol. 5535, 422-427, 2009.

[3] Suchal, J., Návrat, P., Full text search engine as scalable k-nearest neighbour recommendation system. In *Artificial intelligence in theory and practice III, IFIP Advances in Information and Communication Technology*, Vol. 331, 2010.

User Modeling, Virtual Communities and Social Networks

Detecting User Communities Based on Latent and Dynamic Interest on a News Portal

Marián HÖNSCH*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
marian.honsch@gmail.com

One of the drawbacks of today's community-based collaborative recommender systems is that they group users based only on their aggregated similarity. Only those users are assigned to a community, whose profiles match completely the profile of the community. This prevents them from using the wisdom of crowd coming from users that match only parts of their interests. This drawback was also described in former research papers [1], [3]. We assume that recommendations coming from communities that address this issue can significantly improve the quality of recommendations. The demand for accurate recommender systems is very actual. Not only that users prefer personalized portals over non personalized ones, but internet companies can also raise their profits by having users spend more time on their web page thanks to the recommender and personalisation features.

Our target is to find communities that are defined by one particular interest. This community should include all users that share this interest. Depending on how many interests we have discovered in user's profile, she can belong to several communities.

We evaluated the whole approach on a news recommender system by recommending articles based on communities detected by our approach. We also performed an experiment to confirm our interest comparing strategy.

Interests are derived from the analysis of the domains corpus and are expressed as sets of words that have dense relatedness interconnection between each other. Next we cluster the interest based on our similarity metric and so detect communities. An interest of a user is deducted from the articles she has accessed before. Based on the time interval and frequency of accesses we can deduce long term and short term interests. We assume that keywords extracted from accessed articles determine her interests. Similar assumption can be found in [2]. Our interest extraction can be summarized in these steps:

1. Capture all articles accessed by a user,
2. Extract keywords from these articles,

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

Spring 2011 PeWe Workshop, April 1, 2011, pp. 39–40.

3. Create keyword relatedness graph by connecting the keywords based on their relatedness (i.e. semantic relatedness),
4. Apply algorithms of finding virtual communities on the keyword relatedness graph.

Detected keyword groups are the base of user interests. With a similar approach on all articles we can detect interests in the whole corpus. We refer to users interests as local interests and to interests detected in corpus as global interest. Local interests have lesser cardinality than global interests and they are subsets of global interests.

Detecting communities

A community includes users who have one particular interest in common. To find neighbours for one of users interests we compare it to all other identified interests. The neighbours are then all interests that match to a certain threshold. We define two strategies during interest comparison:

- local interest to local interest
- local interest to global interest

The first approach detects all users that have a similar interest to the selected user interest and therefore it is always specific for one user, compared to the second approach which groups interests relevant to a specific global interest. The community size is cut-off by a threshold or top N rule.

A certain boost to both approaches is to consider the article category specified by the news editor (i.e. Sport, Politics) from which the keyword was extracted.

Our main contribution is the novel approach to user interest deduction and the detection of virtual communities based on these interests.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Braak, P.T., Abdullah, N., Xu, Y.: Improving the Performance of Collaborative Filtering Recommender Systems through User Profile Clustering. In *Web Intelligence & Intelligent Agent*s, Washington, DC, USA, IEEE Computer Society, pp. 147–150, 2009.

[2] Grineva, M., Grinev, M., Lizorkin, D.: Extracting key terms from noisy and multitheme documents. In Proc. of the 18th Int. Conf. on World Wide Web, New York, New York, USA, ACM Press, pp. 661–670, 2009.

[3] Lops, P., Degemmis, M., Semeraro, G.: Improving Social Filtering Techniques Through WordNet-Based User Profiles. In User Modeling 2007. Vol. 4511, Lecture Notes in Computer Science, Springer, pp. 268–277, 2007.

Encouragement of Collaborative Learning Based on Dynamic Groups

Ivan SRBA*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
srba07@student.fiit.stuba.sk

Web 2.0 principles became very successful and brought a lot of energy into the development of web applications. One of the new trends is so-called social software. It uses the web as a broker which allows users to collaborate, communicate or share content and opinions [1]. Typical examples of social software are wikis, blogs or social portals. The rising popularity of these applications caused that many users with different interests and social contexts are connected via common applications. If we want these users to collaborate effectively we need to know how to successfully identify users' groups and help users to find appropriate collaborators [2]. This problem is especially important in the domain of Computer-Supported Collaborative Learning (CSCL). There are several methods which solve this problem but they usually use only one source of information about users and do not consider actual context. Also these methods suppose that teacher knows which attributes make collaboration more effective.

Our main goal is to propose a method for creating different types of study groups and observe their dynamic aspects. Basic process schema using the proposed method is displayed on the Figure 1. This method will be able to take many users' characteristics as inputs, i.e. interests, friendship with other students, knowledge of learning objects etc. Also, the proposed method will consider information about an actual student's position in the learning system. The output of the method will be various types of groups [3]. The group members will be for example friends, experts or novices in the problem area. In order to create these groups we will employ several methods (i.e. latent jigsaw method or methods for creating homogenous, heterogeneous and also mixed groups). Students in created groups will be able to communicate and cooperate with all available collaborative tools. We will observe dynamic aspects of created groups, especially how students use these tools to achieve their goals. The result of our observation will be the behavior patterns which can be used as an additional input for the proposed method.

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 41–42.

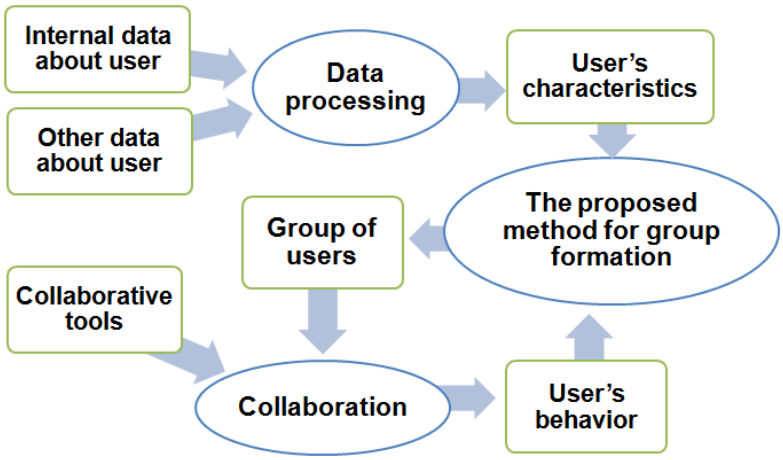


Figure 1. Basic process schema using the proposed method.

We will evaluate the proposed method in the e-learning system ALEF, Adaptive LEarning Framework. In the experiment we hypothesize that there will be a relationship between the group compositions and the way how students mediate collaboration. It means that the result of experiment will be recommendations and observed behavior patterns which can be used in any e-learning system to encourage students' cooperation with respect to available collaboration tools. This type of information can be very useful to improve design of e-learning systems which support effective collaborative learning.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010

References

[1] Rollett, H., Lux, M., Strohmaier, M., Dosinger, G., Tochtermann, K.: The Web 2.0 way of learning with technologies. *Int. J. Learn. Technol.*, Vol.3, No.1, pp. 87-107, 2007.

[2] Liccardi, I., Ounnas, A. et.al.: The role of social networks in students' learning experiences. In *Working group reports on ITiCSE on Innovation and technology in computer science education*, ITiCSE-WGR'07, ACM, New York, NY, USA, pp. 224-237, 2007.

[3] Wessner, M., Pfister, H.: Group formation in computer-supported collaborative learning. In *Proc. of Int. ACM SIGGROUP Conf. on Supporting Group Work*, ACM, New York, NY, USA, pp. 24-31, 2001.

Devising Secure Communication for Decentralized Environment

Márius ŠAJGALÍK*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
sajgalik@live.com

Many of contemporary personalisation systems are developed in the centralized way. The personalisation itself is done almost entirely on the webservers. That has several disadvantages, e.g., low flexibility due to too many data on single place, privacy. Another approach is a decentralized model, in which each client (an agent in the multi-agent system in essence) keeps a model of its user, determines what will be shared with others and personalizes the content and navigation to the actual needs of the user. This is called distributed or decentralized user modelling.

Previously, whilst commencing our research we analysed the most essential requirements for our work, which are:

- A network of distributed components must be able to adapt itself, especially because the same communication partners and technology are not always available.
- Information should be able to move among multiple users and platforms without the need for centralized controlling.
- How to locate an agent who has the relevant model with regard to the context and the purpose for which this model is required?
- How to make sense of potentially inconsistent, contradictory data?
- In general, how to interpret models created by other agents?
- How to ensure durability and the integrity of user models in this environment?
- How to ensure user privacy?

As the decentralized user modelling requires wide range of issues [1] to be addressed and those aforementioned are definitely not all of them, our research should not focus on every one in detail since there has already been done significant amount of work. Instead of that we must analyse and choose the most appropriate existing methods and helper tools. That will allow us to focus on the main goal of our work – to build up the

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

Spring 2011 PeWe Workshop, April 1, 2011, pp. 43–44.

decentralized client-side solution, which would provide the functionality similar to a personalized proxy server.

So far in our analysis we studied several approaches to ensure user privacy. The motivation for privacy care is to protect user private data and to encourage him to safer interaction, which has potential that user will use system more often and frankly without fear since he knows his data is protected and secure. This leads to more information about the user and hence a better basis for personalization.

There are multiple proposed standards for agent communication which have been already used in several works. Knowledge Query and Manipulation Language (KQML) is both a language and protocol originally designed as an interface to knowledge based systems, but later repurposed as an agent communication language. Another agent communication language is FIPA-ACL proposed by the Foundation for Intelligent Physical Agents. As the most recent, the platform for privacy preferences (P3P) protocol has been developed and officially recommended by the World Wide Web Consortium, which allows websites to declare their intended use of information they collect about users who can thus better manage their personal information while browsing.

To devise a communication infrastructure we discovered LoudVoice [2], which is an efficient multi-agent communication platform based on the concept of channelled multicast. Messages are sent on a channel and received by all agents that subscribe to it. Channelled multicast reduces the amount of communication needed when more than two agents are involved in a task. Moreover, LoudVoice presents several other features including the ability to distinguish streams of messages by their theme, and to address agents by their characteristics. Multi-agent systems can benefit from the possibility of broadcasting messages to a wide audience. The audience may include overhearing agents which, unknown to senders, observe conversations and, among other things, pro-actively send suggestions.

LoudVoice has been designed to support the notion of implicit organizations. An implicit organization is a group of agents playing the same role on a given channel and willing to coordinate their actions for the sake of delivering a service. The term “implicit” highlights the fact that there is no need for a group formation phase, since joining an organization is a matter of tuning into a channel. By definition, implicit organizations are formed by agents able to play the same role. LoudVoice allows senders to address messages either to specific agents or to all agents that offer a certain service on a channel, for example providers of a particular type of information.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Lorenz, A.: Agent-Based Ubiquitous User Modeling, In *Proc. of User Modeling*, Springer Berlin, Vol. 3538, LNCS, pp. 512-514, 2005.

[2] Busetta, P., Dona, A., Nori, M.: Channelled Multicast for Group Communications, In *Proc. of the Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems*, Bologna, Italy, 2002.

Modelling a Tutor for E-Learning Support

Peter SVORADA*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
psvorada@gmail.com

E-learning web systems allow students to educate themselves for example by studying materials, solving tests and doing exercises. Students take actions (text reading, answer inputting, text messaging etc.) which change the learning environment in which they work. It is proven that students can learn more if they are included in the learning process and this process is adapted to the needs of individual student. Likewise it is known that student advances faster in the learning process if he/she is guided by someone who acts more like a friendly tutor than a leading authority [1].

The architecture of an intelligent tutoring system can be divided into four parts: the domain model, the student model, the tutoring model and the student's interface. In this research we explore way for improvements of the tutoring model part of this architecture. The tutoring model receives input from both the student model and the domain model and makes decisions about which tutoring strategies and actions to take (Figure 1).

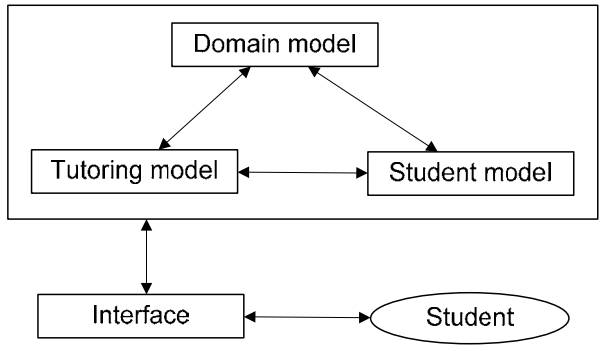


Figure 1. Typical components of an intelligent tutoring system.

These decisions include deciding whether it is time to intervene and when is the best time and how to do so regardless of whether it is a feedback on the correctness of steps taken – not only the final answers, providing error specific feedback etc. Selecting the

* Supervisor: Jozef Tvarožek, Institute of Informatics and Software Engineering

next steps in the learning process (choosing the content and an appropriate form of its presentation – many intelligent tutoring systems present content only as a plain text but some students might prefer different approaches like video, self solving exercises with narration etc.) is also part of the tutoring model's functionality. Other responsibilities of tutoring model include offering the context-sensitive next-step hints currently primarily called upon the student's request what presents itself as a possible area for improvement and automation.

The goal of this research is to propose improvements of the tutoring process, design them and implement a tutor model based on these improvements into an existing learning system [2], *Peoplia*, which is an interactive web-based environment that facilitates student learning by a socially intelligent tutoring agent.

