Proceedings of the 1st Workshop on Semantic Personalized Information Management

SPIM 2010

Workshop Organizers

Ernesto William De Luca
Technische Universität Berlin, DAI Labor, Germany
ernesto.deluca@dai-labor.de

Aldo Gangemi
Italian National Research Council (ISTC-CNR), Italy
aldo.gangemi@cnr.it

Michael Hausenblas
National University of Ireland, DERI, Ireland
michael.hausenblas@deri.org

Technical Chair: Till Plumbaum
Technische Universität Berlin, DAI Labor, Germany
till.plumbaum@dai-labor.de

Preface

Search engines have become an essential tool for the majority of users for finding information in the huge amount of documents contained in the Web. Even though, for most ad-hoc search tasks, they already provide a satisfying performance, certain fundamental properties still leave room for improvement. For example, if users perform general questions, they get frequently lost in navigating the huge amount of documents returned and typically stop their search after scanning a couple of result pages. Basically, results are ranked based on word frequencies and link structures, but other factors, such as sponsored links and ranking algorithms, are also taken into account.

Standard search engines do not consider semantic information that can help in recognizing the relevance of a document with respect to the meaning of a query, so that users have to analyze every document and decide which documents are relevant with respect to the meaning implied in their search. Therefore, they also struggle for matching the individualized information needs of a user.

Since users are different, and want to access information according to their experience and knowledge, different techniques for constructing user models, analyzing user profiles and deriving information about a user for the adaptation of content have been proposed. An emerging approach is to use Semantic Web and Web 2.0 technologies to model information about users.

Program Committee

We would like to express our thanks to the members of the program committee that helped in putting the program together and enhancing the quality of the workshop

Sahin Albayrak, Technische Universität Berlin, Germany
Lora Aroyo, Free University of Amsterdam, The Netherlands
Korinna Bade, University of Magdeburg, Germany
Jie Bao, Rensselaer Polytechnic Institute, USA
Diego Berrueta, Fundación CTIC, Spain
Charles Callaway, University of Haifa, Israel
Juan Manuel Cigarran, National Distance Learning University (UNED), Madrid, Spain
Richard Cyganiak, DERI Galway, Ireland
Harry Halpin, University of Edinburgh, Scotland
Tom Heath, Talis Information Ltd, United Kingdom
Leonhard Hennig, Technische Universität Berlin, Germany
Fabien Gandon, INRIA, France
Hugh Glaser, University of Southampton, UK
Alfio Gliozzo, Semantic Technology Laboratory, Rome
Lalana Kagal, MIT, Cambridge, MA, USA
David Karger, MIT, Cambridge, MA, USA
Michael Kruppa, DFKI GmbH, Berlin, Germany
Andreas Lommatzsch, Technische Universität Berlin
Peter Mika, Yahoo! Research Barcelona, Spain
Eric Prud'hommeaux, MIT/W3C, USA
Georg Russ, University of Magdeburg, Germany
Frank Rügheimer, Pasteur Institute, Paris, France
Conor Shankey, Visual Knowledge, Vancouver, B.C., Canada
Armando Stellato, University of Tor Vergata, Rome
Susie M. Stephens, Johnson and Johnson, USA
Denny Vrandecic, DFKI GmbH, Germany
Ana García-Serrano, National Distance Learning University (UNED), Madrid, Spain

Program

9:00 - 9:30 Opening and Greetings (Workshop organizers)

9:30 - 10:00 Keynote (Organizer)

10:00 - 10:30 ADAPTIVE USER-CENTRIC LAYOUT OF NEWS WITH AESTHETIC CONSTRAINTS
Thomas Strecker and Sahin Albayrak

10:30 – 11:00 Coffee Break

11:00 – 11:30 POPULATION AND ENRICHMENT OF EVENT ONTOLOGY USING TWITTER
Shashi Narayan, Srdjan Prodanovic, Zoë Bogart and Mohammad Fazleh Elahi

11:30 – 12:00 PERSONALIZED LEARNING TO RANK APPLIED TO SEARCH ENGINES
Christian Scheel, Ernesto William De Luca, Sahin Albayrak, Sascha Narr, Orlando Macone, Dennis Egert and Matthias Hohner

12:00 – 12:30 PERSONALIZED MULTI-DOCUMENT SUMMARIZATION USING N-GRAM TOPIC MODEL FUSION
Leonhard Hennig and Sahin Albayrak

12:30 – 13:00 SUMMARIZATION AND VISUALIZATION OF DIGITAL CONVERSATIONS
Vincenzo Pallotta, Rodolfo Delmonte and Marita Ailomaa

13:00 – 14:30 Lunch Break

14:30 – 15:00 SEMANTIC TV ENGINE: AN IPTV ENABLER FOR PERSONALIZED RECOMMENDATIONS
Juri Glass, Torsten Schmidt, Stefan Marx and Fikret Sivrikaya

15:00 – 15:30 A DESKTOP-INTEGRATED SEMANTIC PLATFORM FOR PERSONAL INFORMATION MANAGEMENT
Noemi Scarpato, Maria Teresa Pazienza, Armando Stellato and Andrea Turbati

15:30 – 16:00 Coffee Break

16:00 – 17:00 Open Discussion

Table of Contents

A Desktop-Integrated Semantic Platform for Personal Information Management

Maria Teresa Pazienza, Noemi Scarpato, Armando Stellato, Andrea Turbati

ART Research Group, Dept. of Computer Science,
Systems and Production (DISP) University of Rome, Tor Vergata
Via del Politecnico 1, 00133 Rome, Italy
{pazienza, scarpato, stellato, turbati}@info.uniroma2.it

Abstract

The Semantic Web dream of a real world-wide graph of interconnected resources is – slowly but steadily – becoming a concrete reality. Still, the whole range of models and technologies which will change forever the way we interact with the web, seems to be missing from every-day technologies available on our personal computers. Ontologies, annotation facilities and semantic querying could (and should) bring new life to Personal Information Management, supporting users in contrasting the ever-growing information overload they are facing in these years, overwhelmed by plethora of communication channels and media.

In this paper we present our attempt in bringing the Semantic Web Knowledge Management paradigm at the availability of diverse personal desktop tools (Web Browser, Mail clients, Agenda etc…), by evolving Web Browser Semantic extension Semantic Turkey to an extensible framework providing RDF data access at different levels: java access through OSGi extensions, HTTP access or dedicated JavaScript API for the whole range of tools from the open source suite of Mozilla applications

1. Introduction

The Semantic Web is becoming ever and ever a concrete reality: with SPARQL reaching W3C recommendation early this year (Prud'hommeaux, 2008), languages for data representation and querying have finally reached standardization, while interests and research in Semantic Web technologies have definitely migrated from mere ontology development (which has now met industry standards) aspects to the discovery and devise of applications which can both show and exploit Semantic Web full potential.

Despite this encouraging trend of Semantic Web models and technologies, these seem to be missing from applications which we use every day on our personal desktop computers. Hopefully, they could surely contribute to improve the quality of personally managed data by supporting users with powerful vocabularies (ontologies) which can be extended (by adapting them to personal needs) and shared through different applications and with other people.

Recently, several efforts have been spent towards definition of applications and solutions for implementing the so called Semantic Desktop (Iturrioz, Díaz, Fernández Anzuola, & Azpeitia, 2003; Sauermann, 2005; Groza, et al., 2007).

All the Semantic Desktop approaches cited above usually aim at centralizing an RDF Semantic Repository as a local information management resource, which can be accessed by diverse applications on the desktop sharing common data but providing different services over them.

In this work, we present our proposal for a Semantic Integrated Environment for the Mozilla suite (though it can be exploited also by other applications) of desktop utilities (Firefox, Sunbird, Thunderbird etc…). This project originated from our ontology tool Semantic Turkey (Griesi, Pazienza, & Stellato, 2007), which was originally thought as a Semantic extension for the Firefox Web Browser and lately evolved into a multi-layered extensible framework for Knowledge Management and Acquisition.

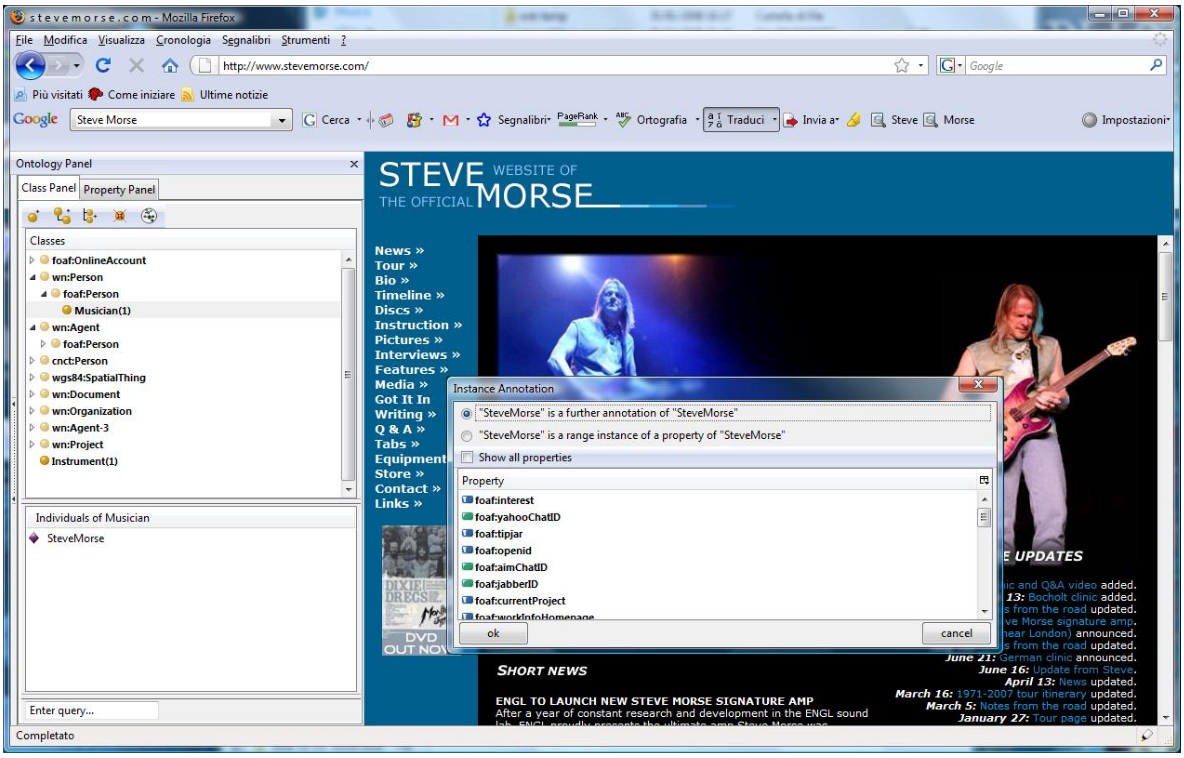
The current framework which still backbones Semantic Turkey, is two-fold in its offer: by first, being of interest for ontology developers and domain experts, since it aims at facilitating the process of knowledge acquisition and development, and, on the other side, providing an extensible infrastructure over which SW applications, needing and relying on rock-solid web browsing functionalities as well as on RDF management capacities, can be developed and deployed. In this paper we present the different service layers which are exposed by current version of Semantic Turkey, and how they can be accessed by Mozilla-based and other external applications to give life to a new multimodal Semantic Desktop.