The learning system also facilities for student collaboration, and the tutor can support and guide this collaboration to archive better results. This can be done by dividing students into appropriate groups (e.g. students who understand the problem can work with students who do not in order to help them advance faster) and encouraging them in helping each other in a way that is useful for all of them – tutor needs to analyze whether the feedback given by tutoring students is helpful for their tutees and help them in providing relevant help [3].

Tutors can also provide support in making the studying more user friendly. Students advance faster if they are guided by a friendly tutor not a strict authority or machine like teacher. Consequently tutor's communication and feedback patterns should feel natural to the students, posing another challenge in this research area.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Heffernan, N. T., Koedinger, K.: An Intelligent Tutoring System Incorporating a Model of an Experienced Human Tutor. In *Proc. of the 6th Int. Conf. on Intelligent Tutoring Systems,* Springer, pp. 596–608, 2002.

[2] Tvarožek, J., Bieliková, M.: Enhancing Learning with Off-Task Social Dialogues. In *Proc. of European Conf. of Technology-Enhanced Learning*, Springer, pp. 445–450, 2010.

[3] Walker, E., Rummel, N., Koedinger, K.: Modeling helping behavior in an intelligent tutor for peer tutoring. In *Proc. of 14th Int. Conf. on Artificial Intelligence in Education*, IOS Press, Brighton, pp. 341–349, 2009.

Combining Different Data-Sources for User Modelling in Personalized Learning

Maroš UNČÍK*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
maros.un@gmail.com

The trend of using e-learning systems is progressively growing and opportunities that the Web 2.0 provides are huge. Nowadays, e-learning systems offer more and richer content, enable communication and collaboration among users. The rise in use of these systems causes information overload. The adaptive e-learning systems are trying to address the most crucial issues, which are related to this overflow: (1) adaptive systems present only information, which is appropriate and/or interesting for him/her at the moment, (2) help user to choose the way to proceed when viewing content, (3) prevent user from getting lost in the content and/or avoid to prevent him/her from forgetting the original objectives.

To allow such personalization web-based e-learning systems monitor characteristics of individual users, including modelling of their skills, knowledge and/or interests. The performance of such systems is derived from an important element – the user model, which is used to minimize error rates and learning time.

A lot of problems in the domain of user modelling were identified. Combinations of several different inputs entering the user modelling process or the use of information about user beyond the adaptive system to enrich the user model are just two of them. Another challenge, the scrutability, concerning the visibility of the user model to users, is also closely related. In most of the systems, the user cannot directly access the user model and cannot provide explicit feedback about him/her, which could be otherwise taken into account. Much work has been devoted to resolve these problems.

Modelling of user model generally includes three steps [1] that take place in cycle (see Figure 1). In our work we deal with two of them: data collection and how it affects the user model. Our aim is to design a transparent user model, with data collection isolated from the construction of user model itself. We work with several sources of input for user modelling.

We also consider the visualization of user model from user's point of view, which allows direct and explicit feedback from students to enrich the user model.

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 47–48.

Additionally, the visualization can answer questions about what the system believes to be true, what it believes to be false and to find relationships between these beliefs, if they exist. It brings also another benefits, since many real-world user models are likely to be large, the visualization helps the user:

1. to get an overview of the whole model,
2. to get a clearer overview of dependencies in the user model, and
3. to adjust the sensitivity of the user model.

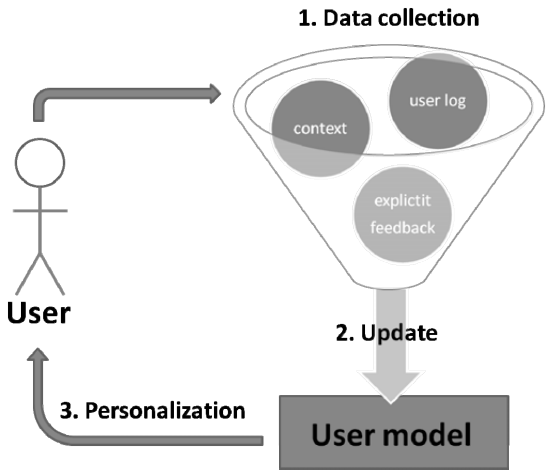


Figure 1. User modelling process can be divided into three steps: user data collection, updating the user model and personalization, which are repeated continuously.

We will experiment with the proposed user model in the Adaptive LEarning Framework (ALEF) in real-world setting, which is used as the e-learning system at the Faculty Informatics and Information Technologies at several courses.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Brusilovsky, P., Millán, E.: The Adaptive Web, chapter User Models for Adaptive Hypermedia and Adaptive Educational Systems, LNCS 4321, Springer, pp. 3–53, 2007.

Modelling Context Relations to Discover Hidden Contexts

Dušan ZELENÍK*

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

Recommender systems or even more general information retrieval systems have been deeply researched over the past decade. Almost every system, which works with documents, images, videos, music or different products had to adapt to new technologies, since there is a huge market competition and only the best stays and profits. Day by day we are inventing new technologies, which help us to live our lives comfortably. Searching the Web seems to be very easy task nowadays. We cannot even remember how it would be without indexed items over the whole Web and how long would a simple user spend by searching for a book he would find interesting. But there is always something new what will be invented and would make our lives even more pleasant. Current evolution of mobile devices indicates our need to be equipped by something smart always and everywhere. Smart devices are becoming our accompanies what also bring advantages we had never thought about. These devices are learning and adapting. They know who we are, what we do, where we are and even what we need.

Situational attributes like location, time, weather etc. are available now for devices whose only goal is to help us. We call these situational attributes simply *contexts*. Contextual information is not a new topic for researchers [2]. It has been already explored in many different areas of interdisciplinary research, but it all started with monitoring users in the scope of the Web. Nowadays this research around contexts rises again because of aforementioned computing power around us (generally called *pervasive computing*). Context is a new dimension in information retrieval as it is in recommending. Contexts such as location or part of the day are widely used to personalize systems. Contexts, and not only simple as location and time, are becoming more and more available what bring the research in the field of *user modelling* to the new and attractive perspective.

In this work we want to focus on the context and research, which has been done in this area. We want to bring an overview of the context and put it in constraints of new

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 49–50.

technologies, recent works. We want to show theoretical state-of-art approaches and link them with modern approaches, trends and perspectives. Our intention is to show new field and reveal possible research interests, which are summed up in the laid theses.

Web is a very specific place for contexts. We are able to recognize user interests by monitoring their interaction with websites. We could easily obtain information about current location (IP address), but there are more contexts which are not so visible. For now, we should abstract from tools of pervasive computing, since we are not able to determine low level contexts using computer. Hence there is new trend to browse the Web using smart devices, but still, we are never sure about complete set of current contexts of the user.

There has been some work done in classifying sentiment or opinion mining e.g., by Bermingham et al. [1]. It takes us to the idea of Web as storage of the users opinions. Furthermore, people spend many hours by contributing to Web what opens a new possibilities. By monitoring their contribution we could even discover contexts of their current state in general. To be more specific our intention is to extract emotion from the contributions. People use Twitter, Facebook or different tools to express their emotions or also other contexts, like what is new for them, where they are, what are they doing and what they feel. People are writing blogs, tweets, they do comment and like. This is all about nature language processing and learning the computer to understand humans. So far, we are introducing the *Web of Emotions*. Here we want to start the new approach in understanding emotions and how they affect user behaviour or interests on the Web. For instance, sad person would possibly like to watch a comedy or play a game to forget about the sadness.

Our goals here could be summed up as:

- Extract emotions from microblogs and social activities on the Web.
- Context of Emotion affecting interests.
- Reveal relations among emotions and other contexts.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Bermingham A, Smeaton AF.: Classifying sentiment in microblogs: is brevity an advantage? In *Proc. of the 19th ACM Int. Conf. on Information and Knowledge Management*, pp. 1833-1836, 2010.

[2] Kantor P., Ricci F., Rokach L., Shapira B. (Eds.): *Recommender Systems Handbook*. Springer, 2010.

Domain Modeling, Semantics Discovery and Annotations

Named Entity Disambiguation using Wikipedia

Martin JAČALA*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
jacala06@student.fiit.stuba.sk

The constantly growing amount of human written textual content available on the Web is a source of interesting and actual information about persons, organisations or places. One of the problems we face when analysing or querying in such content is name ambiguity. Does the word jaguar mean the sports car, the jungle animal or something different? Which Michael Jordan does the text refer to?

The proper names in news articles comprise approximately 10% of text and many of such proper names are ambiguous. In our work we propose an approach to answer these questions by disambiguating named entities using explicit semantics extracted from web-based corpora serving as background knowledge. We follow Miller and Charles distributional hypothesis [4] stating that similar entities appear in similar contexts even across multiple documents.

The problem of disambiguating named entities found in common, human written textual resources (usually referred as Named Entity Disambiguation) is a well established task in the natural language processing community. This task originated at the 6th Message Understanding Conference and has come a long way since then. Various approaches has been proposed over time, such as creating clusters of similar entities within set of given documents, or approaches specific to problem domain (e.g., geographical names or persons). However, these methods are rather limited when dealing with constantly changing open web data.

Using Wikipedia data as background knowledge for disambiguation has been proved successful by mapping entities on Wikipedia articles [1]. For each string containing an ambiguous entity they extract all articles that can be referred with the entity. We compute tf-idf cosine similarity measure with the ambiguous string for each retrieved article. The documents are further extended with term vectors from documents belonging to the same category. Evaluation of the system on various Wikipedia articles gives precision of approximately 80%. Similar approach use different context generation method together with secondary measure based on Wikipedia's category taxonomy, improving the precision to about 88% on Wikipedia articles.

* Supervisor: Jozef Tvarožek, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 53–54.

In our work we use the explicit semantics already present in Wikipedia data to extract all possible meanings of currently analysed entity. The network of redirects and page links helps us to resolve possible synonyms. We further use disambiguation pages to extract articles corresponding to various meanings of this entity (and any other extracted via redirects).

Instead of directly comparing the fragments of documents using similarity measure such as cosine similarity, we build a 'semantic space' using Explicit Semantic Analysis [3]. This method is similar to the Latent Analysis, however it does have certain assumptions on processed data. The most notable difference is that in ESA each document corresponds to an 'explicit concept' instead of inferring latent concepts from large, un-labelled dataset with LSA. The created semantic space is in fact a term-document matrix containing weighted list of Wikipedia concepts with respect to the individual keywords.

Further in the analysis process, we transform the analysed document and each of Wikipedia article retrieved in the previous step using disambiguation pages into the concept space. The document vector is computed as the running total of intermediate vectors computed while creating the semantic space. With vectors transformed, we compare them using cosine similarity metric normalised with Euclidean distance.

Finally, we assume that most similar article will describe the correct meaning of analysed entity. In our preliminary evaluation, we were able to obtain precision up to 74% on news article texts. We experienced similar behaviour as [2] when the method failed to rank the correct meaning first, even when using semantic spaces during the comparison. In most cases the correct meaning was ranked among the top three entities from whole list of possible meanings, containing usually 20 to 50 'candidates'.

We plan to further improve the results with additional classifier based on Wikipedia category structure. Our final goal is to prepare working web-based demonstration of proposed method, which require further optimisations and design decisions to speed up the current experimental implementation.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Bunescu, R., Pasca, M.: Using encyclopedic knowledge for named entity disambiguation. In *Proc. of EACL*, Vol. 6, pp. 9–16, 2006.

[2] Cucerzan, S.: Large-scale named entity disambiguation based on Wikipedia data. In *Proc. of EMNLP-CoNLL*, pp. 708–716, 2007.

[3] Gabrilovich, E., Markovitch, S.: Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. Evaluation, pp. 1606–1611, 2006.

[4] Miller, G.A., Charles, W.G.: Contextual correlates of semantic similarity. *Language and Cognitive Processes*, Vol. 6, No. 1, 1–28, 1991.

Discovering Keyword Relations

Peter KAJAN*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
peto.kajan@gmail.com

One of the popular approaches for discovering new knowledge about the users or the content on the Web is using the wisdom of the masses. We already know folksonomies, resulting from collaborative assignment of tags or keywords to resources. Analyses of this information can reveal relations between users, resources or tags. Similar results can be achieved by analyzing how users navigate on the Web and access web resources. For example, if significant number of users visit page with keyword A and right after they visit page with keyword B, we can conclude that this keywords may describe similar concept. In our work we investigate the potential of folksonomies and similar data structures for acquiring various relations between keywords.

We know different kinds of possible relations. If only used type of relation is a *parent-child relation* forming a hierarchy of keywords, we get a taxonomy. By adding other relations like *synonyms*, the relation between keywords with similar meaning, a taxonomy turns into an ontology.

Relations between the keywords could be also found using web resources like the lexical database Wordnet [1] or users created encyclopaedia, Wikipedia [2]. Our idea is to take advantage of such resources on the Web, which provide at least basic semantic grounding. For instance, Linked Data initiative associates data represented by URIs over the Web and there are plenty of tools and services, built on the top of Linked Data. Useful for dealing with homonyms might be Open Calais which is able to disambiguate a keyword with multiple meanings by using its context.

Another idea is to explore the dynamics of the Web usage. Users' clickstreams may provide sequences of keywords with similar meaning, describing the same concept or keywords forming hierarchy.

There are plenty of works presenting algorithms for taxonomies creation which are based on statistics, clustering or set theory [1,3]. Many authors also published their approaches of turning folksonomies into ontologies by either automatic [4] or semiautomatic processes [2].

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

Spring 2011 PeWe Workshop, April 1, 2011, pp. 55–56.

Our goal is to propose a method of turning folksonomies into ontologies, which combines several approaches in order to build a knowledge base characteristic for a group of users, who contributed to the creation of underlying folksonomy. The method consists of two steps: ontology creation and ontology enrichment. Our approach takes a folksonomy-like structure based on the dynamic records of the Web usage as an input and employs existing solution combined with usage dynamics analysis in the first step. Next, semantic web services are used for finding new relation and enrich the emergent data structure. This process is outlined in the Figure 1.

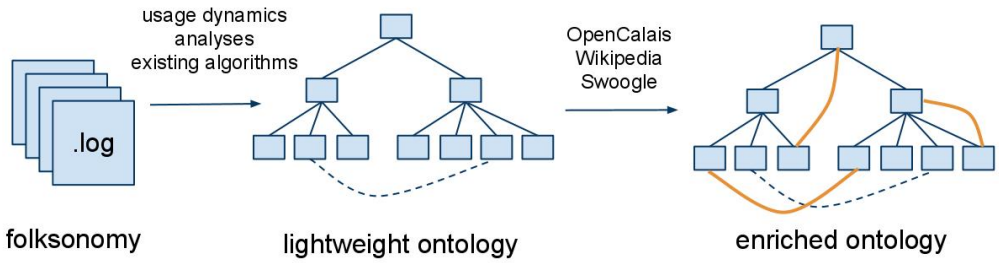


Figure1. Proposed process of turning folksonomies into ontologies

Another idea is to analyse the relations, their properties like commutativity and transitivity or relations between relations. Evaluation will be done in the domain of the personalized web search. Personalized search results will be presented by analysing the keywords in search queries and extending them using created knowledge base.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 345-032STU-4/2010

References

[1] Barla, M., Bieliková, M.: On Deriving Tagsonomies: Keyword Relations Coming from Crowd. In *Proc. of the 1st Int. Conf. on Computational Collective Intelligence*. Semantic Web, Social Networks and Multiagent Systems, Springer-Verlag, Berlin, Heidelberg, pp. 309-320, 2009.

[2] Van Damme, C., Siorpaes, K.: FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies, In *Bridging the Gap between Semantic Web and Web 2.0*, SemNet 2007, pp. 57-70, 2007.

[3] Lu, C., Chen, X., Park, E.K.: Exploit the tripartite network of social tagging for web clustering, In *Proc. of the 18th ACM Conf. on Information and Knowledge Management*, New York, NY, USA, ACM, pp. 1545–1548, 2009.

[4] Tang, J., Leung, H., Luo, Q., Chen, D., Gong, J.: Towards ontology learning from folksonomies, In *Proc. of the 21st Int. Jont Conf. on Artifical Intelligence*, San Francisco, CA, USA, Morgan Kaufmann Publ. Inc., pp. 2089–2094, 2009.

Automatic Photo Annotation Based on Visual Content Analysis

Eduard KURIC*

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

Automatic photo annotation is the process by which a computer system automatically assigns metadata to a target photo. With increasing popularity of digital and mobile phone cameras, the need for quick and exact searching grows, for example by general category or focusing on a specific object. Manual creation of annotations is very time-consuming and produces often subjective results thus making automatic photo annotation a very desirable albeit challenging task.

Generally, approaches for automatic annotation are categorized into two scenarios [1]: learning-based methods primary focus on determining complex categories or groups of specific objects and web-based methods that use crawled web image data to obtain relevant annotations.

In learning-based methods, a statistical model is built to teach a classifier. Automatic face recognition in photos is good a example of automatic annotation of specific objects. One possibility is that a retrieval process uses a robust dictionary of visual terms to identify people. Similarity can be evaluated by comparing local descriptors which are computed over local features such as edges, small patches around points of interest. The local descriptors are much more precise and discriminating than global descriptors. For searching for specific objects, this feature is useful, but by searching complex categories it can be an obstacle. Another obstacle is the need to store the huge number of the extracted features.

While the Web provides unlimited vocabulary for web-based methods, its problem is the initial query and the lack of information about target photos. Without providing information such as key caption, searching for similar photos to a target photo on the Web is like finding a needle in a haystack. Additional important drawbacks include performance and noise in obtained annotations.

In our method (Figure 1), we combine local and global features to retrieve the best results. For detection and extraction of local features, we use Scale Invariant Feature Transform (SIFT). To computation of the global descriptors, we use Joint

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 57–58.

Composite Descriptor (JCD). With their combination, we are able to ensure robustness and generalization needed for complex queries.

We place great emphasis on real-time performance. To cope with the huge number of extracted features, we implemented disk-based locality-sensitive hashing to index descriptors. By searching for similar candidates during extraction of keywords, we focus on photos analysis in terms of probability, that the retrieved photos contain the right keywords for the target photo. For example, we prefer photos where extracted objects of interest from the target photo are dominant in retrieved photos or their frequency of occurrence is greater.

Our method is designed for versatile use, e.g. identifying key objects in specific photo albums; complex automatic photo annotation in large web photo galleries; searching similar photos according to objects of interest (query by image content).

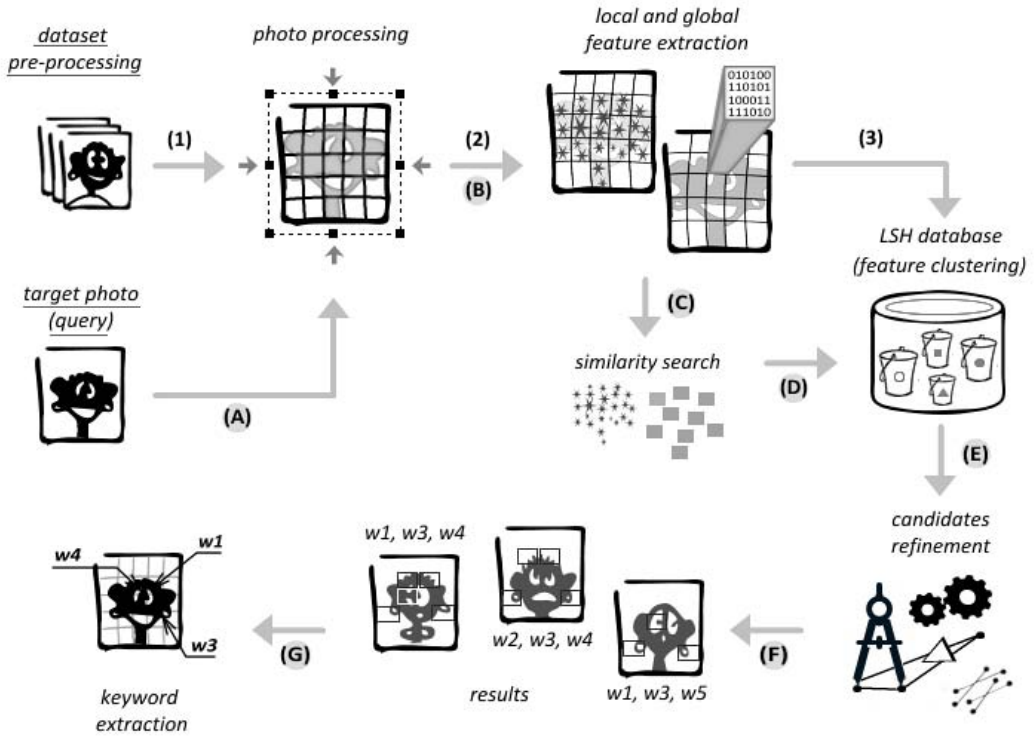


Figure 1. A scheme of our annotation method consists of two independent phases, namely dataset pre-processing (1) – (3) and processing of a target photo (A) – (G).

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] Wang, X., et al.: AnnoSearch: Image Auto-Annotation by Search. In *Proc. of the IEEE CS Conf. on Computer Vision and Pattern Recognition – Vol. 2*, CVPR '06, Vol. 2. IEEE CS, Washington, pp. 1483-1490, 2006.

Relations Discovering in Educational Texts based on User Created Annotations

Vladimír MIHÁL*

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

To increase effectiveness of learning, web based e-learning systems employ tools to personalize the content, navigation and allow students to collaborate and contribute to the educational content. To efficiently utilize personalization tools it is necessary to maintain a domain model. However manual creation of the domain model requires expert knowledge of the domain and requires much effort. Once the model is created, it is also necessary to correctly link the learning content to the domain model. Therefore it is essential to support the authoring of the domain model and its enrichment with additional relations and concepts.

E-learning systems often allow students to create text annotations [2] to support participation of students within a learning course and enhance collaboration between students. Moreover, the annotations inserted into learning content are a source of feedback related to the learning content [1], and they can be used to guide students and encourage them to perform interactive and collaborative learning tasks.

Students typically search for additional information sources on the Web, collect links to most interesting and helpful content and share them among their friends. Inserting such links into related learning texts as annotations helps organizing the links to useful resources and makes the sharing of these links easier. Teachers may insert links to provide students with additional learning resources and/or to motivate the students to insert and share their own links.

To control the quality of inserted links we provide the students with the ability to rate the usefulness of the inserted external (re)sources. We distinguish feedback we receive from students from the feedback we receive from teachers. Ratings assigned by the students represent the resource's popularity, students can either like or dislike an external resource, expressing how useful is the resource for them. Ratings assigned by teachers are interpreted as approval or disproval of the suitability of an external source in regard to the learning text. Approved sources are displayed at the top and can be further processed while the removed sources are no longer visible to students.

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 59–60.

Inserted external sources related to the learning course enrich the course with additional learning texts. However besides enriching the learning content, they better explain portions of the document where they are inserted, by the means of covering either similar topic as a part of the document, or wider range of topics, usually covering larger part of a learning course.

Using this information we derive relations between learning object and concepts from existing domain model. To process information sources in various formats and languages, we use an automatic translation service (Google Translate) and a text extraction service to extract readable text from documents and presentations. We analyze the content of external sources to link them with appropriate concepts and construct a graph from learning objects, external sources and concepts according to the known relations. We apply spreading activation algorithm on the graph to compute similarity of entities (learning object or concepts) to each learning object. According to the computed similarity score we create weighted relations and merge them into the existing domain model.

We have implemented a widget within the ALEF framework to collect external sources and explicit feedback regarding quality of sources from students and teachers. We collected more than 700 external sources from students during the Procedural programming course. For further evaluation we construct a subset of collected external sources which will be analyzed and used to derive relations. We will then separately evaluate analysis of external sources and creation of new relations.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Bancroft, P., Roe, P.: Program annotations: feedback for students learning to program. In *Proc, of the 8th Australian Conf. on Computing Education*, ACM Vol. 165. Australian Computer Society, Darlinghurst, Australia, pp. 19-23, 2006..

[2] Bieliková, M., Šimko, M., Barla , M., Chudá, D., Michlík, P., Labaj, M., Mihál V., Unčík, M.: ALEF: Web 2.0 Principles in Learning and Collaboration. In *Proc. of the Int. Conf. on e-Learning and the Knowledge Society*, e-Learning´10, Riga, Latvia, pp. 54–59, 2010.

Personalized Text Summarization

Róbert MÓRO*

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

One of the most serious problems of present-day Web is information overload. Since we can find almost everything on the Web, it has become very problematic to find what we actually want or need – to find relevant information. Also, the term "relevant information" is subjective, because users differ in their interests, goals and knowledge. Automatic text summarization aims to address the information overload problem. The idea is to extract the most important information from documents, which can help users to decide, whether it is relevant for them and they should read it or not.

The problem with classical (generic) automatic text summarization methods is that they do not consider different user goals, interests or knowledge. Our idea is to personalize text summarization since much information about user interests can be inferred from their browsing (or in this case reading) behavior [1]. If a user reads a document about a particular topic, it serves as implicit feedback that he or she is interested in the topic [2]. Also, with the arrival of Web 2.0, users are no longer passive consumers of web content, but they can create content and add metadata, such as annotations or tags. These can be used as another important source for personalization.

Annotation of documents is a technique widely used by people especially when reading printed documents. They highlight or underline important parts of text, add explanations or different formulations or even references to other documents. This way, annotations can indicate reader (or user's in the context of the web) interest in that particular part of the document [3]. We can take into account not only user annotations but also those of similar users, including the users' collaboration into the process of text summarization.

Tags, as a special type of annotations, can also be considered. They are usually generalized descriptions of the topics contained in a document and directly reflect the users' vocabulary and their understanding of the document.

We propose a personalized summarization method for web documents. Figure 1 shows the summarization process, which is common for all extractive summarization methods. The main difference, by which various methods can be distinguished, is usually in the keyword scoring and sentence selection. Also, additional preprocessing

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 61–62.

of documents and/or post-processing of produced summaries can be employed to improve readability and text cohesion.

Our goal is to explore what additional information inferred from user provided metadata can be used (and how) to give higher scores to keywords reflecting user interests, goals or knowledge thus resulting in personalized summarization; and also to produce summaries adapted to actual user needs or context (e.g., reflected in summary length). We plan to evaluate our proposed method in the domain of e-learning in ALEF (Adaptive Learning Framework); however, the method itself will be designed and implemented as domain-independent using an enhanced proxy server [1].

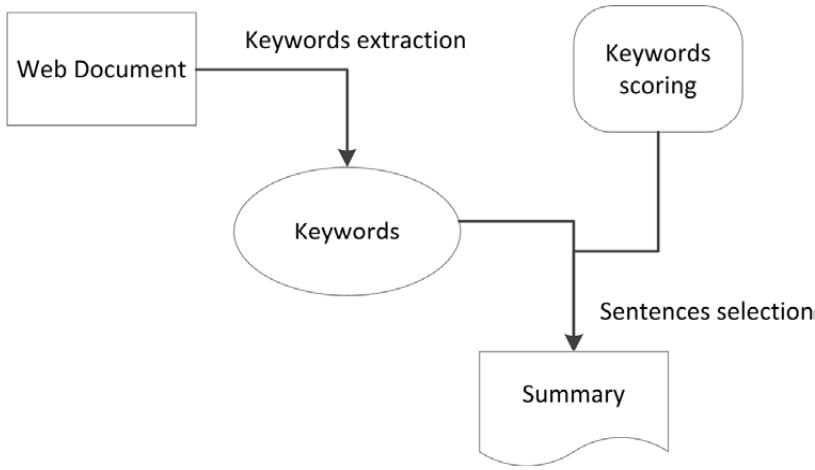


Figure 1. Overview of web document summarization.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Barla, M., Bieliková, M.: Ordinary Web Pages as a Source for Metadata Acquisition for Open Corpus User Modeling. In *Proc. of IADIS WWW/Internet 2010*. IADIS Press, pp. 227-233, 2010.