2. Other works

Beside the main research stream which is conducted in this field, other researchers are focusing on finding new powerful and versatile ways of interaction with the user, which can exploit the advanced possibilities given by the Semantic Desktop. as in (Iturrioz, Díaz, & Fernández Anzuola, 2008) where the seMouse (Semantic Mouse) offers a Mouse extension (cabled at Operating System level) allowing for easy classification, authoring, retrieval etc… of files on the desktop and of their textual content. Since it is acting at OS level, this mouse extension is not limited to any specific working environment/application: no matter whether the user is working with Word, Power-Point, Netscape, etc, the semantic button is available for annotation/authoring and the user does not have to move to a new dedicated editor when annotating.

Though intuitions such as the one of seMouse centered the limitations of past approaches with respect to their concrete usability in real life, most recent trends tend to favor the centralization of core knowledge services, thus giving the possibility to all desktop applications to feature even very specific and advanced functionalities while interacting together with (and possibly be coordinated by) the central semantic repository.

The most recent (and sensible) effort following this trend has been represented by the FP6 EU funded project NEPOMUK (Groza, et al., 2007) where a massive range


stevemorse.com - Mozilla Firefox
File Modifica Visualizza Cronologia Segnalibri Strumenti ?
http://www.stevemorse.com/
Più visitati Come iniziare Ultime notizie
Google Steve Morse Cerca Segnalibri PageRank Ortografia Traduci Invia a Steve Morse Impostazioni
Ontology Panel
Class Panel Property Panel
Classes
foaf:OnlineAccount
wn:Person
foaf:Person
Musician(1)
wn:Agent
foaf:Person
cnct:Person
wgs84:SpatialThing
wn:Document
wn:Organization
wn:Agent-3
wn:Project
Instrument(1)
Individuals of Musician
SteveMorse
Enter query...
Completato
STEVE MORSE WEBSITE OF THE OFFICIAL
News » Tour » Bio » Timeline » Discs » Instruction » Pictures » Interviews » Features » Media » Got It In Writing » Q & A » Tabs » Equipment Store » Contact » Links »
Instance Annotation
Property
foaf:interest
foaf:yahooChatID
foaf:tipjar
foaf:openid
foaf:aimChatID
foaf:jabberID
foaf:currentProject
foaf:workInfoHomepage
ok
cancel
UPDATES
SHORT NEWS
ENGL TO LAUNCH NEW STEVE MORSE SIGNATURE AMP
DIXIE DREGS DVD OUT NOW

● "SteveMorse" is a further annotation of "SteveMorse"

○ "SteveMorse" is a range instance of a property of "SteveMorse"

□ Show all properties

Fig. 1 Semantic Bookmarking with Semantic Turkey

of technologies comprehended several extensions for existing applications centered around an RDF Data server activated by the Operating System.

Eventually, a Semantic Desktop could probably rely on a combination of both approaches, which are not in contrast with each other.

Another important aspect of research is the definition of the metamodels which should contradistinguish such powerful organization systems: in PIMO (Sauermann, van Elst, & Dengel, 2007) a multilayered ontology model is presented. The PIMO (Personal Information Models) Ontology offer a first distinction between three conceptual categories: Native Resources (files, e-mails, contacts etc…), Native Structures (representing organizational schemas for the above, such as folders, bookmark folders, tags etc…) and lastly the Mental Model provides a cognitive representation of the knowledge a user is intended to manage, which is indipendent of (though may be linked to) the above.

PIMO is the structured according to five layers which account for different levels of specification (such as for the first three levels: *PIMO-Basic*, *PIMO-Upper* and *PIMO-Mid*) as well as for the specific exigencies of the user (*PIMO-User*) and of the working/social environment where he acts (*Domain ontologies*).

The necessity for addressing different facets of knowledge in organization systems is also present (though in a less general perspective, which is specifically aimed at enterprise organizations) in (Apostolou, Mentzas, & Abecker, 2008), where a single Knowledge Object (KO) may be characterized according to descriptors which are provided by different facets of the whole ontology. These facets are: Business, Domain, Community, Context and Content, describing *where* a KO may be used, according to *which conditions* its use is suggested, the *range of users* which may be interested in it, and the like.

3. From Semantic Bookmarking to Knowledge Management and Acquisition

Semantic Turkey was born inside a national project – funded by the FILAS agency (Finanziaria Laziale di Sviluppo) under contract C5748-2005 – focused on innovative solutions for browsing the web and for collecting and organizing the information observed during navigation.

The prototype for the project immediately took the form of a Web Browser extension allowing users to annotate information from visited web sites and organize it according to a personally defined domain model: Semantic Turkey paradigmatic innovation was in fact to "obtain a clear separation between (acquired) knowledge data (the WHAT) and web links (the WHERE)" pointing to it. That is, to be able, through very easy-to-use drag'n'drop gestures, to *select* textual information from web pages, *create* objects in a given domain and *annotate* their presence in the web by keeping track of the selected text and of its provenience (web page *url*, *title* etc…). We coined the expression "semantic bookmarking" for this kind of activity.

Due to its proverbial extendibility, the Firefox platform (http://www.mozilla.com/en-US/firefox/) had been chosen as the hosting browser for our application, while Semantic Web standards and technologies were the natural candidate for representing its knowledge model.

Semantc Turkey (Fig 1) was thus born. Standing on top of mature results from research on Semantic Web technologies, like Sesame (Broekstra, Kampman, & van Harmelen, 2002) and OWLim (Kiryakov, Ognyanov, &

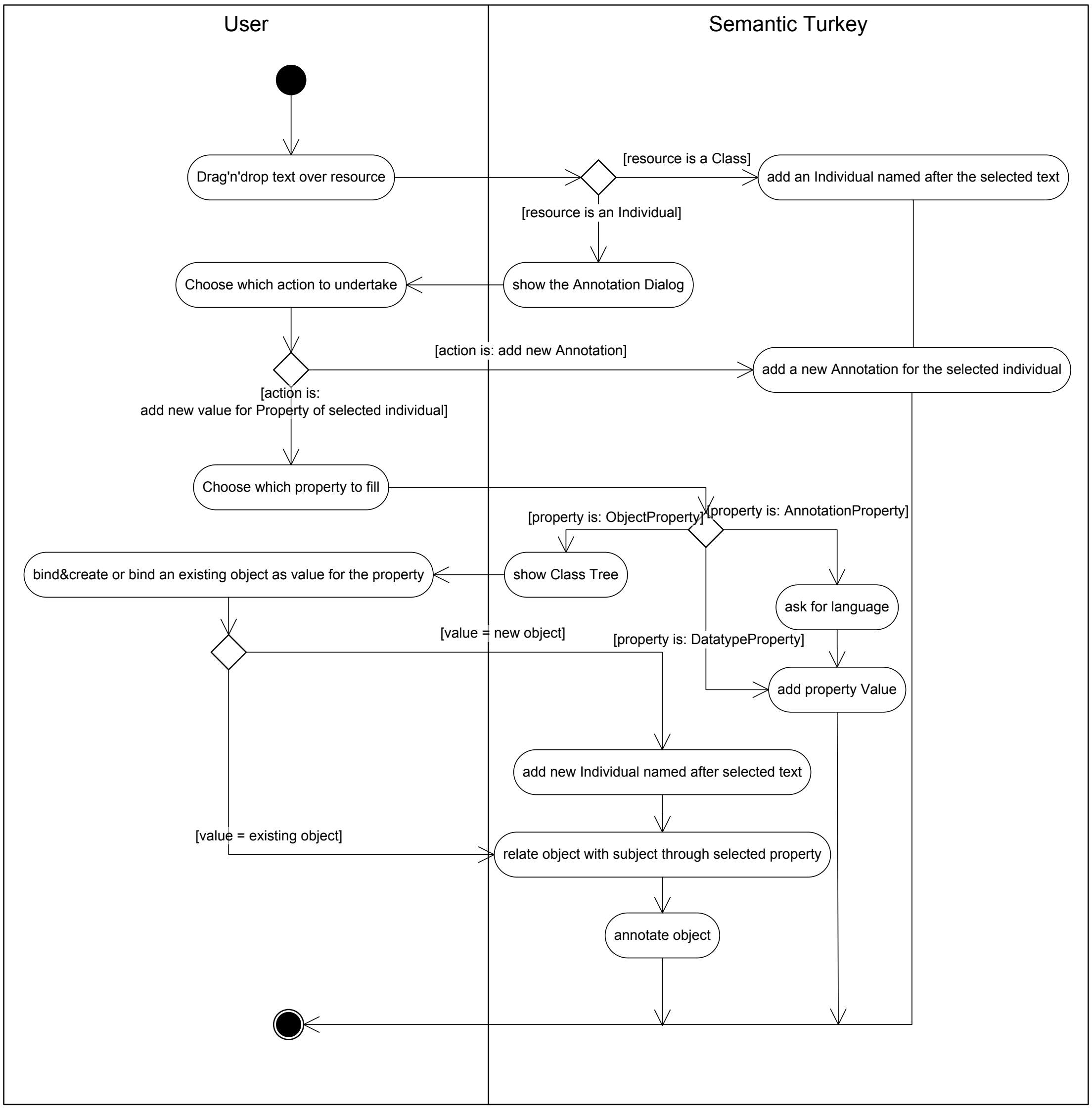

User
Semantic Turkey
Drag'n'drop text over resource
[resource is a Class]
add an Individual named after the selected text
[resource is an Individual]
show the Annotation Dialog
Choose which action to undertake
[action is: add new Annotation]
add a new Annotation for the selected individual
[action is:
add new value for Property of selected individual]
Choose which property to fill
[property is: ObjectProperty]
[property is: AnnotationProperty]
show Class Tree
bind&create or bind an existing object as value for the property
ask for language
[value = new object]
[property is: DatatypeProperty]
add property Value
add new Individual named after selected text
[value = existing object]
relate object with subject through selected property
annotate object

Fig. 2 Activity diagram for semantic bookmarking/annotation

Manov, 2005) as well as on a robust platform such as the Firefox web browser, ST (Semantic Turkey) differentiated from other existing approaches which are more specifically tailored respectively towards knowledge management and editing (Gennari, et al., 2003), semantic mashup and browsing (Dzbor, Domingue, & Motta, Magpie: Towards a Semantic Web Browser, 2003; Huynh, Mazzocchi, & Karger, 2005) and pure semantic annotation (Ciravegna, Dingli, Petrelli, & Wilks, 2002; Kahan & Koivunen, 2001), by introducing a new dimension which is unique to the process of building new knowledge while exploring the web to acquire it.