[2] Campana, M., Tombros, A.: Incremental Personalized Summarization with Novelty Detection. In *Proc. of the 8th Int. Conf. on Flexible Query Answering Systems*, Springer-Verlag Berlin, Heidelberg, pp. 641-652, 2009.

[3] Zhang, H., Ma, Z.C., Cai, Q.: A Study for Documents Summarization Based on Personal Annotation. In *Proc. of the HLT-NAACL Workshop on Text Summarization*, Association for Computational Linguistics Morristown, NJ, USA, pp. 41-48, 2003.

Metadata Acquisition via Interactive Games

Balázs NAGY*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
nagybalazs@zoznam.sk

Metadata play an important role in today's web- based world. Every time we search for web pages, pictures or videos, we get results that were found by using metadata. Its collection and creation is a challenging and complex job, which can be facilitated via *the power of human computation* [1].

We devised a game for automatic acquisition of metadata for photos. The basis of our project is the memory game PEXESO, which was reinvented to be useful for image annotation (Figure 1). We extended the game with hint for players in such a way that *players can write notes* (Figure 1, B) *to unveiled images* (Figure 1, C). The next step was to persuade the players to use the offered opportunity with increasing the size of the table up from 8x8 (Figure 1, A). Players need not memorize the exact location of pictures since by knowing their rough positions they can find the pairs via tags they input, which are shown as tooltips.

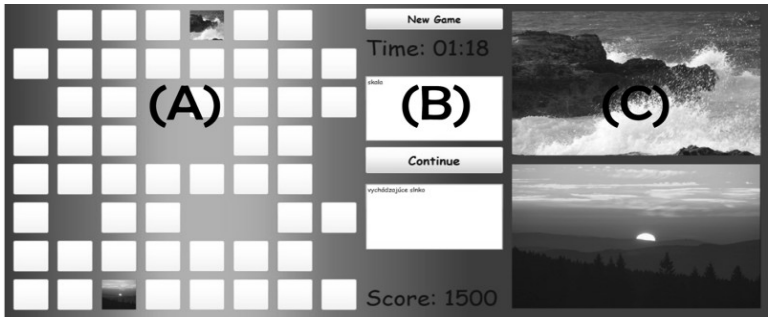


Figure 1. PexAce interface: main board (A), block with controls (B) and reversed photos (C).

We measure only time spent for searching for suitable cards, i.e. when 2 cards are inverted, players can properly label them as that stops time measurement. Once the player clicks ‘Continue’ or presses ‘Enter’, time starts to run again and the player can continue to search cards. As a cheating precaution, tooltips do not appear above cards

* Supervisor: Michal Tvarožek, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 63–64.

during annotation as this could allow smarter players to quickly reverse two cards and stop time measurement with tags shown.

To get reasonable annotation coverage for evaluation as soon as possible, we narrow down the number of photos used to 600 and in each new game select pictures from this narrowed range. Once we collect enough tags for an image, we put it on a blacklist and select only photos that are not yet sufficiently annotated in new games. After getting basic annotations, we also select photos with similar content to make the game more difficult and *produce more precise tags for given photos*.

With our prototype of the game PexAce we gained 56 registered users, and some users that played our game as guests. We know use the first 600 photos from our dataset, which have so far been annotated with 1,599 annotations, corresponding to 2,605 words, what means that we created in average 2.67 annotations and 4.43 words for one photo.

The quality of labels is very promising, because only few of them contain misspellings or are written in other languages. Even these tags are not lost as we fix them via the Google Translate and WordNet services.

We accept and save user input as tags if 3 different players agree on the same tag for a photo. With this method we eliminated the problem of multiple words given by the same user to the same photo, because these words are counted only once. Guests are treated as a single user.

Today's results and research suggest that we cannot underestimate the potential of Games With A Purpose [2], which *can utilize lots of time wasted by playing normal games* for benefit in many areas that can be supported by GWAPs [3].

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] von Ahn, L.: *Human Computation*. Pittsburgh, December 7, 2005.

[2] von Ahn, L., Dabbish, L.: Designing Games With Purpose. *Communications of the ACM.* pp. 58-67, 2008.

[3] Šimko, J., Tvarožek, M., Bieliková, M.: Little Google Game: Creation of Term Network via Search Game. In *Proc. of Datakon 2010,* 2010 (in Slovak).

Web Content Annotation in Educational Web-Based Application

Jakub ŠEVCECH*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
sevo_jakub@yahoo.fr

Today we are facing the problem of information overload. It is thus important to provide information to web page readers in such a way, that it can be used most effectively. Annotations attached to documents are frequently used as a way for organising and providing additional information.

In our work we propose a method for automatic extending the information content of web pages by adding annotations to keywords in the text of the pages. The method is designed to be able to insert annotations into the text written in Slovak. The process of creating annotations for a web page consists of several steps:

1. elimination of redundant parts and selection of text to be annotated,
2. extraction of candidate words for assignment of annotations,
3. search for information to the annotations, and
4. visualization and personalization of the annotations.

Before creating the annotations, it is necessary to analyze the document and find the words to which it is appropriate to assign the annotations. As a first step it is necessary to remove redundant parts of the web page as various navigation elements, advertisement banners, etc.

The second step of the analysis of document is the search for candidate words for assignment of annotations. For this step, it is possible to use various services for keyword extraction [1]. Since these services provide best results when English text is processed, method for creating annotations uses text of the web page translated into English. To assign annotations to words in the original text, we propose a method for mapping of equivalent words between the Slovak text and its translation into English.

A method for mapping equivalent words between text translations requires an extensive bilingual dictionary. The size of the dictionary required for this task may be prohibitively large; hence we use a different approach. We employ a much smaller dictionary and match different word forms using Levenshtein distance. We calculate

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 65–66.

the Levenshtein distance of the words in the dictionary and the words in the text. If this distance is lower than a selected threshold, we consider these words to be equivalent to each other.

Information for the annotations is obtained through publicly available services for information retrieval. Using these services, the proposed method does not depend on any particular domain. Annotations created using different services can take different forms, depending on the services used to fill them. We employ services providing definitions of keywords and links to web pages concerning these keywords, but other services that provide multimedia information can be used as well.

Annotation personalization is performed on the basis of implicit feedback, gathered from user clicks on links in the content of the annotations. Whereas these links are presented in a list, user is affected by its order when deciding which link to follow. Consequently, we do not assign same weights to clicks on different positions in the list, but we treat them as indications that the clicked link is better than the other [3]. We use different strategies to partially order the links using the knowledge which links were clicked and which not. We then build a graph from this partial ordering and we use adapted PageRank algorithm to find the ratings of links for annotation purposes. The links in an annotation are displayed sorted by their rating.

In the future work we will evaluate these methods on education system ALEF [2], on the course of software engineering.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Barla, M., Bieliková, M.: Ordinary Web Pages as a Source for Metadata Acquisition for Open Corpus User Modeling. In *Proc. of IADIS WWW/Internet 2010*. IADIS Press, pp. 227–233, 2010.

[2] Šimko, M., Barla, M., Bieliková, M.: ALEF: A Framework for Adaptive Web-based Learning 2.0. In *Proc. of IFIP Advances in Information and Communication Technology, Key Competencies in the Knowledge Society*, Springer, Vol. 324, pp. 367–378, 2010.

[3] Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting clickthrough data as implicit feedback. In *Proc. of the 28th annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, New York, USA, ACM, pp. 154–161, 2005.

Harnessing Manpower for Semantics Acquisition

Jakub ŠIMKO*

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

Nowadays, the amount of information on the Web grows extremely fast. In order to be able to search the Web and utilize its content, we require scalable methods for acquiring information about individual resources. This information, opposite to the heterogeneity of web resources, must be homogeneous to be easily processed by machines. Nowadays, the role of such homogeneous meta-layer above the Web is played by the keyword search indexes within web search engines.

Semantic Web principles were created to provide a worldwide framework for creating richer web resource annotation than keyword indexes. The Semantic Web can be seen as a meta-layer on the common web: a collection of web resource annotations unified under universal and widely accepted domain models. With such structure, the web invisibility problem [2] is easy to solve – it can be used to create various forms of web abstractions to browse. The solving of the problem of sophisticated queries becomes also trivial – besides the structured information and knowledge, Semantic Web standards provide logical reasoning frameworks suitable for question answering. Although there is much work done in the field of automatic semantics acquisition, human effort on building the semantics is still a need. In response to that we study approaches for creating semantics on the Web with accent on games with a purpose (GWAP), which are computer games that transform human intelligence tasks into entertaining games [5, 3].

In our previous research, we introduced *Little Google Game* (LGG), a game with the purpose of acquiring a general term relationship network. Little Google Game is a web search game of query formulation, in which the player's task is to minimize number of search results returned for the given single-term query (e.g., query "sea" which yields approximately 2 billion results). The player can expand the query using only *negative* search terms that reduce the original result set of web resources decorated with negative terms (e.g., "sea –blue –fish –deep"). The game is single player and exercises a ranking system to motivate competition among players [4].

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 67–68.

We have validated the semantic soundness of the relationships within the network and shown that the game can discover relationships that remain hidden to automated corpora mining methods. Now we look at the LGG network as a potential source of ontological facts. We conducted experiments to disclose how many of the relationships are present in the major knowledge base of Wikipedia using the Wikipedia Miner tool and what kind of relationships are mostly present in the network, by evaluating them manually and also comparing them to the facts in the knowledge base of ConceptNet. We propose two methods for naming the LGG network relationships: first, a modification of Little Google Game itself forcing players to disclose predicate-like terms related to existing bigrams and second, automated sentence mining with web search engines helping to gather relevant sentences.

We are also interested in the design of games with a purpose for the Semantic Web in general. GWAPs are a potent tools for harnessing human knowledge and hypothetically applicable to any problem. Creating a GWAP is however, not so straightforward and is very specific with each problem. Our goal is to reach a set of best practices and guidelines to unify the creation of GWAPs (for the Semantic Web) since many design aspects recur (e.g., reaching the entertainment, elimination of cheating, validation of user inputs). One possible way may be a modification of a classical game design methodology, such as MDA [1], or methodology for transformation of existing classical games to games with a purpose.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Hunicke, R., Leblanc, M., Zubek, R.: MDA: A formal approach to game design and game research. In *Proc. of the Challenges in Games AI Workshop, Nineteenth National Conf. of Artificial Intelligence,* pp. 1–5, 2004.

[2] Marchionini, G.: From finding to understanding. *Communications of the ACM* Vol. 49, No. 4,, 41-46, 2006.

[3] Siorpaes, K., Hepp, M.: Games with a purpose for the semantic web. *IEEE Intelligent Systems*, Vol. 23, 50–60, 2008.

[4] Šimko, J., Tvarožek, M., Bieliková, M. Little Google Game: Creation of Term Network via Search Game. In *Proc. of Datakon 2010,* 2010 (in Slovak).

[5] von Ahn, L., Dabbish, L.: Designing games with a purpose. *Communications of the ACM*, Vol. 51, No. 8, 58–67, 2008.

Automated Metadata Extraction for Adaptive Social Learning

Marián ŠIMKO*

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

The emergence and spread of Web 2.0 technologies changed the face of web-based learning. It enabled to fulfil the "Read-Write Web" vision. Students cease being passive consumers of information. They benefit from learning and collaboration in a social learning environment and can contribute to the content by various forms of annotations: tags, ratings, comments or feedback. The distinction between teacher and student is being reduced [2].

In order to make the learning process more effective, educational systems tailor learning material to student's goals, needs and characteristics. Adequate adaptation requires a domain description enabling adaptation engines to make at least basic reasoning. The bottleneck of adaptive educational systems is the complexity of domain model creation and update. Identifying concepts – domain knowledge elements – or defining hundreds or even thousands of relationships between them is difficult and almost impossible for humans. The complexity of domain model update is notable especially in the case of student-generated content when considering Adaptive Web-based Learning 2.0. To the best of our knowledge, there are only few works related to automatic metadata acquisition in adaptive educational web-based systems. State-of-the-art approaches rely mostly on domain experts or teachers, who supply an adaptive system with necessary semantic descriptions – e.g. [1].

In our work we aim for an unsupervised approach requiring teacher assistance for fine-tuning the automatically generated domain model only. We consider heterogeneous sources of information to process and extract relevant domain descriptions. We particularly focus on learning objects (created by a teacher), social annotations (created by students) and links between them. The method we propose consists of the three major steps: (1) resources preprocessing, (2) relevant domain term (RDT) extraction, and (3) relationship discovery (see Figure 1). In our approach we employ methods and techniques of text mining (statistical, linguistic processing), graph analysis and set theory. We proposed several variants of relationship discovery, each providing a unique view of the actual domain model state.

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 69–70.