By focusing on this aspect, we went beyond the original concept of Semantic Bookmarking and tried to amplify the potential of a new Knowledge Management and Acquisition System: we thus aimed at reducing the impedance mismatch between domain experts and knowledge investigators on the one side, and knowledge engineers on the other, providing them with a unifying platform for acquiring, building up, reorganizing and refining knowledge.

Fig. 2 shows the different annotation/knowledge acquisition possibilities offered by the functionalities based on interaction with the hosting web browser. In the new version of ST, support for all kind of properties has been introduced and reflected in the bookmarking facility: when a portion of text is selected from the page and dragged over an individual, the user may choose (as in the old version) to add a new annotation for the same individual or to use the annotation to fill one property slot for it. In the second case, the user can now choose from a list of properties (see small window in) the one which will be filled: this list includes those properties having their rdfs:domain including one of the types of the selected instance, but may be extended to cover all properties (letting the inference engine do the rest). If the property selected for enrichment is an object property, the user is prompted with a class tree (rooted on the rdfs:range of the selected property) and is given the possibility of creating a new individual named after the text selected for the annotation or to choose an existing one: in both cases the selected individual is bound –

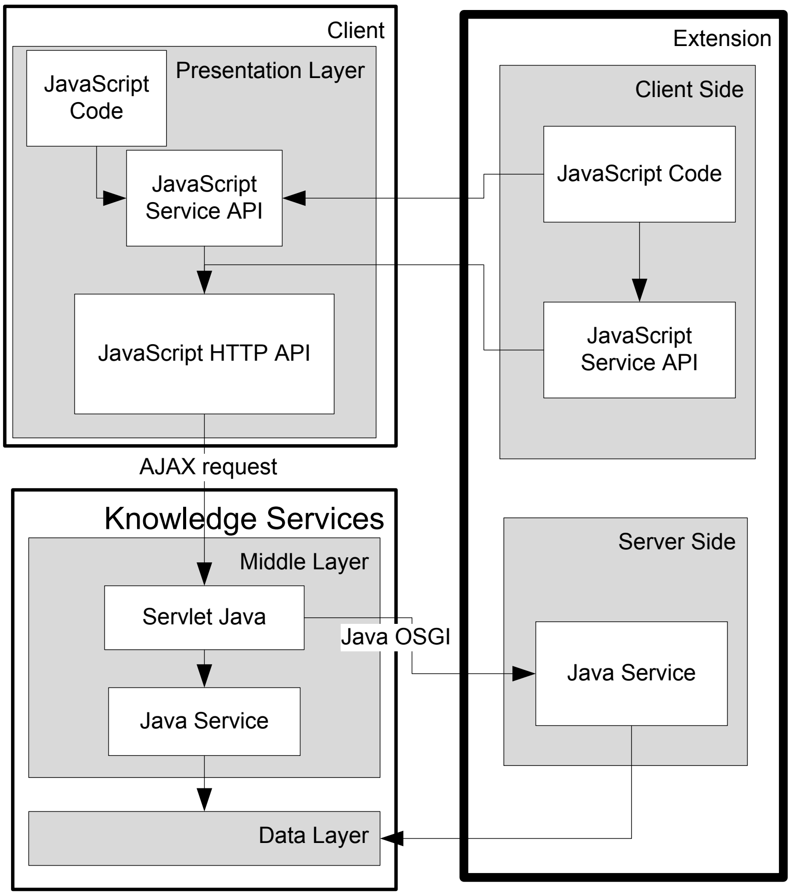


Fig. 3 Architecture of the different Access layers for Mozilla Semantic Desktop

through the chosen property – to the one where he originally dropped the text; a bookmark is also added for it, pointing to the page where the object has been observed. Even in this case, the user may choose to visualize the entire class tree and not the one dominated by the range of the property: the inference engine will automatically assign the pointed instance to that range.

The above interaction modalities for knowledge acquisition/annotation/bookmarking can be used in the main Ontology Editor tool, as well as be exported as *pluggable functional objects*, into other client applications willing to adopt them in simpler user-centered environments for Personal Data Management. The next sections describes the different service layers which are available through Semantic Turkey and how they can be used to propel Semantic based desktop applications.

4. Service Layers for Applications

The main underlying application consists of an RDF framework made of an HTTP application server (which in Semantic Turkey is automatically started through Firefox) based on Java technologies and of a set of client layers facilitating access by users or third party applications.

The whole extension mechanism of the framework is implemented through a proper combination of the Mozilla extension framework (which is used to extend the user interface, drive user interaction, add/modify application functionalities and provide javascript API for the whole set of Mozilla desktop utilities) and the OSGi java extension framework (OSGi RFC0112, 2005) which provides extension capabilities for the service and data layers of the architecture. A comprehensive description of Semantic Turkey architecture can be found in (Griesi, Pazienza, & Stellato, 2007) and in (Pazienza, Scarpato, Stellato, & Turbati, 2008). In this section we focus instead on the different layers (see Fig. 3 above) and extension points which characterize Semantic Turkey as an open RDF framework with specialized functionalities for Personal Information Management.

4.1. Javascript extensibility

Thanks to javascript dynamic programming paradigm, where functions are first-class citizens of the language, functionalities such as the annotation resolver described in section 3, can be dynamically imported and associated to logically coherent events in different client applications of the Mozilla suite. The *pluggable functional objects* mentioned in section 3 can thus be considered independent components which can be exported and be reused in web browser as well as in email clients. For example, highlighting text in a web page within Firefox, and dropping it over a class, could invoke the same behavior when selecting text in emails from within Thunderbird. Conversely, reacting to changes in the underlying knowledge could produce different effects depending on the client platform which is connected to the

Semantic Desktop: finding RDFa (Adida & Birbeck, 2007) data on a web page from within the web browser, detailing scheduled events, could lead to the import of that data inside the semantic desktop's ontology, and the consequent export of this data inside other desktop applications for calendar management such as Lightning or Sunbird[1].

4.2. OSGi extensibility

OSGi compliance is obtained through the OSGi implementation developed inside the Apache Software Foundation, called Felix (felix.apache.org/).

Two main extension points have been introduced: an *OntologyManager Extension* and a *Service extension*.

The *OntologyManager* Extension point allows different triple-store technologies implementing low level RDF data storage, to be plugged to the system. Current implementations provide support for Sesame2, OWLIM and Jena (McBride, 2001) – through its NG4J extension (Bizer, Cyganiak, & Hartig) supporting named graphs – technologies.

The *service extension* point allows new java services to be plugged to the system, this way further desktop applications can automatically deploy and add their functionalities to the main service.

The set of services offered by the Knowledge Server provide high-level, macro operations, other than standard ontology management ones. The pure triple-level RDF data layer is not obfuscated by macro-operations, and is directly accessible through java API as well as replicated in a set of basic knowledge services for RDF manipulation.

A third extension point allows for the registration of plug-ins: these act as collectors for set of services sharing a common logical ratio. While standard service extensions are sort of add-ons to the main application and are always available unless deactivated or removed, extensions bound to plug-ins are activated/deactivated according to the status of the plug-in. Plug-ins are assigned to projects and their status and persistent information is stored with the metadata for each project.

The project-based behavior of the platform comes from its ontology-editor ancestry, while when it is being used as Semantic Desktop Server, a single project (called *main-project*), is always active and automatically started at system initialization. Each application based on the Semantic Desktop and needing customized services thus registers itself as a plug-in and installs all of its required services via OSGi.

Finally, a data extension point allows for the declaration of *support* and *application* ontologies which are loaded by the system to drive its behavior and the one of its extensions and connected applications. These ontologies are not treated the same way as imported domain/user ontologies and are explicitly registered for their role. Registering an ontology through this extension point has a variety of consequences: these are loaded automatically at system startup even if they are not explicitly imported by the edited domain ontologies and application ontologies' content (and content classified after application ontologies' concepts) is not shown explicitly but only managed and exposed indirectly through applications' services.

We enabled this classification of ontologies since all the data which is available through the Mozilla Semantic Desktop (MSD from now on) is available as RDF triples: it was thus mandatory to separate the knowledge which is being managed by the user, from the one which is being used by the Semantic Desktop to coordinate its activities. Despite this "conceptual separation" – ontology spaces are managed through the use of *named graphs* (Carroll, Bizer, Hayes, & Stickler, 2005) – having a single RDF cauldron where all triples are being stored allows for more tight connection between these spaces, so that, for example, data in the application space could be used to organize the domain information according to different *facets*, or add annotations which should not be available as domain ontology. As an example of application ontology, the basic version (i.e. no extensions installed) of MSD declares an *application ontology* called *Annotation*[2] describing the textual occurrences from which entities submitted by the user have been annotated, together with details about the document (type of document, url for web pages, title etc…) where these annotations have been taken. An example of *support ontology* is instead provided by the Sesame2 implementation of the *OntologyManager* extension point: Sesame2 library does not support OWL reasoning nor includes the OWL vocabulary; since Mozilla Semantic Desktop relies on the OWL vocabulary, this is being declared as a support ontology and dynamically added to the core knowledge.

Data defined upon vocabulary from the Annotation ontology (since it is an application ontology) is thus not shown by default in all ontology editing interfaces, and its content is made available to the user through MSD's functionalities (such as those for retrieving documents associated to ontology resources, or for highlighting all the annotations taken in a document), while resources from the OWL vocabulary (being it a support ontology) are shown but are kept separate from user data (owl vocabulary is not saved together with user data nor it is explicitly imported by user ontology).

4.3. HTTP Access

All of OSGi services are available via AJAX through HTTP request. The response to these requests is codified in XML or (in some cases) in JSON, depending on request type, available standards and compactness of the content.

Due to its complete platform/technology independence, this is the layer which can be exploited by any application which has no direct connection with the service layer and is not compatible with Mozilla technology.

4.4. Mozilla JavaScript API

Upon the above layer, a set of JavaScript API, completely hiding the HTTP request/response interaction, has been built by using Mozilla technology. These are the API which are currently used inside Semantic Turkey Semantic Web Browser.

These API are coded as exportable functions into Mozilla modules, a proprietary Mozilla solution for JavaScript allowing for persistence (JavaScript objects inside a module persist upon different imports of the same

[1] http://www.mozilla.org/projects/calendar/

[2] http://art.uniroma2.it/ontologies/annotation

module) and hiding/encapsulation (a module's developer must choose which objects/functions are exported by users of the module and which ones just serve as hidden internal machinery).

These JavaScript Modules (roughly paired with their service counterparts in the service layer) can thus easily be imported into any sheet of a Mozilla based application (or extension). In the following example:

```
Components.utils.import(
  "resource://stservices/SERVICE_Cls.jsm",semanticturkey
)
```

all the objects and functions exposed by the SERVICE_Cls module are imported into the variable semanticturkey: this is a good practice to prevent variable clashing, as Mozilla extensions share a common space where all script code (from main application and all of its extension) is pooled.

Once the above statement is explicated in a script document, API methods contained in SERVICE_Cls can be used in the same sheet, like in the following:

```
semanticturkey.STRequests.Cls.getInstanceList(clsName)
```

where all instances of class identified by clsName are retrieved and returned by the method.

HTTP masking is handled by a common module:

```
resource://stmodules/SemTurkeyHTTP.jsm
```

which is shared by all API methods. The SemTurkeyHTTP.jsm module contains convenience methods for composing GET and POST requests, for unmarshalling received XML/JSON over HTTP responses and recomposing them in terms of dedicated JavaScript objects.

Due to the masking of HTTP details by Mozilla JavaScript Semantic API, all of their methods return explicit JavaScript exceptions. These are classified as:

- *errors*: error JavaScript exceptions mask HTTP communication errors as well as exceptions thrown at run time by the invoked service and caught by the HTTP Server. Usually it is not easy for the common user to discover the problem which has been generated, and these kind of exceptions are considered as severe application faults
- *exceptions*: JavaScript exceptions marked as application exceptions are due to predictable java exceptions which occurred at server level. Usually they contain understandable messages which may be explicitly communicated to the user. Also, specific management of these exceptions depending on their type and the context where these occurred can be performed by the application invoking the method which threw them.

Developers of new applications based on the Mozilla framework can thus invoke the underlying services and handle exceptions depending on the context of invocation, thus following a traditional structured programming approach and producing readable "narrative scripting" code, instead of writing complex code for client-server interaction.

Application Developers willing to add further APIs for interfacing with their software, can extend the service layer through OSGi and then build new modules for the JavaScript API, relying on the common SemTurkeyHTTP.jsm infrastructure.

4.5. Reusable widgets for Semantic Applications based on this Mozilla Semantic Desktop

Applications exploiting the Mozilla Semantic Desktop which are based on the same Mozilla technology, can beneficiate of exportable widgets expressly dedicated to Ontology Management. We are currently expanding this aspect, which is currently limited to reusable widgets for class and property trees, and for resource editors (class, property, instance and ontology resource editor widgets) to cover a whole range of widgets for ontology maintenance and editing.

Also, to satisfy the more complex needs of end-user applications, which should hide the ontology editing aspects and show custom widgets more close to their specific nature, we are considering the addition of a dedicated UI generator based on the Fresnel model (Pietriga, Bizer, Karger, & Lee, 2006) for browser independent visualization of RDF graphs. Our UI generator will provide a Fresnel parser and UI generation facilities based on the XML User Interface Language XUL, which is adopted by the suite of Mozilla tools.

5. Conclusions

We have presented here our ongoing work for a fully-extensible RDF based platform realizing the Semantic Desktop paradigm.

The strength of Mozilla Semantic Desktop is not in the whole range of end-user services (which are currently limited to the Semantic Bookmarking services offered by its originating platform Semantic Turkey), but in the wide spectrum of connections that are exposed to future applications willing to interact with it.

A second point is on the depth and completeness of its ontology management capabilities, providing a solid platform with convenience methods for ontology editing, disburdening the application developer from the non-trivial effort of maintaining the underlying ontology. Keeping the RDF graph clean (free from potential redundancies and from dangling triples, i.e. triples out of the reachability of any application insisting on them) is in fact a non-trivial aspect from which applications should abstract and which is not supported by default triple-store systems. Advanced layers for RDF management should consider the kind of triple-store they are using, the level of reasoning which is supported (and, where necessary, the "trivial reasoning" which should be computed by them to present data in a readable way) etc.. to provide an homogeneous interaction layer for the application developer.

These "advanced management" requirements are not limited to pure graph maintenance. RDF/OWL pushed forward concepts such as explicit semantics, shareability and interconnectivity: platforms supporting shared knowledge for cooperation of RDF-based applications, should be able to provide powerful tools for meta-management: modularization, multi-faceted perspectives, visualization, are all fundamental aspects which should contradistinguish the layering of future RDF based frameworks, and in special case for Semantic Desktop Platforms.

References

Adida, B., & Birbeck, M. (2007, October 26). *RDFa Primer*. Retrieved from World Wide Web Consortium - Web Standards: http://www.w3.org/TR/xhtml-rdfa-primer/

Apostolou, D., Mentzas, G., & Abecker, A. (2008). Managing Knowledge at Multiple Organizational Levels Using Faceted Ontologies. *Journal of Computer Information Systems , Winter 2008-2009*, 32-49.

Bizer, C., Cyganiak, R., & Hartig, O. (Eds.). (n.d.). *NG4J - Named Graphs API for Jena*. Retrieved May 14, 2009, from NG4J - Named Graphs API for Jena: http://www.wiwiss.fu-berlin.de/suhl/bizer/ng4j/

Broekstra, J., Kampman, A., & van Harmelen, F. (2002). Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. *The Semantic Web - ISWC 2002: First International Semantic Web Conference* (p. 54-68). Sardinia, Italy: Springer Berlin / Heidelberg.

Carroll, J. J., Bizer, C., Hayes, P., & Stickler, P. (2005). Named Graphs, Provenance and Trust. *WWW '05: Proceedings of the 14th international conference on World Wide Web* (p. 613-622). New York, NY, USA: ACM Press.

Ciravegna, F., Dingli, A., Petrelli, D., & Wilks, Y. (2002). User-system cooperation in document annotation based on information extraction. *13th International Conference on Knowledge Engineering and Knowledge Management, EKAW02.* Springer Verlag.

Dzbor, M., Domingue, J., & Motta, E. (2003). Magpie: Towards a Semantic Web Browser. *2nd International Semantic Web Conference (ISWC03).* Florida, USA.

Gennari, J., Musen, M., Fergerson, R., Grosso, W., Crubézy, M., Eriksson, H., et al. (2003). The evolution of Protégé-2000: An environment for knowledge-based systems development,. *International Journal of Human-Computer Studies , 58* (1), 89–123.

Griesi, D., Pazienza, M. T., & Stellato, A. (2007). Semantic Turkey - a Semantic Bookmarking tool (System Description). *4th European Semantic Web Conference (ESWC 2007).* Innsbruck, Austria.

Groza, T., Handschuh, S., Moeller, K., Grimnes, G., Sauermann, L., Minack, E., et al. (2007). The NEPOMUK Project - On the way to the Social Semantic Desktop. In T. Pellegrini, & S. Schaffert (Ed.), *Proceedings of I-Semantics' 07* (pp. 201-211). JUCS.

Huynh, D., Mazzocchi, S., & Karger, D. (2005). Piggy Bank: Experience the Semantic Web Inside Your Web Browser. *Fourth International Semantic Web Conference (ISWC05),* (p. 413-430). Galway, Ireland.

Iturrioz, J., Díaz, O., & Fernández Anzuola, S. (2008). Toward the Semantic Desktop: The seMouse Approach. *IEEE Intelligent Systems , 23*, 24-31.

Iturrioz, J., Díaz, O., Fernández Anzuola, S., & Azpeitia, I. (2003). The Semantic Desktop: an architecture to leverage document processing with metadata. In S. Guier (Ed.), *Multimedia and Data Document Engineering (MDDE'03).* Berlin, Germany.

Kahan, J., & Koivunen, M.-R. (2001). Annotea: an open RDF infrastructure for shared Web annotations. *WWW '01: Proceedings of the 10th international conference on World Wide Web* (pp. 623-632). Hong Kong, Hong Kong: ACM.

Kiryakov, A., Ognyanov, D., & Manov, D. (2005). OWLIM – a Pragmatic Semantic Repository for OWL. *Int. Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2005), WISE 2005.* New York City, USA.

McBride, B. (2001). Jena: Implementing the RDF Model and Syntax Specification. *Semantic Web Workshop, WWW2001.*

OSGi RFC0112. (2005). Retrieved from http://www2.osgi.org/Download/File?url=/download/rfc-0112_BundleRepository.pdf

Pazienza, M., Scarpato, N., Stellato, A., & Turbati, A. (2008). Din din! The (Semantic) Turkey is served! *Semantic Web Applications and Perspectives.* Rome, Italy.

Pietriga, E., Bizer, C., Karger, D. R., & Lee, R. (2006). Fresnel: A Browser-Independent Presentation Vocabulary for RDF. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, et al. (Ed.), *The 5th International Semantic Web Conference (ISWC06). LNCS 4273*, pp. 158-171. Athens, GA, USA: Springer Verlag.

Prud'hommeaux, E. :. (2008, January 15). *SPARQL Query Language for RDF.* Retrieved from World Wide Web Consortium - Web Standards: http://www.w3.org/TR/rdf-sparql-query/

Sauermann, L. (2005). The Gnowsis Semantic Desktop for Information Integration. *1st Workshop on Intelligent Office Appliances(IOA 2005): Knowledge-Appliances in the Office of the Future.* Kaiserslautern, Germany.

Sauermann, L., van Elst, L., & Dengel, A. (2007). PIMO - A Framework for Representing Personal Information Models. In T. Pellegrini, & S. Schaffert (A cura di), *Proceedings of I-MEDIA '07 and I-SEMANTICS '07 International Conferences on New Media Technology and Semantic Systems as part of (TRIPLE-I 2007).* Graz, Austria.

Adaptive User-Centric Layout of News with Aesthetic Constraints

Thomas Strecker, Sahin Albayrak

DAI-Labor, TU Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{thomas.strecker,sahin.albayrak}@dai-labor.de

Abstract

Traditional newspapers are laid out by professionals and require a lot of manual effort which multiplies with every reuse of content. In contrast, the personalized news lists of a user are mostly just that: lists with pointers to content which is laid out in a wide range of styles. Electronic Paper proposes to bridge the gap between the traditional newspaper and the personalized news each user collects for their own use by allowing for adaptivity of content and layout.