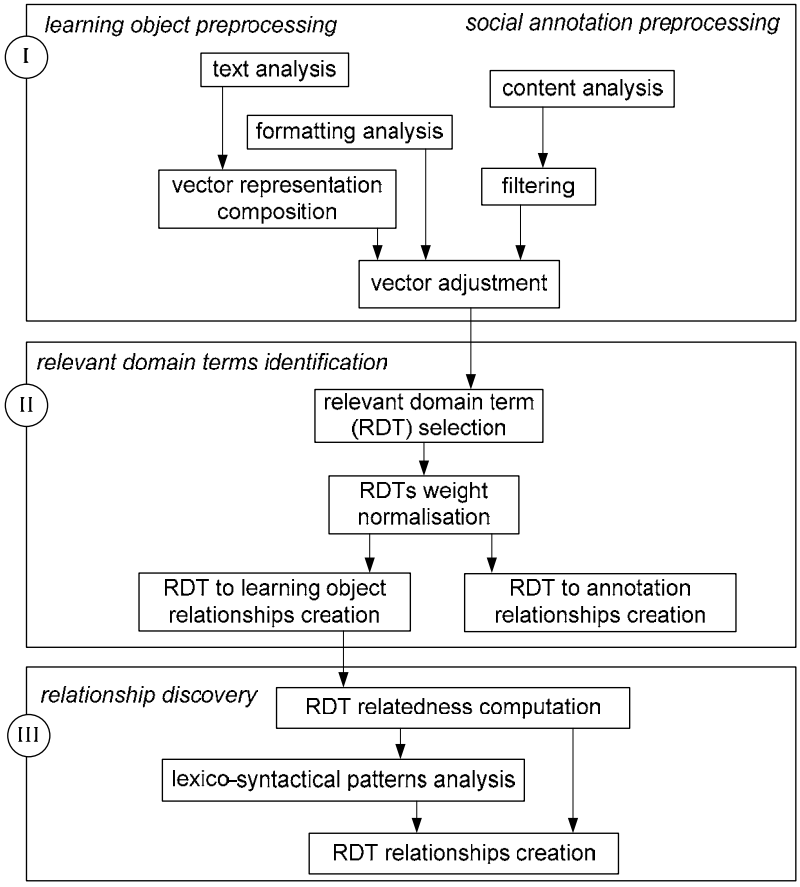


Figure 1. The method for automated metadata extraction.

We have integrated the method into the ALEF (Adaptive LEarning Framework) [3] and already conducted several real-world experiments. In the current stage of the research we evaluate the proposed method by performing recommendation simulations.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Cristea, A. I., de Mooij, A.: Designer Adaptation in Adaptive Hypermedia Authoring. In *Proc. of the Int. Conf. on Information Technology: Computers and Communications ITCC'03*. Las Vegas, US, IEEE, pp. 444–448, 2003.

[2] Downes, S. E-learning 2.0. In *eLearn magazine*. ACM, No. 10, p. 1, 2005.

[3] Šimko, M., Barla, M., Bieliková, M. ALEF: A Framework for Adaptive Web-based Learning 2.0. In *KCKS 2010, IFIP Advances in Information and Communication Technology*, Vol. 324. Springer, pp. 367–378, 2010.

Automated Recognition of Writing Style in Blogs

Martin VIRIK*

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

In the current web, blogs represent a genre that stands between static pages and live forums. They have became a tool for ordinary users to share information, ideas or even emotions, creating a heterogeneous mass of user generated content filled with unique information related to individual as well as society-wide issues. Since the blog articles are typically weakly structured, the extraction of valuable information is becoming increasingly difficult to handle.

Even though there are no restrictions or limits to content or form in blogs, users seem to spontaneously create writing styles and genres reflecting their intention and current emotional state. Modern search services are aware of these genres and consider them as highly valuable for several tasks. For example, for filtering articles with low information value or recognizing sentiment about a specified object.

In our research we follow a distribution between informative and affective articles [1] and for affective articles we suggest two dichotomies of further differentiation: reflective vs. narrative and emotional vs. rational. By their combination, we receive four categories, which we will use for our classification (direct reaction on an event, rational reflective article, direct reaction on writer's day or a story from past, i.e. vacation, a tale from childhood). We have gathered a dataset for initial experiments of about 16 thousand blog posts and managed to manually classify a subset in a user experiment including several participants.

In our work we focus on linguistic characteristics of blog articles. We propose a novel method for Slovak blogs classification that considers not only word usage and lexical and morphological attributes (typical for state-of-the-art approaches [2]), but also more complex features such as sentence syntax or text structure obtained during a pre-processing step. We have improved the morphological tagging by considering word's position in sentence enabling syntactic analysis, as well as capturing the usage of each word class. We have built a lightweight syntactic analyzer, which transforms morphologically tagged text into a structure based on an object model depicted in

* Supervisor: Marián Šimko, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 71–72.

Figure 1. A result of this transformation is the localization of predicate candidates, which are the verbs that we have used to identify the sentences in the previous step. In our feature set we investigate their dominant tense, person and number categories along with the intensity of this dominance.

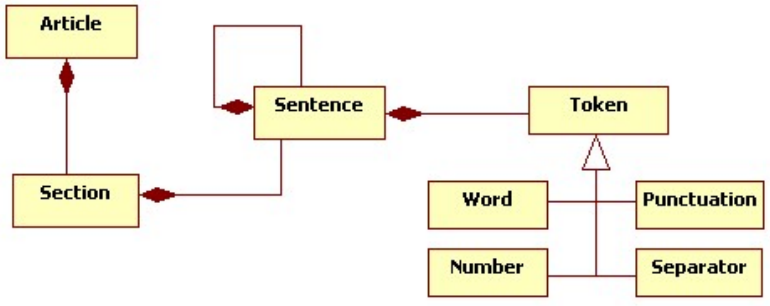


Figure 1. Class diagram of article structure model.

Besides the simple structural features (i.e. number of sections, average number of sentences per section) we measure also the modified standard deviation of section length. We have modified the standard deviation formula by counting the proportion of difference and average, so now it reflects the variation of section length and the consistency of the whole text.

To evaluate our method, we have suggested a compound classifier based on three binary classifiers for each pair of classes. We have conducted an initial experiment, in which each article was represented by a 28-dimensional feature vector and for each classifier we have integrated and configured Naïve Bayes (as baseline classifier), Suport Vector Machine and k-Nearest Neighbours classifiers from Weka tool. By applying 10-fold cross-validation evaluation method we have gathered first result with precision 68–81%.

In the current phase of our project we are gathering a larger training set of manually classified articles and improving lightweight syntactic parsing methods and methods for feature selection. We are experimenting with classification algorithms and classifier comities in order to boost the accuracy of classification.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 345-032STU-4/2010.

References

[1] Ni, X., Xue, G., Ling, X., Yu, Y., Yang, Q.: Exploring in the Weblog Space by Detecting Informative and Affective Articles. In *Proc. of 16th Int. Conf. on World Wide Web*, ACM, pp., 281–290, 2007.

[2] Argamon, S., Koppel, M., Pennebaker, J.W., Schler, J.: Automatically profiling the author of an anonymous text. *Communications of the ACM*, Vol. 52, No. 2, 119–123, 2009.

Workshop Events Reports

PeWeProxy Workshop

Michal BARLA and Tomáš KRAMÁR

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
{barla,kramar}@fiit.stuba.sk

1 Motivation and expected outcomes

For the first time in the history of PeWe workshops, we decided to include a hack-day-like activity into the program. The initial motivation was to allow workshop attendees to play and experiment with various available platforms, services, or APIs and take advantage of innovative potential of PeWe members to invent something new and interesting in the field of Adaptive Social Web.

We chose our specialized proxy server [1] *PeWeProxy* to play the role of framework upon which workshop participants were asked to build their solutions. *PeWeProxy* serves as a man-in-the-middle between user's browser and remote web servers and provides all required means for manipulating HTTP messages flowing through the proxy (see Fig. 1 for an overview).

Manipulation of HTTP messages is done by two types of plugins, inserted into the processing pipeline

- *Request processing plugins*, which operate on the request from users' browsers. A plugin can decide upon further processing of the request – whether it should be treated normally, modified, redirected to other web server or completely dismissed.
- *Response processing plugins*, which operate on the response from the remote web server. Similarly to previous case, a plugin can decide upon modifications or other treatment of the response (e.g., process the response asynchronously in order to extract certain features from the content).

The proxy comes with a set of services, which allow plugins to perform basic operations on the request and response messages and provide core functionality that other plugins can build upon. More, the proxy is bundled with a set of core plugins, which are responsible for basic operations such as user identification or logging of user's accesses to web pages.

This flexible architecture allows for almost limitless possibilities when it comes to enhancing existing pages. Their content may be altered, removed, or even completely replaced.

Spring 2011 PeWe Workshop, April 1, 2011, pp. 75–86.

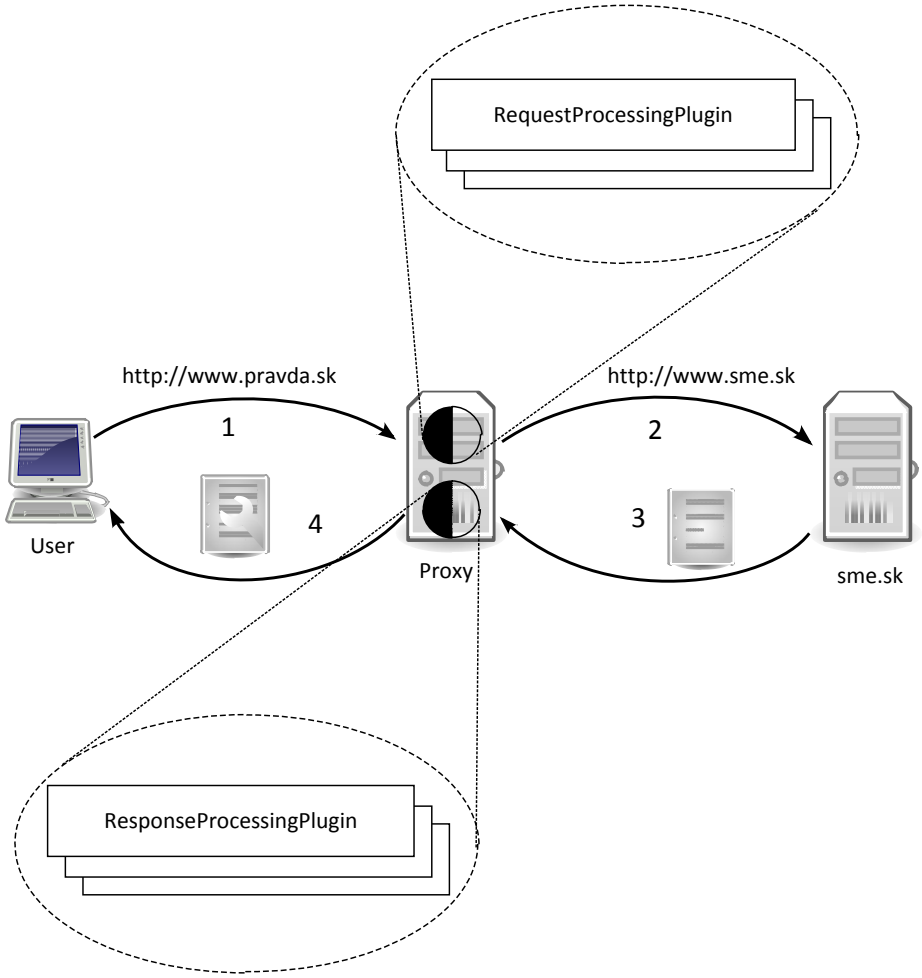


Figure 1. Basic outline of the PeWeProxy operation. The request from user's browser is sent via proxy (1), where it is processed by a chain of request processing plugins. Next, the request is forwarded to the remote web server (2). The remote response is sent back to the proxy (3) where it is processed by a chain of response processing plugin. The (possibly) modified response is sent back to user's browser (4).

The task given to the participants of the hack-afternoon session held at 9th PeWe workshop was to create small applications (plugins of the PeWeProxy), that would enhance and improve existing Web pages, to make the browsing experience more social, adapted and generally more attractive to the end users. We expected not only to collect interesting ideas on useful PeWeProxy plugins, but our aim was also to show to PeWe members how easily they can realize their ideas in the PeWeProxy platform and thus how easily they can use this platform as a basis for their future work.

2 Team reports

The hack-afternoon was preceded by an introductory pre-session the day before the actual event, where we discussed the basic concepts and ideas of the PeWeProxy platform. This kick-off session was supposed to make the attendees familiar with the possibilities and provide a short overview of existing proxy API, so the participants could start hacking immediately.

Participants were supposed to work in teams, with each team having a leader – a doctoral student, who was asked to come with an initial idea of Web surfing enhancement, which could be realized via PeWeProxy platform. Teams were created right after the tutorial part of the kick-off session, by drawing names from the pools of participants. We strived for balanced distribution of master and bachelor student across all teams.

In this section, we present ideas and results achieved by individual teams.

2.1 Colour your Web!

Team leader: Jozef Tvarožek
Team members: Anton Balucha, Peter Krátky, Štefan Mitrík, Martin Virík

Web users are overloaded by the sheer volume of hyperlinked text that is continuously emerging on the Web. Typically a single user follows only a limited number of topics that are relevant to him, however, it takes him considerable amount of time to quickly detect and process (i.e., read and understand via follow-up searches) all of it when visiting previously unknown areas of web sites.

The idea of our approach is to cognitively enhance user's web browsing experience with the use of colours. Each colour represents a different context. When browsing a hyperlinked text via the PeWeProxy (Fig. 2), the text paragraphs are augmented with three coloured buttons – red, green, blue. User can choose to colour the accompanying text using the coloured buttons. After the button is clicked, the nearby text is assigned to a context of the colour of the button pressed. On the basis of this colouring, text on web pages is semantically analysed and coloured according to the colour of the context that contains the most similar terms. Similarly with links on the current page, the contents of the pages is retrieved, analysed, and coloured according to the most prevailing context.