In order to fill the gap, this work proposes an approach which incorporates the knowledge traditional newspaper design, reuses existing contents and provides means to personalize the selection of information which is presented. The result of this approach is a newspapers-like layout of information which is relevant for the specific user, presented in a pleasing way and potentially ready to be deployed on ePaper devices.

1. Introduction

The process of identifying relevant and discarding irrelevant information and duplicates in the current information universe is only one step in dealing with the “information overload”. The other step is the consumption which typically inherits the heterogeneous nature of the information sources, i.e. the user has to log in to the different systems and consume the information in the form the system defines. In contrast, traditional printed newspapers, magazines and journals provide relevant and well-researched information in a clean, readable and structured format, which has been exercised and refined for hundreds of years and provides superior consumability. However, to have every newspaper carefully prepared by a set of editors has at a price: every reader gets the same information no matter what her specific interests are, i.e.they do not adapt to the user at all.

The increasing amount of divergent and heterogeneous information published on the Internet and the consolidated and integrated style of newspapers are obviously the two extremes of the current situation. With the new ePaper devices, however, a third option emerges: personalized information presented in the traditional form of a newspaper.

In order to implement this option a number of tasks have to be solved, including the collection and consolidation of news from heterogeneous sources, the computation of relevance of any piece of information for the user, the adaptation of the information to resemble a newspaper and the delivery of the results to the user which could employ optimized clients in order to provide interactive elements, thus combining the presented information with exploration, retrieval and management mechanisms. Last but not least, the interaction of the user with a client can be used to obtain feedback which can be used to adapt the user’s preferences.

This paper presents a system which addresses all of the aforementioned tasks, of which the focus is the layout process, however. Chapter 2. presents the general overview of the system, detailing on the content preprocessing, preparation and delivery. Chapter 3. describes the theoretical foundation of the layout algorithm, while Chapter 4. describes the actual implementation and Chapter 5. presents the results obtained in the experiments we conducted. The last Chapter concludes the paper by summarizing the results and giving an outlook of further research.

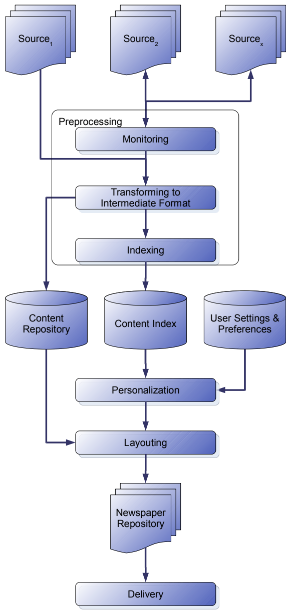


Figure 1: Abstract Architecture of the System

2. The Information Management System

A system which addresses all of the tasks listed in the introduction, must essentially consist of a number of dedicated sub-systems which are interacting to produce the desired results. Figure 1 provides an overview of the system which we designed for our experiments. For the sake of simplicity we will, however, only address in detail the sub-systems which are especially relevant to the tasks of information management and personalization, which are Preprocessing, Delivering and Personalized Retrieval.

2.1. Preprocessing

The preprocessing of contents consists of two steps. The first takes the incoming contents and transforms it into the format which is used for storage. The second step consists of updating the content repository and index with the new information.

Because the system deals with news contents, the input format of choice was NITF[1] an industry standard by the IPTC[2]. NITF is a XML-based format which provides meta-information about the news, e.g. subjects, urgency, keywords, revision history, etc, and the actual news item in a single document. It does not, however, contain information about the style of the news item. Therefore, anyone wishing to present the news to a user has the freedom to create their own presentation. Any content not compliant with this standard may be transformed to NITF with the proper tools.

The implemented preprocessing component separates actual contents from meta data by discarding everything that is inside the NITF header element, except for the identifier and the source information, and retaining only the headlines, paragraphs and media references of the body element. All other data and meta data is not needed in the layout process. However, some of the meta data is used in the index in order to improve the retrieval functions. News items are indexed using Apache Lucene[3], a high-performance full-text search engine. The obvious advantage of Lucene is its built-in extensible set of Natural Language Processing (NLP) functions, i.e. it is possible to perform language-dependent stop word elimination and stemming both on incoming news items and for subsequent queries by users.

2.2. Personalization

So far, the process described is the same for all users. The question therefore is, how the newspaper will be personalized. For this task three options are implemented. The first is the incorporation of interests; the second is the desired media ratio of the newspaper; the last option is the total number of pages to create.

The restriction on the total number of pages is important because not all articles may be placed on the resulting pages. If more pages are available articles have more opportunities to be placed even if the number of their layouts are limited. A small number of pages require more articles to be discarded, thus enforcing each article to be of high importance and, therefore, the need to ensure that highly relevant articles can be placed on a page in at least one way by providing a great number of layout options for each article.

The media ratio personalization allows the users to select a type of newspaper to generate. The implemented options are a choice between mostly textual contents, mostly media contents or a balanced version. The idea behind these options is the typical media-driven layout of the yellow press in contrast to serious newspapers such as *New York Times*, *Frankfurter Allgemeine Zeitung* and others, while the third option allows for a type of newspaper between the extremes.

A user's interest is expressed by setting preferences for specific topics, potentially associated with specific providers. In the implemented system, this consists of the possibility to select from the available information sources and the information topics they offer. For this purpose the IPTC subject codes [4] are used. They define a taxonomy of about 1.400 subjects in three levels. However, users may only select from the first level which contains 17 subjects. The second and third levels are reserved for the use in implicit topic selection where a system may refine the user-specified topics with details learned from the interaction of the user with the system, e.g. preferences concerning specific kinds of sports, cultural events or social matters.

Before the layout process starts a relevance score is computed for every active article which has been provided by or updates any of the user-selected news providers. Additional personal preferences, e.g. letting the user define the desired number of pages per category or the weighting of topics are not considered as part of the personal preferences for the layout process. This is no restriction, though, because these preferences can be simulated by adjusting the relevance of articles accordingly: articles belonging to topics more relevant to a user should receive a higher relevance score. Because one goal of the optimization process is to produce pages with high relevance, it will lead to pages containing more of the articles with boosted relevance.

2.3. Delivery and Interaction

In addition to merely displaying the generated newspaper, a client can be used to monitor the user to obtain explicit as well as implicit feedback which can be used to further refine the initially defined preferences: for each action the client allows the user to perform the attributes are stored and sent back to the system at regular intervals. Examples of such actions are:

- Marking an article
- Discarding an article
- Searching for a keyword
- Searching for articles related to a given article
- Requesting additional articles for a topic
- Requesting a detail view of an article
- Skipping a topic or page

Each of these actions is translated into an adaptation of the user's profile, e.g. marking an article results in an increase in the relevance of the article's topics and terms, while discarding an article will result in a decrease. In addition, the actions can be used to compute user-user similarities for collaborative filtering and recommendation/exploration. For the current system, however, this was out of scope.

The Figures 2, 3, 4, and 5 depict the client at various states of interaction with the user.

[1] News Industry Text Format, see `http://www.nitf.org/`

[2] International Press Telecommunications Council

[3] See `http://lucene.apache.org/`

[4] See `http://cv.iptc.org/newscodes/subjectcode/`

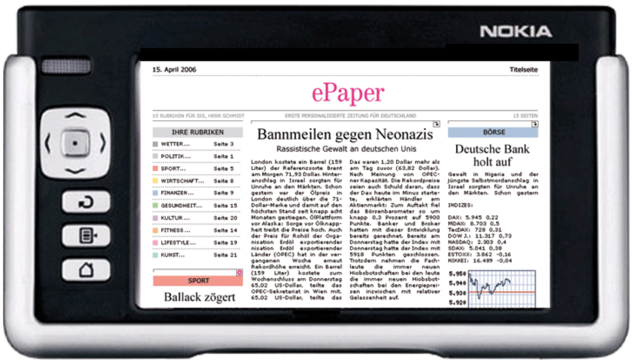


Figure 2: Main View on an Exemplary Device

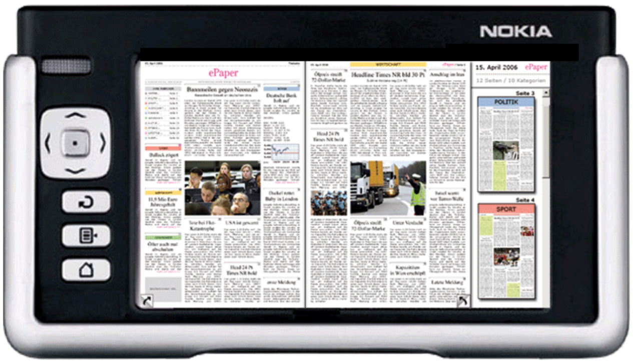
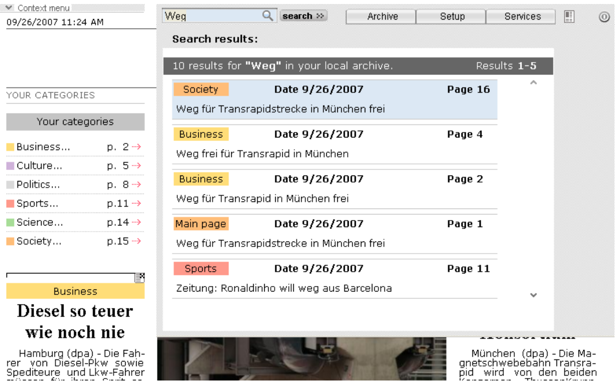


Figure 3: Alternative View for Navigation



09/26/2007 11:24 AM

YOUR CATEGORIES

Your categories

Business...	p. 2 →
Culture...	p. 5 →
Politics...	p. 8 →
Sports...	p.11 →
Science...	p.14 →
Society...	p.15 →

Business

Diesel so teuer wie noch nie

Hamburg (dpa) - Die Fahrer von Diesel-Pkw sowie Spediteure und Lkw-Fahrer

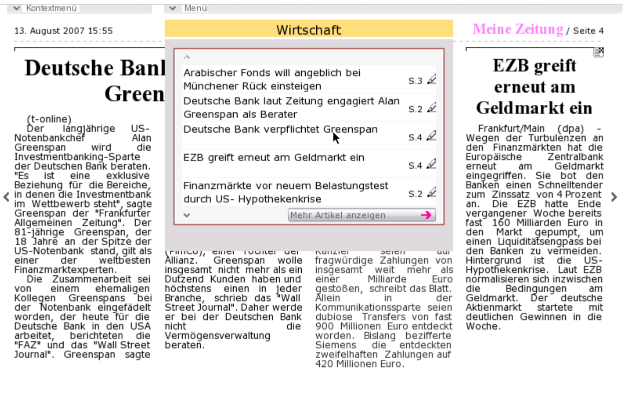
München (dpa) - Die Magnetschwebebahn Transrapid wird von den beiden

Search results:

10 results for **"Weg"** in your local archive. Results **1-5**

Society	**Date 9/26/2007**	**Page 16**
Weg für Transrapidstrecke in München frei		
Business	**Date 9/26/2007**	**Page 4**
Weg frei für Transrapid in München		
Business	**Date 9/26/2007**	**Page 2**
Weg für Transrapid in München frei		
Main page	**Date 9/26/2007**	**Page 1**
Weg für Transrapidstrecke in München frei		
Sports	**Date 9/26/2007**	**Page 11**
Zeitung: Ronaldinho will weg aus Barcelona		

Figure 4: Search for News



13. August 2007 15:55

Wirtschaft

Meine Zeitung / Seite 4

Arabischer Fonds will angeblich bei Münchener Rück einsteigen	S.3
Deutsche Bank laut Zeitung engagiert Alan Greenspan als Berater	S.2
Deutsche Bank verpflichtet Greenspan	S.4
EZB greift erneut am Geldmarkt ein	S.4
Finanzmärkte vor neuem Belastungstest durch US- Hypothekenkrise	S.2

Deutsche Banl Green

(t-online) Der langjährige US-Notenbankchef Alan Greenspan wird die Investmentbanking-Sparte der Deutschen Bank beraten. "Es ist eine exklusive Beziehung für die Bereiche, in denen die Investmentbank im Wettbewerb steht", sagte Greenspan der "Frankfurter Allgemeinen Zeitung". Der 81-jährige Greenspan, der 18 Jahre an der Spitze der US-Notenbank stand, gilt als einer der weltbesten Finanzmarktexperten.

Die Zusammenarbeit sei von einem ehemaligen Kollegen Greenspans bei der Notenbank eingefädelt worden, der heute für die Deutsche Bank in den USA arbeitet, berichteten die "FAZ" und das "Wall Street Journal". Greenspan sagte

(PimCo), einer Tochter der Allianz. Greenspan wolle insgesamt nicht mehr als ein Dutzend Kunden haben und höchstens einen in jeder Branche, schrieb das "Wall Street Journal". Daher werde er bei der Deutschen Bank nicht die Vermögensverwaltung beraten.

Kunzler seien auf fragwürdige Zahlungen von insgesamt weit mehr als einer Milliarde Euro gestoßen, schreibt das Blatt. Allein in der Kommunikationssparte seien dubiose Transfers von fast 900 Millionen Euro entdeckt worden. Bislang bezifferte Siemens die entdeckten zweifelhaften Zahlungen auf 420 Millionen Euro.

EZB greift erneut am Geldmarkt ein

Frankfurt/Main (dpa) - Wegen der Turbulenzen an den Finanzmärkten hat die Europäische Zentralbank erneut am Geldmarkt eingegriffen. Sie bot den Banken einen Schnelltender zum Zinssatz von 4 Prozent an. Die EZB hatte Ende vergangener Woche bereits fast 160 Milliarden Euro in den Markt gepumpt, um einen Liquiditätsengpass bei den Banken zu vermeiden. Hintergrund ist die US-Hypothekenkrise. Laut EZB normalisieren sich inzwischen die Bedingungen am Geldmarkt. Der deutsche Aktienmarkt startete mit deutlichen Gewinnen in die Woche.

Figure 5: Overview of Content for a Category

3. A Short Theory of Automatic Layout and Aesthetics

The goal of the proposed system is to provide a substitute for the layout process performed by the editors and the familiar reading experience of a newspaper. Therefore, the questions which must be addressed are ”What makes up a good newspaper?” and ”How can this be automated?”.

The history of aesthetics reaches as far back as Plato and Aristoteles, maybe even further. However, aesthetics were considered impossible to formalize and therefore typically taught/learned by giving rules which the designer was to follow, modern examples here are (Rehe, 1985) and (Collier, 1990). Until today, most editors use a mixture of intuition and style guides to achieve the unique look of their newspaper.

At the beginning of the last century, however, the first formal theories started to emerge, which claimed that the aesthetic quality of a design could be computed mathematically. The starting point was Birkhoff’s ”Aesthetic Measures” (Birkhoff, 1933) which was further refined by Bense (Bense, 1965a; Bense, 1965b; Bense, 1969), Greenfield (Greenfield, 2005) and others.

Most of these theories, however, started with some more or less intuitive notion of aesthetics, and therefore consistently failed to capture some of the important aspects. The change came towards the end of the last century when empirical user studies (Streveler and Wasserman, 1984; Sears, 1993; Ngo, 1994; Tullis, 1997; Ngo and Byrne, 1998) evaluated the influence of aesthetics and thus eliminated some of the subjectivity of the approaches, and established mathematical theories of interface aesthetics. Of these theories Ngo et al (Ngo et al., 2003) and Harrington et al (Harrington et al., 2004) can be considered as unifying the others and providing a comprehensive set of measures which serve to define the aesthetic quality of any interface.

Most of the measures proposed by the two papers have corresponding measures in the other, although they may have a different name or a slightly different computation. They can be roughly divided into two groups: measures which consider the basic optical aspects of the layout and measures which consider cognitive aspects. Examples of the first group are balance, equilibrium, density, white-space fraction and free-flow, which compute the optical density (or weight) of the elements on the page and their distribution. The second group, containing for example alignment, regularity, homogeneity, cohesion, proportion and economy, consists of measures which take into account aspects which deal with the overall perception and reading guidance.

In order to computer the quality of a layout, the different measures are computed and combined to a final score which is higher the better the layout is.

An alternative approach to achieve a high aesthetic quality consists of learning from samples of good (and bad) layouts. Buhr et al (Buhr, 1996) used Artificial Neural Networks to learn the optimal distribution of optical weight, but yielded only poor results, most likely because of the relative simplicity of the network used. Soon after, Bernard et al (Bernard and Lirsia, 1998) proposed an inductive constraint logic to learn rules which create visually pleasing layouts. While the claim is that their algorithms could learn a constraint which would place a given set of articles (of a given size) on the page, they give no examples of the resulting layouts, thus making the evaluation difficult. In general, approaches which learn what defines a good layout, require a lot of training samples. This reduces their applicability to cases where such training data is available, which is seldom the case.

Another alternative for generating high quality layouts is the use of templates. Each template describes ways of arranging the elements of a page to obtain an appealing layout and typically consists of interdependent constrains which generate the typical look. Jacobs et al (Jacobs et al., 2003; Jacobs et al., 2004) use a set of templates which describe adaptive page layouts of which the system then selects the best sequence, given a particular document. The goodness of a chosen sequence is measured in terms of constraint satisfaction, while the specific values for all parameters were computed with the help of constraint solvers. Similarly, Lin (Lin, 2006) proposes an algorithm which is restricted to single pages, but allows for the placement of multiple elements on the page. Schrier et al (Schrier et al., 2008) proposed an XML-based layout language which allows the recursive definition of templates, conditional elements and grouping of elements, which is based on the approach of Jacobs et al. Templates are chosen based on preconditions and scores and subsequently finalized with a constraint solver. The major drawback of all template-based approaches, however, is their need of an extensive set of templates which the system can choose from and a deep understanding of the constraint definition language and the constraint solving mechanisms in order to achieve high quality results.

Apart from the pure scientific applications of automatic newspaper layout, a set of real-world application has come to life in he recent years, e.g. Tabbloid, FeedJournal, PersonalNews and niiu (Tabbloid, 2009; Martinsson, 2009; Syntops GmbH, 2009; InterTi GmbH, 2010). In general, they achieve only unpleasing results because they are too inflexible or do not impose suitable constraints. The exception are the PersonalNews and niiu systems, which compose a personal newspaper of whole pages taken from original newspapers, thus avoiding the layout problem at the cost of not providing a consistent look across the complete newspaper.

4. The Layout Algorithm

Based on the observations made in the previous Section, we decided to implement the layout algorithm as filling a page with a selection of articles from a content base and optimizing the aesthetic quality. This can be further refined to the sub-tasks of coming up with possible arrangements of article contents and the consequent selection which of the arrangements of articles to choose and where to place them. The selection should reflect the user's preferences and the relevance of the articles for the user and the arrangement and placement should result in a newspaper-like look and feel while simultaneously maximizing the aesthetic quality. This description is very similar to the family of Cutting & Packing (C&P) problems which are addressed in the field of Operations Research (cf. (Wäscher et al., 2007; Kellerer et al., 2004). C&P problems arise in many practical situations, e.g. loading containers, cutting stock, scheduling tasks, assigning courses, or VLSI wiring, but are $\mathcal{NP}$-hard in general (Garey and Johnson, 1979). Exact solutions for these problems are impractical, therefore. Approximation algorithms try to find solutions close to the optimum at greatly improved computational complexity (Sahni, 1975; Beasley, 2004; Julstrom, 2005; Hiley and Julstrom, 2006; Saraç and Sipahioglu, 2007; Singh and Baghel, 2007).

In order to achieve the desired amount of flexibility in the layout and of aesthetic quality, the proposed algorithm for the layout combines two concepts which were identified during the research on traditional layout and aesthetics. The first, is the use of a style guide for defining valid variations for the layout of each article. The second concept is the employment of an objective function which combines the applicable aesthetic measures and a score which reflects the similarity to the personal preferences of a user. The task for an algorithm which automates the layout process then is to find an arrangement of articles which maximize the objective function of the generated layout. These concepts are described in the next sections.

4.1. Style Guide

The style guide which was developed, is based on the concept of a grid as the underlying structure of each page, i.e. each page is into w columns and h rows resulting in $w * h$ identical cells (cf. Figure 6). Any article on the page may occupy a rectangle of an arbitrary but integer number of cells. In addition, rules for text and page margins, the style of article headlines and the size and placement of media elements were defined. This very simple set of rules automatically optimizes some of the aesthetic measures and provides consistency among the laid out articles. How this is achieved, is shown in reference to the aesthetic measures of Ngo and Harrington (Ngo et al., 2003; Harrington et al., 2004). Cohesion, unity, simplicity and regularity are measures based on the number of alignment points and available forms. With the grid structure the number of different aspect ratios is effectively limited and the aspect ratios themselves are related because every dimension is a multiple of the basic underlying grid dimension, thus supporting the cohesion of the appearance. In addition, the grid reduces the alignment points between the articles, while the alignment of text lines in neighboring cells further minimizes the alignment points, thus achieving a high simplicity. Regularity, i.e. the similarity of the distances between alignment points, is also achieved by the aforementioned grid and text alignments.

Because of the gap between columns and the margins introduced at the top (above the headline) and bottom (below the last text lines or images), which separate the article contents from the rest, related information is grouped. This corresponds to the unity of the layout.