Semantic analysis of the text is performed using a latent semantic index devised from the DMOZ.org directory of web pages. The index contains around 700k documents categorized in 80k different categories. Processing of a single page visit takes up to 2 seconds.

2.2 Motivator

Team leader: Dušan Zeleník
Team members: Michal Lohnický, Jakub Ševcech, Róbert Móro

The most attractive websites such as Facebook or Youtube are the most time consuming as well. Besides fun which we experience with these sites, we should think about the efficiency of our browsing and productivity of activities on the web. We often do not know how much time we spent playing games or exploring friend's

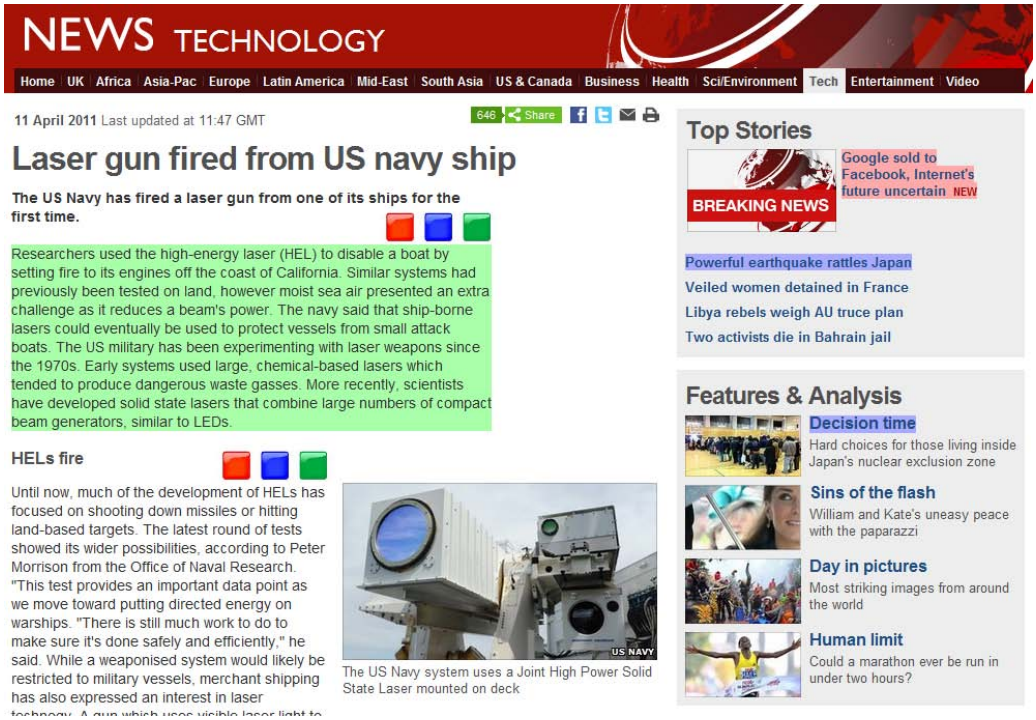


Figure 2. Coloured buttons provide an easy contextual feedback facility. After some text is coloured (button at a paragraph of text is clicked), text paragraphs on the current page, and links to other pages are coloured according to the similarity to the user's preferred contexts.

profile pictures. On the other hand, besides reading tutorials, reading papers or learning at Wikipedia we need some time for resting. But who knows what the best ratio is? How much could one play and how much should one work? We are here to solve this problem by simple, non-invasive motivation which is based on measuring the effective time spent on the web.

Our idea consists of monitoring the sites which the user visited, and time he spent on the site. Secondly, we evaluate whether it is productive or not. Efficiency of the time is calculated as a ratio of productive time and total time. Results are displayed to the user in the top right corner of the site he visits along with tendency of his efficiency (see Fig. 3). To decide whether the site is productive or not we applied more techniques starting with manual decision made by user, through collaborative assignment and ending with automated keyword-based recognition of the site. Our further work consists of the statistics and graphs which would be displayed to the user more detailed.

To sum up our contribution, we bring a simple solution to help users of the web to track their productivity and get to know about spending their precious time on the web. They are able to see their productivity over time what would possibly motivate them to use their time more efficiently. Furthermore, we are able to monitor even employees and their productivity what would have positive impact on the efficiency in businesses.

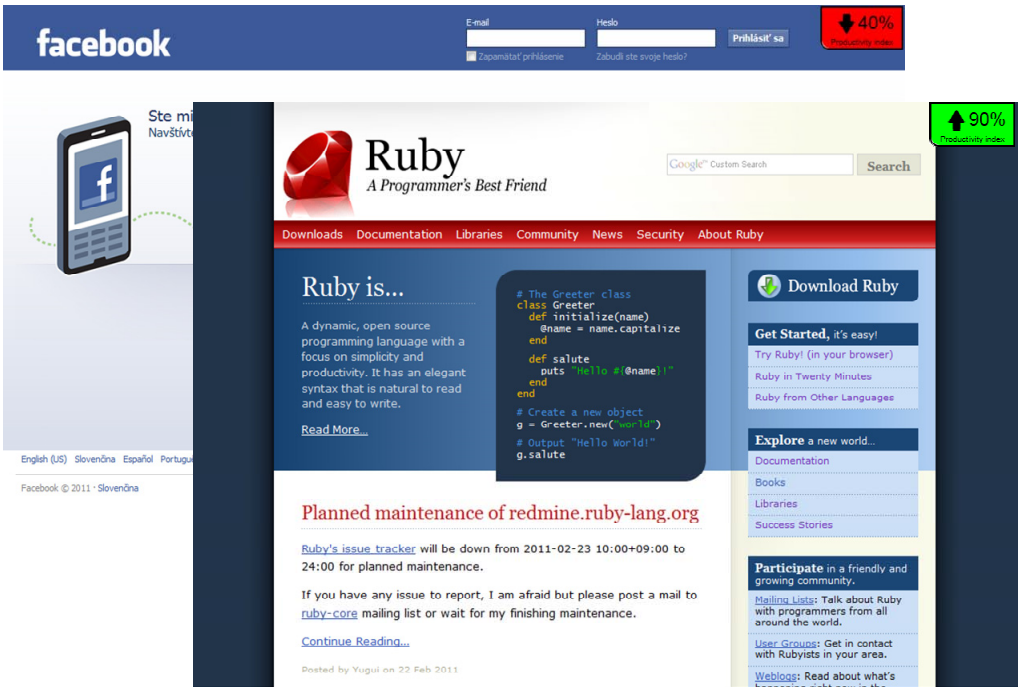


Figure.3. Widget is added to the top right corner of the web site.

2.3 Bublinátor – the real Social Web

Team leader: Ján Suchal
Team members: Martin Jačala, Marek Láni, Nagy Balázs, Márius Šajgalík

Since Web 2.0 started to gain attention, social aspect of the Web has become an important part of the way we interact with other people. We follow what our friends are doing on Facebook or Twitter, we tag and comment photos of our families, we plan meetings and discuss topics with our colleagues online. However in real life, there is at least one more social aspect that is current Web still lacks – real social navigation.

When we seek for a good restaurant in a foreign city, we surely will not land in one that is empty at lunchtime, but on the contrary, we would look for one that is nearly full. We are surely not the only one browsing the Web, but where are the others right now?

Our approach uses PeWeProxy in two ways: a) to track where all users are currently browsing and b) enriching all websites with information about where other users currently are. The latter is done by adding “bubbles” with count of current visitors over links that user can see on a webpage (Fig. 4). Furthermore we also provide and update real time global aggregation statistics about most popular content and websites on the Web.

Figure 4. Real time navigation on the newspaper portal. The green bubbles show how many people are currently reading the article that link points to.

2.4 DeNERDizer

Team leader: Marián Šimko
Team members: Vladimír Mihál, Anton Benčič, Róbert Horváth, Matúš Tomlein

The Web today is wild and savage place, while kept open for everyone. Being its greatest advantage on one hand, Web openness and unlimited access resulting in a potential negative influence constitutes a threat for the whole society on the other hand. In our work we focus on PeWeProxy advisory and supervisory capabilities.

Being able to store a user model and to track every web page a user aims to visit, the proxy can intervene according to predefined actions. The extent of intervention can vary. Unlike traditional state-of-the-art approaches, it is not limited to preventing access to a page representing a threat. The proxy can change the page and enrich it with additional information with respect to a user's context. This way, we can censor porn pages for children, crazy shops for drug addicts, or arbitrary type of content for any defined type of user.

At the workshop we created a demo application named deNERDizer, which prevents bad habits caused by browsing inappropriate web sites. In particular, we focused on people with increased interest in computers and technology at the expense of social life and one's own health. DeNERDizer comes in two versions: *Lite* and *Full*. Both versions consider a web page being accessed and the context of the visit. DeNERDizer Lite considers the context as a current weather at the time of browsing. By categorizing the web page it estimates the level of inappropriateness, i.e. it reveals if the page falls within Computer & Internet, Gaming or Science & Technology categories. DeNERDizer Full considers both current weather and user's schedule extracted from his calendar. It detects user's intent to participate in the inappropriate events during browsing (by processing the web page content). As a consequence,

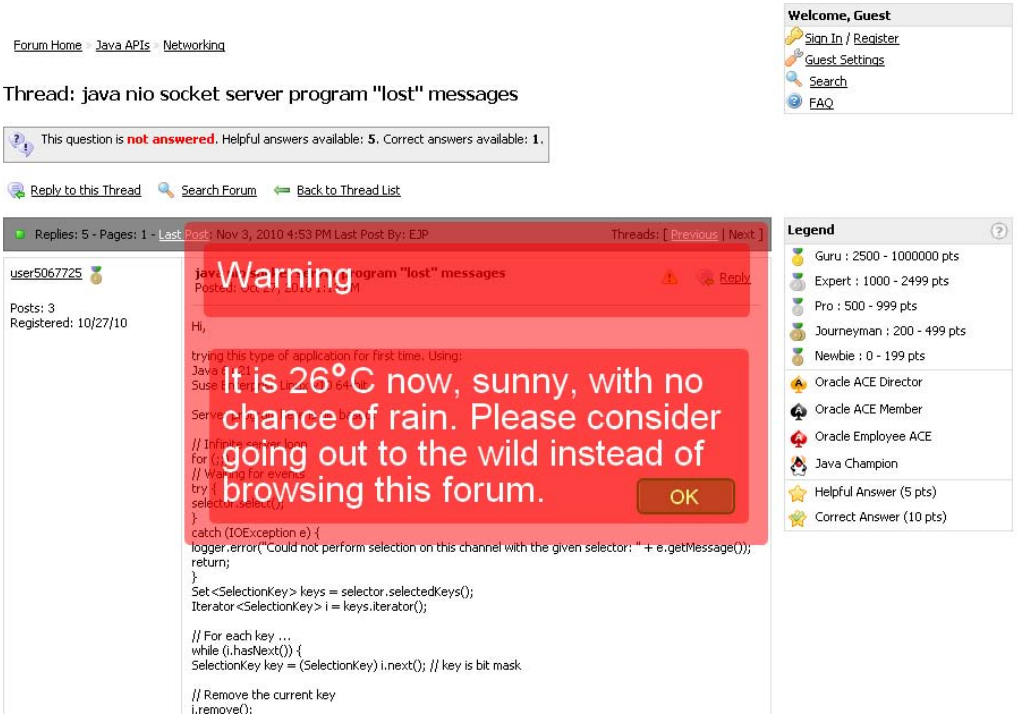


Figure 5. Example of proxy intervention when browsing inappropriate content.

deNERDizer enriches the page by warning the user of bad habit, which can potentially harm him (see Fig. 5).

In our work we use AlchemyAPI to obtain a web page category and Google Calendar API to access user events and weather forecast.

2.5 Actual news topics

Team leader: Michal Kompan
Team members: Marián Hönsch, Maté Fejes, Ivan Srba

The amount of information on the Web is increasing continuously. In the news domain there are several aggregation portals developed in the last decade (GoogleNews, Megite, NewsBrief etc.). When user wants to find actual or "hot" topics, the only chance how to do it is to visit such an aggregation portal. The placement of an article on the front page may be considered a kind of hotness too. However, there are usually plenty of other not so actual and important news included in the front page.

Our solution connects aggregation portal with any web site (primarily news portal) and displays information about actual article attractiveness (popularity of the article across all influential news sources). Figure 6 presents typical example of informing user about article attractiveness progress.

Thanks to PeWeProxy this information is displayed wherever the link to the actual article is included. In such a way users can easily not only see, which articles are important and discussed in the society actually, but they are resistant to some kind of editor's influence.

As a source for "hot" news we used NewsBrief.eu, which monitors hundreds of sources across the Web in various languages. These aggregations are updated every 10 minutes, which allows us to respond to changes in almost real-time.

In the future we can easily extend presented information to displaying e.g. article abstract or peaks of article attractiveness. Similarly, we can display all sources of the article, and users can choose which one will they read, without time-consuming search.

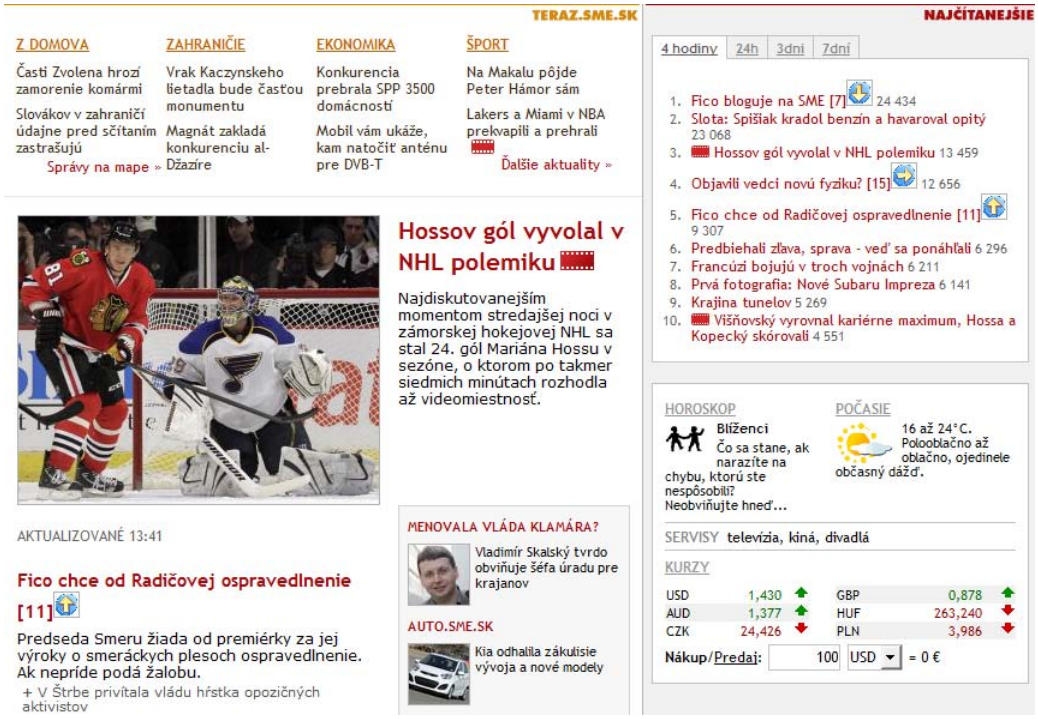


Figure 6. Visualization of the news attractiveness. Arrows indicate decrease, increase or no change in news "hotness" followed by number of sources sharing the news – number in brackets.

2.6 Web revisitation support

Team leader: Michal Tvarožek
Team members: Martin Lipták, Karol Rástočný, Peter Študent, Michal Tomlein

Effective revisitation of web resources is often required by many users during exploratory search tasks to track changes in the evolving Web [2]. Existing approaches include browser aids, such as history lists, bookmarks, and more advanced solutions such social bookmarking sites. Most approaches however only focus on revisitation support on the document / web page scope and do not address user revisitation needs with finer granularity (e.g., document paragraphs or parts of web pages).

Historically in the desktop environment, change tracking was supported by tracking changes in documents via specialized tools or by using utilities that specifically highlighted changes between two documents (e.g., the popular diff tool). In the web environment, revisitation is usually only supported in browsers by highlighting

already explored links with a different colour, while leaving revisitation support on individual web sites to site designers (e.g., via a News page showing recent changes, or RSS feeds aggregating all site modifications).

We address resource revisitation support by adaptively highlighting new (content) parts of web pages with different background colour based on their age (see Fig. 7). In our approach, we first divide the web page into different regions and identify relevant content regions while ignoring the rest (e.g., banners, ads, menus). We record hash values for the page as a whole and for individual page parts along with timestamps of their acquisition.

Figure 7. Revisitation support on the newspaper page.

During regular browsing we augment the presented web pages with highlighted page parts, which have changed since the last user visit in order to provide a quick overview of what has changed and needs to be explored again. In this way, we improve user experience by decreasing information overload by guiding users to new and/or relevant content and also support information recall as users have a better understanding of what information they have explored previously.

2.7 Link summarizer

Team leader: Jakub Šimko
Team members: Martin Labaj, Maroš Unčík, Pavol Bielik

During the exploratory search tasks like exploration of an unknown domain (e.g., during learning), the web user needs an adequate browsing guidance. This includes especially the visualization of the information space to provide browsing options for the user and to support his decision making concerning navigation.

To put this paradigm into practice, we introduce the *Link Summarizer* PeWeProxy plugin. Its aim is to inform the user about the content of textual web resources hiding behind the links on the web page he is currently on. The information is displayed in context frame upon mouse-over-the-link event in the browser (the content is loaded asynchronously and stored for further usage). It enables a quick review of the target contents displaying title, extract, resource keywords, relevance to the current resource and the transform factor indicating whether other users usually use this link (see Fig. 8). This may help user answering frequent questions and help him to decide whether the resource is relevant. The attributes are acquired as follows:

- Title. Straightforwardly taken from resource HTML.
- Extract. Contains most representative sentences from the resource. Various text summarizers could be used to extract these. We used the *Open Text Summarizer*.
- Keywords. Using the *Metall* web service, the relevant terms for the resource are extracted.
- Relevance to the current resource is computed as cosine similarity of vectors of characteristic terms of the current and target resource. Alternative similarity metrics could be used, such as *Wikipedia miner* (based on linkage between the articles), if we consider usage over Wikipedia.
- Transfer factor. Based on existing logs recorded by PeWeProxy, each link in the current document has a relative rate of usage, which may be categorized as none, low, medium and high.

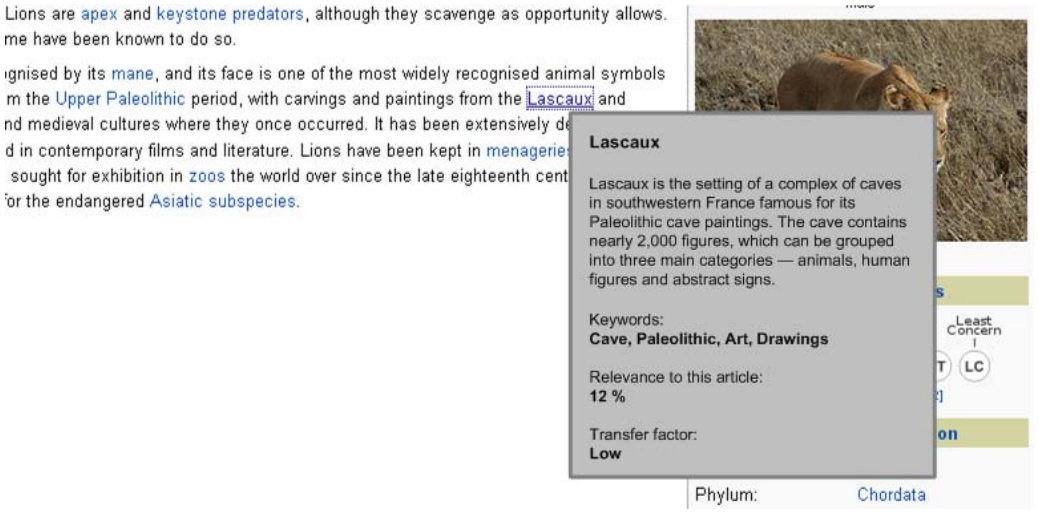


Figure 8. Link summarizer displaying the summary of a Wikipedia page

2.8 Decision support on a website of travel agency

Team leader: Michal Holub
Team members: Jakub Kříž, Eduard Kuric, Peter Macko

The Web is not only storage of information; it is more and more becoming a platform for doing business. Many people use it for shopping for various items. However, many

websites lack some decision support for the user when he is about to make a purchase. In this work we present a simple yet powerful expansion of a travel agency's website.

When a person plans his trip or vacation he usually considers a travel agency. They offer various trips with the details about the destination and activities, price of the trip and possible dates. It is up to the user to make the appropriate calculations concerning his financial situation. We introduce a feature allowing the user to calculate the time he needs to earn enough money for the trip. We calculate this from the automatically parsed price of the trip and the salary given by the user. Moreover, we highlight the first possible date when the user can travel (see Fig. 9). We also calculate possible savings from the travel considering different prices of the same products in various countries.

We add a JavaScript which displays an edit box next to the price of a trip. The user enters his hourly wage (the amount he can save) and after clicking the button the script displays how many days he needs to work (considering 40 hours working week) in order to earn enough money. Next, from the number of days and the current date we compute the earliest possible date when the user can travel. From the third party website we obtain the prices of beer in Slovakia and in the country to which the user wants to travel. From the price difference we compute the money savings when staying a longer time. All these features help the user with the decision.

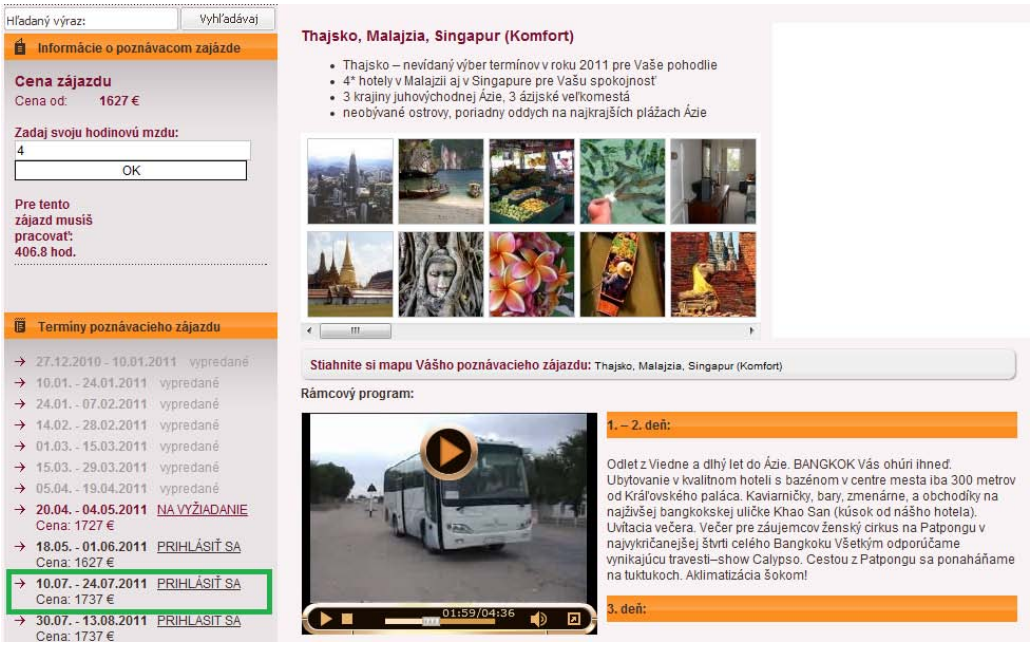


Figure 9. Screenshot of a web page with details about a trip. Calculation of the work time is in the top left corner, bottom left corner contains highlighted date of first travel.

3 Summary

Since this was the first time we tried to organize a hack-day like activity within our PeWe group, we did not know what to expect. Finally, we were very pleased with the course of the whole event, the motivation and enthusiasm of all participants. Our expectations on the actual implementations results were rather low, given the limited time and teams which were not used to work together. Therefore, we were astonished by the quality of all projects, which were all fully functional or was really close to such state. We also appreciate the quality of final presentations of the projects, given by team leaders.

The jury, which had to choose a winning project, was not in an easy situation. However, after some discussion, we decided to grant the first place to the project *Bublinátor – The Real Social Web*, led by Ján Suchal with the project *Actual News Topics*, led by Michal Kompan as the runner-up.

Finally, we would like to thank all participants for excellent work and ideas. We would also like to thank Professor Mária Bieliková and Professor Pavol Návrat for being members of the jury and especially professor Mária Bieliková for her great help with organisation of the event and motivating all participants.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grants No. VG1/0675/11 and VG1/0508/09.

References

[1] Barla, M., Bieliková, M.: Personalization of "wild" Web: Adaptive proxy server. In *Proc. of WIKT 2009, 4th Workshop on Intelligent and Knowledge oriented Technologies,*Equilibria, Košice, Slovakia, pp. 48–51, 2009. (in Slovak)

[2] Marchionini, G.: Exploratory search: from finding to understanding. *Communications of the ACM*, Vol. 49, No. 4, 41-46, 2006.

Group Recommendation

Michal KOMPAN*

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

Numerous of our everyday activities have a social character. We need to cooperate with other humans. An example of such cooperation is watching a movie in the cinema, or taking part in a group activity during holiday. There are several situations when we are forced to act socially or we are forced to share a common space, items or services and may lead to not so positive feelings. By recommendation we want to make us to feel more comfortably in all of these situations.

Several approaches to group recommendation have been proposed. However, the aggregation of single user preferences is more researched and used nowadays. Several researches have shown that preferences are changing during the recommendation process [1]. So called emotional contagion can be observed. This is connected to the group homogeneity and the type of relationship within group members. The sequence of recommendation can greatly influence user's feelings and overall satisfaction also.

If we want to model user's behavior and predict his/her satisfaction to related content, it is necessary to discover the real behavior. To study the emotional contagion, the social status impact, optimal recommendation sequence and aggregation functions used by users we designed a questionnaire given to PeWe Workshop participants.

In total 39 respondents were asked to answer 4 questions based on simulated designed group of three people wanting to watch film together (Table 1).

Respondents were divided into four groups, which differed in social status of group members – Daniel was older then rest of the group, Daniel had a birthday and Daniel was a company director.

Table 1. Simulated group with preferences.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Peter	10	4	3	6	10	9	6	8	10	8
Daniel	1	9	8	9	7	9	6	9	3	8
Lucia	10	5	2	7	9	8	5	6	7	6

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 87–88.

We compared obtained results with similar experiment [2] in order to discover some anomalies. Firstly, the aggregation strategy used by real users was studied. The average, average without misery and least misery strategies were used by most users. This is identical to Masthoff experiment. Interesting is, that when the social status changed to the job position and special occasion approval voting was the second most used aggregation strategy followed by "no" strategy (dictatorship with other members consideration).