The number of design elements which are produced by the application of the style guide, is rather limited: only a single font is used for the text and the headlines differ only in the size of the font; all media elements are rectangular. Colors other than the default text color are used only

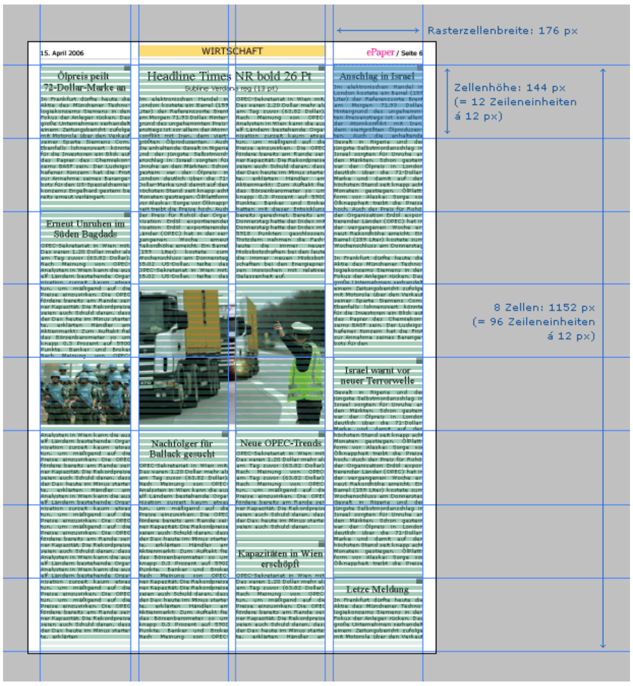

Rasterzellenbreite: 176 px
15. April 2006
WIRTSCHAFT
ePaper / Seite 6
Ölpreis peilt 72-Dollar-Marke an
Headline Times NR bold 26 Pt
Subline Verdana reg (13 pt)
Anschlag in Israel
Zellenhöhe: 144 px (= 12 Zeileneinheiten á 12 px)
Erneut Unruhen im Süden Bagdads
8 Zellen: 1152 px (= 96 Zeileneinheiten á 12 px)
Israel warnt vor neuer Terrorwelle
Nachfolger für Ballack gesucht
Neue OPEC-Trends
Kapazitäten in Wien erschöpft
Letze Meldung

Figure 6: The basic grid with margins and sample contents

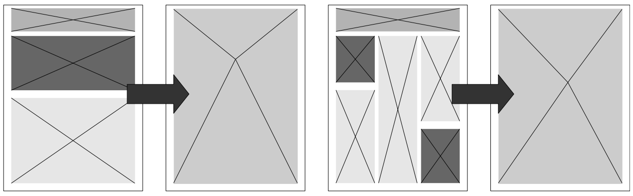


Figure 7: Computation of optical weight for an article from its parts. The level of gray denotes the optical weight.

to highlight important information and placed outside the main text areas. Therefore, the resulting layouts are very economic in their use of colors, fonts and forms.

The measure of rhythm is rather hard to grasp. The structure of the grid and the contained articles, which consist of a headline, text and optional images, introduce a systematic change in the arrangement of elements on the page. This can be considered as the style guide's contribution to rhythm.

4.2. Objective Function

The objective function must be designed to capture both the aesthetic quality of the generated layout and the closeness of the selected content and layout to the user's preferences. The goodness of the content selection can be measured by the relevance score of each article (cf. Section 2.2.). The layout's quality and relation to the user defined media aspect ratio, requires the computation of a measure closely related to the formal theories of aesthetic: the optical density (M_c) and center (x_c, y_c). For each generated variation of an article layout, is is computed as a weighed sum of the parts (cf. Figure 7):

$$x_c = \frac{\sum_i x_i * M_i * a_i}{\sum_i M_i * a_i}$$

$$y_c = \frac{\sum_i y_i * M_i * a_i}{\sum_i M_i * a_i}$$

$$M_c = \frac{\sum_i M_i * a_i}{a_i}$$

where i iterates over all elements of the article (headline, text parts and, if available, images) and x_i, y_i, M_i and a_i the optical center and weight, and the area of the part, respectively. For each part the center is assumed to be identical to the center of the rectangular area it occupies, and the density is fixed and computed as $M_i = 1.0 - brightness$[5]. Based on these properties for each article layout, the different aesthetic measures (cf. Section 3.) can be computed, as well as the overall relevance of a page for a user:

- Relevance, i.e. topical and term-based correlation with user preferences (r_i: relevance of article i, a_i area of article i): $RM = \sum_i r_i * a_i$
- Balance ((x_c, y_c): desired center of gravity, $x_m = \frac{\sum_i x_i * M_i}{\sum_i M_i}, y_m = \frac{\sum_i y_i * M_i}{\sum_i M_i}$: optical center of the page): $BM = 1 - \sqrt{\frac{(\frac{x_m - x_c}{W})^2 + (\frac{y_m - y_c}{H})^2}{2}}$
- Sequence, i.e. "Relevance Order" ($x_r = \frac{\sum_i x_i * r_i}{\sum_i r_i}, y_r = \frac{\sum_i y_i * r_i}{\sum_i r_i}$: "relevance center" of the page): $SM = 1 - \sqrt{\frac{(\frac{x_r}{W})^2 + (\frac{y_r}{H})^2}{2}}$
- Density: $DM = \frac{D_{target} - \sum_i a_i * M_i}{D_{scale}}$ where $D_{scale} = max(M_i) - min(M_i)$
- Fullness, i.e. minimization of Whitespace: $WM = \frac{\sum_i a_i * c_i}{W * H}$
- Page score (weighted sum of scores): $f = \frac{1}{\alpha + \beta + \gamma + \delta + \epsilon} * (\alpha * RM + \beta * BM + \gamma * SM + \delta * DM + \epsilon * WM)$

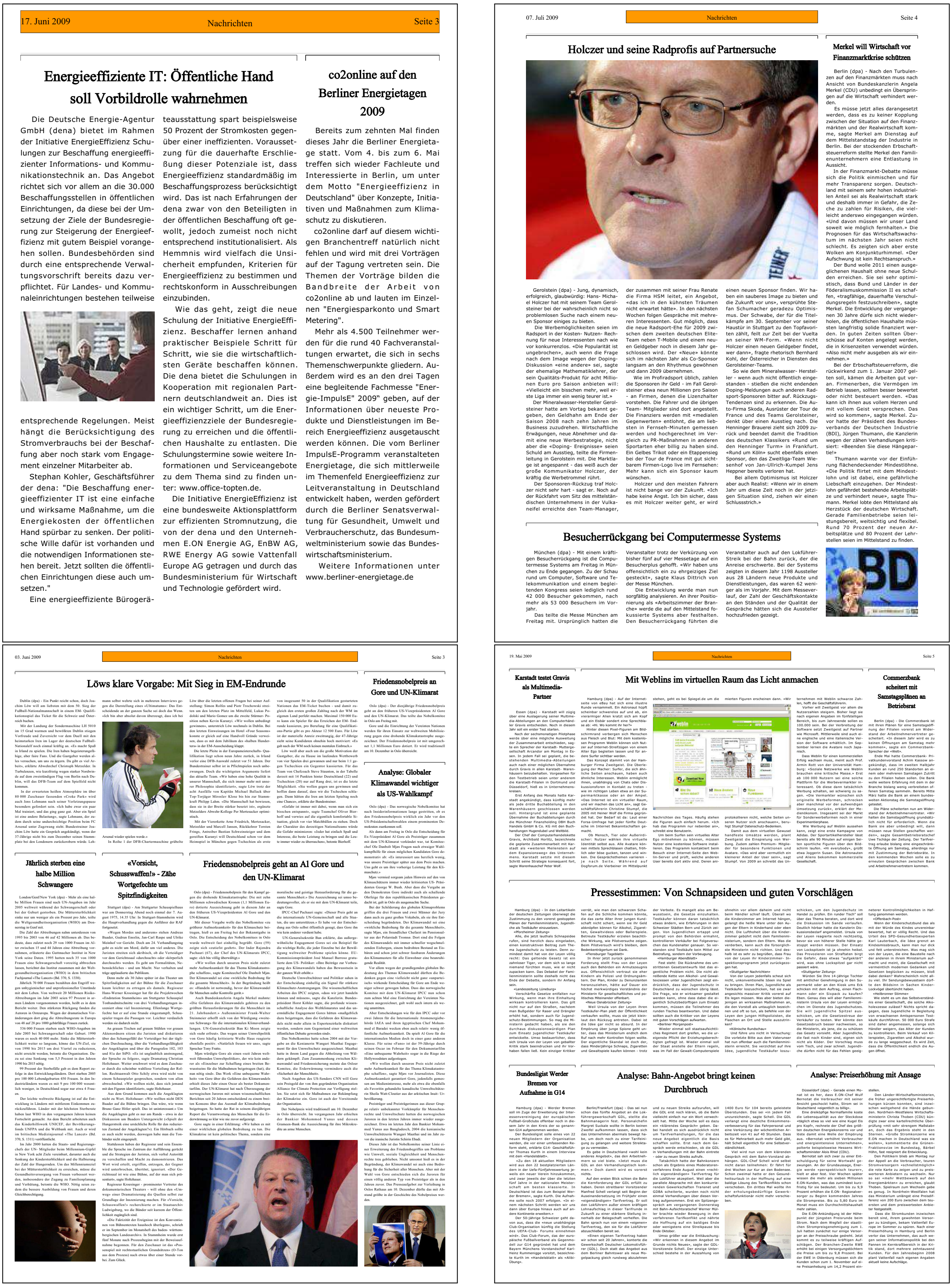
Because the system collects news contents from multiple sources, the issue of redundant news must be addressed, too. The objective function can be extended to contain a term which imposes a penalty for the redundancy of the included items. This can be done by computing the cosine similarity of the content (as a term frequency vector) of any pair of items and including the measure into the objective function:

- Diversity (sim_{ij}: similarity of article i and j, n: number of articles on page): $DivM = 1 - \frac{\sum_{i \neq j} sim_{ij}}{n^2}$

5. Experimental Results

Based on the objective function presented in Section 4.2., different optimization algorithms were used. As Strecker et al showed in (Strecker and Hennig, 2009), relatively simple approximation algorithms could be used for simple objective functions. Using these algorithms in combination with the present objective function yielded only poor results, which we mainly attribute to the huge problem space

[5] The brightness of an element is computed by converting its content from RGB to HSB color space and averaging the brightness values.