Next the emotional contagion was studied. Users were asked to answer simple question: "If you are watching a movie with someone who not enjoys it, will this change your satisfaction?" This question discovers negative emotional contagion, while the reverse question, the positive emotional contagion respectively. It is interesting that most of users report the negative emotional contagion, while Masthoff shows that the positive emotional contagion occurs more often.

Last question investigated the optimal sequence for recommendation. We proposed four basic sequences: average to good ratings (B2); good to average ratings (C3); average – good – average ratings (D4); good – average – good ratings (A1). As we can see the sequence A1 is most preferred and brings high satisfaction followed by B2 and C3, while D4 brings no satisfaction (Figure 1).

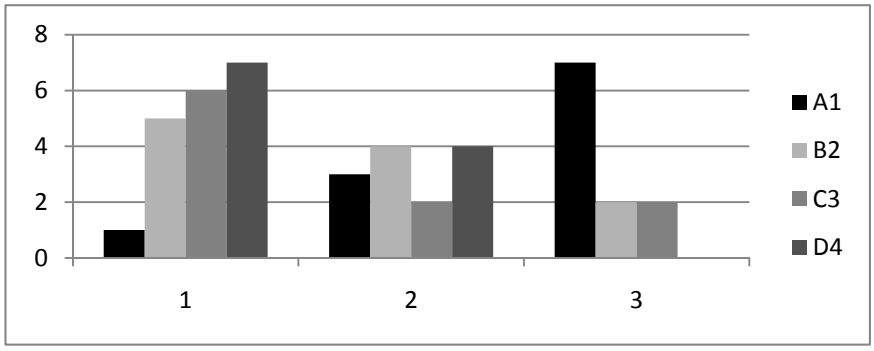


Figure 1. Satisfaction with sequence of recommendation.

Results clearly show that users consider social status or special occasions of group's members. We also discover optimal sequence order when as most important is beginning and the end of sequence. We support Masthoff in the most used aggregation function (average, average without misery and least misery), but users considering special occasions e.g. birthday use also approval voting.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Sumiyoshi H, Sano M, Goto J. CurioView: TV recommendations related to content being viewed. *(BMSB), 2010 IEEE*. 2010.

[2] Masthoff J, Gatt A. In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *UMUAI*. 2006.

Seminar Bingo at PeWe Workshop

Jakub ŠIMKO

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

As an accompanying event of our workshop, we conducted the *Seminar Bingo* event. This recessive event was inspired by the popular *PHD Comics* series[1], which describes the life of the academic researchers in a humorous, appealing way. Such kind of humour is also appreciated in our community and the event was therefore cheerfully welcomed.

The *Seminar Bingo* is a game principle applicable for various seminaries, conferences or other events, where speakers give their presentations and the audience asks questions. During such occasions, various events like disturbances, noise, speech mistakes or technical defects are occurring normally. These events are the basis for *Seminar Bingo*, where players – attendees of the seminar – receive game sheets with a grid of 5x5, with description of such common events in each cell. As the seminar proceeds, the players may mark the events on their sheets upon occurrence. When a first player has a full column or a row or a diagonal marked out, he is the winner of the game.

At our workshop, each participant received a game sheet, similar to one in the Table 1 (with shuffled cells). We used events tailored to our own research group and printed them easily by a generic bingo generator[2]. The game was played during the whole workshop, ending with a victory of a doctoral student Ján Suchal, who was the first in managing to achieve the winning conditions.

We consider the game as being welcome and refreshing for the workshop participants. It also serves as a good attention keeper – the players, eager to win, pay more attention on what is happening and are less distracted. Also, especially junior members of the research community could learn about bad practices during presentations and avoid them in the future. For those reasons, we believe that this game is suitable for casual events and groups similar to ours. We, however, discourage it during formal events, as it requires an agreement of all participants, especially speakers.

[1] http://www.phdcomics.com/comics/archive.php?comicid=847
[2] http://print-bingo.com

Spring 2011 PeWe Workshop, April 1, 2011, pp. 89–90.

Table 1. An example of the Seminar Bingo sheet used on the PeWe workshop.

B	I	N	G	O
The "meta-question" was asked	The first question after presentation is asked by chair	Speaker realizes, that he already told the content of a slide he currently turned in	"We haven't finished that yet"	A student bobs head fighting asleep
Speaker bashes related work	A student sleeps (a superman pose)	"Maybe I do not understand this correctly, but…"	A student was just taught a lesson by some postdoc	"Do you really think that…"
Loud boot sequence of an OS	A question to audience that is not answered	The live presentation of a software fails	Unscrupulous speech interpose	One of the professors falls asleep
Blatant typo	The graphics on the slide are messed up	"I was not paying attention, but…"	"Kúsoček" (a special quote of a certain member of the group)	Presentation is made in the prezi.com tool
"Hmmm"	Speaker runs out of time	For more than 5 min, the discussion is about an unimportant detail	A proper language usage interruption	The speaker uses inappropriate amount of gesticulation

SeBe 4.0: Towards Ubiquitous Savouring

Marián ŠIMKO, Michal BARLA, Jakub ŠIMKO

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
{simko,barla,jsimko}@fiit.stuba.sk

The Beer Driven Research (BDR) phenomenon has had a glorious existence of over past years and has revolutionized the way we work and live [2, 3, 4]. It has been a catalyst for proliferating research across boundaries, enabling necessary stimuli leading to a beerformation based society. For the fourth time in the history, researchers, practitioners, and policy makers of BDR from the whole country met together and exchanged on the current challenges and advances in the field.

The call for bottles announced the upcoming Populate the Beertology project supported by the Eight Framework Programme[1]. In order to participate in the project, organizing and programme committee agreed to extend the number of tracks, which matched the number of concepts (98) of upper layer beertology. A long-term goal of the project is to improve and outperform state-of-the-art approaches in the field of ontology population [1].

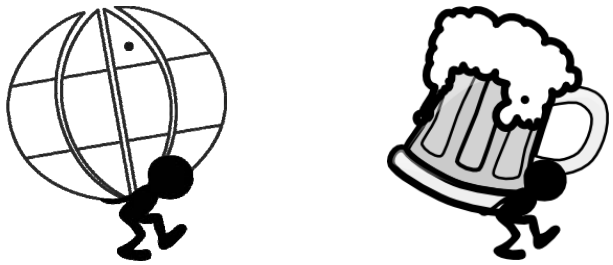
During the workshop opening, a philanthropy mission of SeBe initiative was presented. SeBe realizes its responsibility in the current anthroposociogenetic context. We esteem values and aim to support research. Thus, SeBe decided to grant a MVEC (Most Valuable Experiment Contributor) award for young researchers – PeWe members – significantly contributing to and helping with experiments of their peers.

The main part of the workshop started with traditional LittleBeerGame® introducing the beer in playful and attractive way. Four fellows accepted the challenge and competed in degustation of the essential substance of BDR. Not only they earned respect, enlightened by the degustation, Tomáš Kramár, Marián Hönsch, Štefan Mitrík and Martin Lipták zealously participated in the workshop until the very end.

The great part of the programme was covered by a collaborative social game with a purpose named *ya SeBeC* (Yet Another SeBe Competition). Finding inspiration in the ancient times, it divided participants into the two teams: the Lager house and the Ale house. Both houses were instructed to (virtually) get rid of other house's members. Motivated by the fact that game winners will become members of *best submission jury*, members discussed and collaborated on current research challenges, while eliminating

[1] The Eight Framework Programme (FP8) is multinational chief instrument for funding beer driven research and technological development over the period 2011 to 2012.

Spring 2011 PeWe Workshop, April 1, 2011, pp. 91–92.



(a) *(b)*

Figure 1. Original PeWe logo and SeBe logo representing both the burden of research we carry and the vision of coping with it (b).

their competitors. The game increased interaction intensity and contributed to (subconscious) formation of new research communities. At the end, Ján Suchal, Jozef Tvarožek, Martin Virik and Peter Študent were recognized as absolute winners. They had the privilege to degust submissions and to select the best of them. They agreed to give two prizes for the most valuable contributions to the field of BDR:

1. Peter Bartalos – best bottle award,
2. Martin Jačala – best runner-up.

At SeBe 4.0, experts and young researchers in Beer Computer Science, Beer Education, Beer Psychology, Beer Cognitive Science, and Beer Social Science, as well as entrepreneurs had the opportunity to establish collaborations, strengthen their links and cross-fermentize their core disciplines. SeBe pushed further the Ubiquitous Savouring paradigm by not only tackling the challenges of exploiting new trendy and tasty brands (in various contexts), but also by investigating ways to meet and support formally and informally the BDR devotees in their research playgrounds and social environments thanks to innovative scenarios.

Acknowledgement. This work has been totally supported by organizing chairs, selected SeBe devotees: Ján Suchal, Tomáš Kramár, Dušan Zeleník, Michal Kompan, Michal Holub and *all* SeBe submitting authors. We would like to thank Mária Bieliková for giving us opportunity to organize SeBe at PeWe workshop, and for her great support.

References

[1] Cimiano, P. Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer-Verlag, 347p (2006).
[2] Šimko, M. Barla, M. SeBe: Merging Research with Practice. November 2009, Modra, Harmónia, Slovakia (2009).
[3] Šimko, M. Barla, M. SeBe 2.0. The Emergence of Beer Science. April 2010, Smolenice, Slovakia (2010).
[4] Šimko, M. Barla, M., Šimko, J. Zeleník, D. SeBe 3.0: Means of Beernovation. November 2010, Modra, Harmónia, Slovakia (2010).

Index

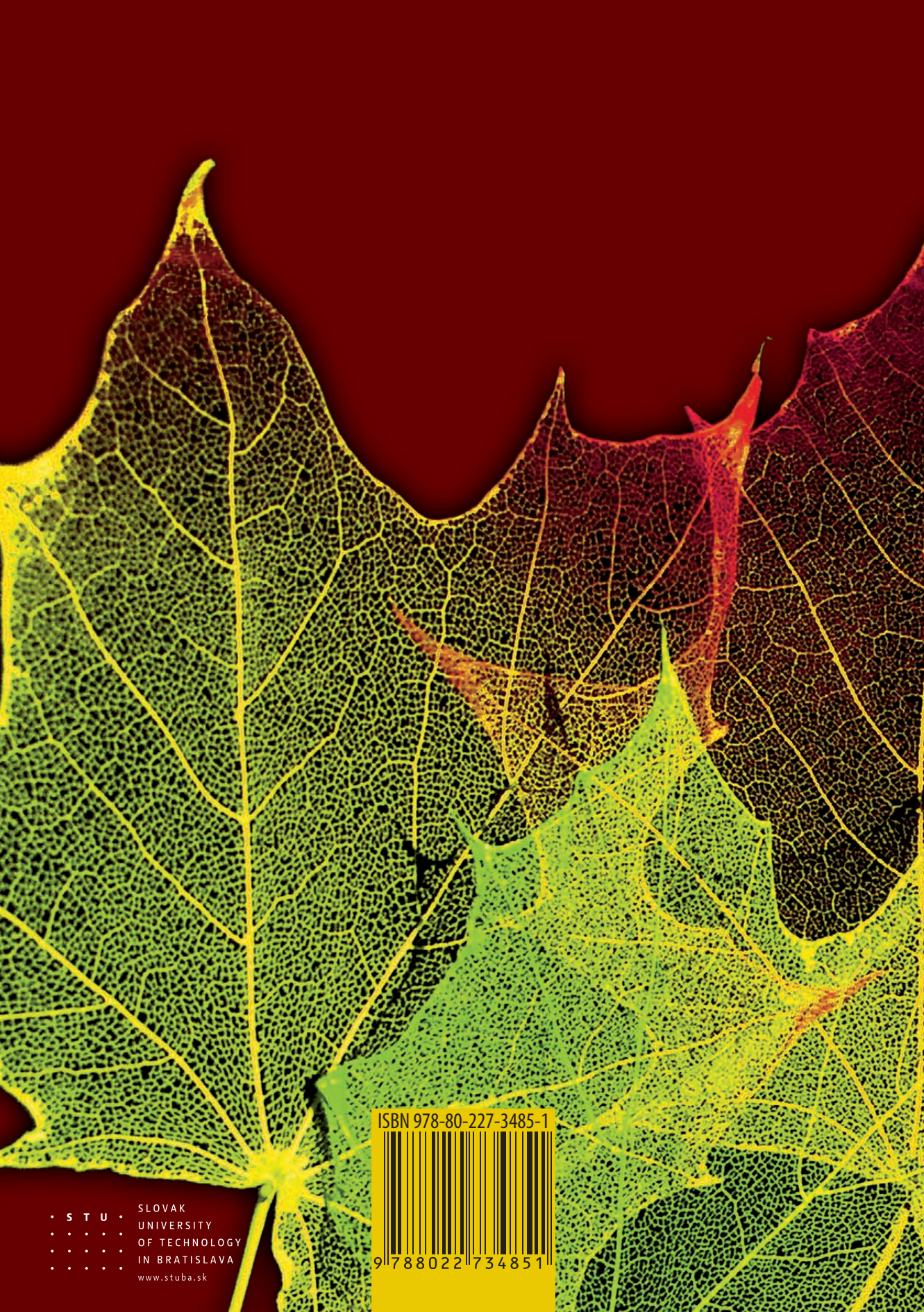
Mária Bieliková, Pavol Návrat,
Michal Barla, Jozef Tvarožek. Michal Tvarožek (Eds.)

Personalized Web – Science, Technologies and Engineering
9th Spring 2011 PeWe Workshop

1st Edition, Published by
Slovak University of Technology in Bratislava

110 pages, 50 copies
Print Nakladateľstvo STU Bratislava
2011

ISBN 978-80-227-3485-1



ISBN 978-80-227-3485-1

S T U
SLOVAK
UNIVERSITY
OF TECHNOLOGY
IN BRATISLAVA
www.stuba.sk