17. Juni 2009 Nachrichten Seite 3

Energieeffiziente IT: Öffentliche Hand soll Vorbildrolle wahrnehmen

Die Deutsche Energie-Agentur GmbH (dena) bietet im Rahmen der Initiative EnergieEffizienz Schulungen zur Beschaffung energieeffizienter Informations- und Kommunikationstechnik an. Das Angebot richtet sich vor allem an die 30.000 Beschaffungsstellen in öffentlichen Einrichtungen, da diese bei der Umsetzung der Ziele der Bundesregierung zur Steigerung der Energieeffizienz mit gutem Beispiel vorangehen sollen. Bundesbehörden sind durch eine entsprechende Verwaltungsvorschrift bereits dazu verpflichtet. Für Landes- und Kommunaleinrichtungen bestehen teilweise entsprechende Regelungen. Meist hängt die Berücksichtigung des Stromverbrauchs bei der Beschaffung aber noch stark vom Engagement einzelner Mitarbeiter ab.

Stephan Kohler, Geschäftsführer der dena: "Die Beschaffung energieeffizienter IT ist eine einfache und wirksame Maßnahme, um die Energiekosten der öffentlichen Hand spürbar zu senken. Der politische Wille dafür ist vorhanden und die notwendigen Informationen stehen bereit. Jetzt sollten die öffentlichen Einrichtungen diese auch umsetzen."

Eine energieeffiziente Bürogeräteausstattung spart beispielsweise 50 Prozent der Stromkosten gegenüber einer ineffizienten. Voraussetzung für die dauerhafte Erschließung dieser Potenziale ist, dass Energieeffizienz standardmäßig im Beschaffungsprozess berücksichtigt wird. Das ist nach Erfahrungen der dena zwar von den Beteiligten in der öffentlichen Beschaffung oft gewollt, jedoch zumeist noch nicht entsprechend institutionalisiert. Als Hemmnis wird vielfach die Unsicherheit empfunden, Kriterien für Energieeffizienz zu bestimmen und rechtskonform in Ausschreibungen einzubinden.

Wie das geht, zeigt die neue Schulung der Initiative EnergieEffizienz. Beschaffer lernen anhand praktischer Beispiele Schritt für Schritt, wie sie die wirtschaftlichsten Geräte beschaffen können. Die dena bietet die Schulungen in Kooperation mit regionalen Partnern deutschlandweit an. Dies ist ein wichtiger Schritt, um die Energieeffizienzziele der Bundesregierung zu erreichen und die öffentlichen Haushalte zu entlasten. Die Schulungstermine sowie weitere Informationen und Serviceangebote zu dem Thema sind zu finden unter: www.office-topten.de.

Die Initiative EnergieEffizienz ist eine bundesweite Aktionsplattform zur effizienten Stromnutzung, die von der dena und den Unternehmen E.ON Energie AG, EnBW AG, RWE Energy AG sowie Vattenfall Europe AG getragen und durch das Bundesministerium für Wirtschaft und Technologie gefördert wird.

co2online auf den Berliner Energietagen 2009

Bereits zum zehnten Mal finden dieses Jahr die Berliner Energietage statt. Vom 4. bis zum 6. Mai treffen sich wieder Fachleute und Interessierte in Berlin, um unter dem Motto "Energieeffizienz in Deutschland" über Konzepte, Initiativen und Maßnahmen zum Klimaschutz zu diskutieren.

co2online darf auf diesem wichtigen Branchentreff natürlich nicht fehlen und wird mit drei Vorträgen auf der Tagung vertreten sein. Die Themen der Vorträge bilden die Bandbreite der Arbeit von co2online ab und lauten im Einzelnen "Energiesparkonto und Smart Metering".

Mehr als 4.500 Teilnehmer werden für die rund 40 Fachveranstaltungen erwartet, die sich in sechs Themenschwerpunkte gliedern. Außerdem wird es an den drei Tagen eine begleitende Fachmesse "Energie-ImpulsE" 2009" geben, auf der Informationen über neueste Produkte und Dienstleistungen im Bereich Energieeffizienz ausgetauscht werden können. Die vom Berliner ImpulsE-Programm veranstalteten Energietage, die sich mittlerweile im Themenfeld Energieeffizienz zur Leitveranstaltung in Deutschland entwickelt haben, werden gefördert durch die Berliner Senatsverwaltung für Gesundheit, Umwelt und Verbraucherschutz, das Bundesumweltministerium sowie das Bundeswirtschaftsministerium.

Weitere Informationen unter www.berliner-energietage.de

07. Juli 2009 Nachrichten Seite 4

Holczer und seine Radprofis auf Partnersuche

Gerolstein (dpa) - Jung, dynamisch, erfolgreich, glaubwürdig: Hans- Michael Holczer hat mit seinem Team Gerolsteiner bei der wahrscheinlich nicht so problemlosen Suche nach einem neuen Sponsor einiges zu bieten.

Die Werbemöglichkeiten seien im Radsport in der Kosten- Nutzen- Rechnung für neue Interessenten nach wie vor konkurrenzlos. «Die Popularität ist ungebrochen», auch wenn die Frage nach dem Image wegen der Doping- Diskussion «eine andere» sei, sagte der ehemalige Mathematiklehrer, der sein Qualitäts-Produkt für acht Millionen Euro pro Saison anbieten will: «Vielleicht ein bisschen mehr, weil erste Liga immer ein wenig teurer ist.»

Der Mineralwasser-Hersteller Gerolsteiner hatte am Vortag bekannt gegeben, den Geldhahn am Ende der Saison 2008 nach zehn Jahren im Business zuzudrehen. Wirtschaftliche Erwägungen, neue Abnehmer und damit eine neue Werbestrategie, nicht aber die «Doping- Ereignisse» seien Schuld am Ausstieg, teilte die Firmenleitung in Gerolstein mit. Die Marktlage ist angespannt - das weiß auch der große Kommunikator Holczer, der kräftig die Werbetrommel rührt.

Der Sponsoren-Rückzug traf Holczer nicht sehr hart - sagt er. Noch auf der Rückfahrt vom Sitz des mittelständischen Unternehmens in der Vulkaneifel erreichte den Team-Manager, der zusammen mit seiner Frau Renate die Firma HSM leitet, ein Angebot, «das ich in den kühnsten Träumen nicht erwartet hätte». In den nächsten Wochen folgen Gespräche mit mehreren Interessenten. Gut möglich, dass die neue Radsport-Ehe für 2009 zwischen dem zweiten deutschen Elite-Team neben T-Mobile und einem neuen Geldgeber noch in diesem Jahr geschlossen wird. Der «Neue» könnte sich im nächsten Jahr als Co-Sponsor langsam an den Rhythmus gewöhnen und dann 2009 übernehmen.

Wie im Profiradsport üblich, zahlen die Sponsoren ihr Geld - im Fall Gerolsteiner etwa neun Millionen pro Saison - an Firmen, denen die Lizenzhalter vorstehen. Die Fahrer und die übrigen Team- Mitglieder sind dort angestellt. Die Finanziers werden mit «medialen Gegenwerten» entlohnt, die am liebsten in Fernseh-Minuten gemessen werden und hochgerechnet im Vergleich zu PR-Maßnahmen in anderen Sportarten eher billig zu haben sind. Ein Gelbes Trikot oder ein Etappensieg bei der Tour de France mit gut sichtbarem Firmen-Logo live im Fernsehen: Mehr kann sich ein Sponsor kaum wünschen.

Holczer und den meisten Fahrern ist nicht bange vor der Zukunft. «Ich habe keine Angst. Ich bin sicher, dass es mit Holczer weiter geht, er wird einen neuen Sponsor finden. Wir haben ein sauberes Image zu bieten und die Zukunft vor uns», versprühte Stefan Schumacher geradezu Optimismus. Der Schwabe, der für die Titelkämpfe am 30. September vor seiner Haustür in Stuttgart zu den Topfavoriten zählt, feilt zur Zeit bei der Vuelta an seiner WM-Form. «Wenn nicht Holczer einen neuen Geldgeber findet, wer dann», fragte rhetorisch Bernhard Kohl, der Österreicher in Diensten des Gerolsteiner-Teams.

So wie dem Mineralwasser- Hersteller - wenn auch nicht öffentlich eingestanden - stießen die nicht endenden Doping-Meldungen auch anderen Radsport-Sponsoren bitter auf. Rückzugs-Tendenzen sind zu erkennen. Die Auto-Firma Skoda, Ausrüster der Tour de France und des Teams Gerolsteiner, denkt über einen Ausstieg nach. Die Henninger Brauerei zieht sich 2009 zurück und beendet damit die Tradition des deutschen Klassikers «Rund um den Henninger Turm» in Frankfurt. «Rund um Köln» sucht ebenfalls einen Sponsor, den das Zweitliga-Team Wiesenhof von Jan-Ullrich-Kumpel Jens Heppner bereits verloren hat.

Bei allem Optimismus ist Holczer aber auch Realist: «Wenn wir in einem Jahr um diese Zeit noch in der jetzigen Situation sind, ziehen wir einen Schlussstrich.»

Merkel will Wirtschaft vor Finanzmarktkrise schützen

Berlin (dpa) - Nach den Turbulenzen auf den Finanzmärkten muss nach Ansicht von Bundeskanzlerin Angela Merkel (CDU) unbedingt ein Überspringen auf die Wirtschaft verhindert werden.

Es müsse jetzt alles darangesetzt werden, dass es zu keiner Kopplung zwischen der Situation auf den Finanzmärkten und der Realwirtschaft komme, sagte Merkel am Dienstag auf dem Mittelstandstag der Industrie in Berlin. Bei der stockenden Erbschaftsteuerreform stellte Merkel den Familienunternehmern eine Entlastung in Aussicht.

In der Finanzmarkt-Debatte müsse sich die Politik einmischen und für mehr Transparenz sorgen. Deutschland mit seinem sehr hohen industriellen Anteil sei als Realwirtschaft stark und deshalb immer in Gefahr, die Zeche zu zahlen für Risiken, die vielleicht anderswo eingegangen würden. «Und davon müssen wir unser Land soweit wie möglich fernhalten.» Die Prognosen für das Wirtschaftswachstum im nächsten Jahr seien nicht schlecht. Es zeigten sich aber erste Wolken am Konjunkturhimmel. «Der Aufschwung ist kein Rechtsanspruch.»

Der Bund wolle 2011 einen ausgeglichenen Haushalt ohne neue Schulden erreichen. Sie sei sehr optimistisch, dass Bund und Länder in der Föderalismuskommission II es schaffen, «tragfähige, dauerhafte Verschuldungsregeln festzuschreiben», sagte Merkel. Die Entwicklung der vergangenen 30 Jahre dürfe sich nicht wiederholen, die öffentlichen Haushalte müssten langfristig solide finanziert werden. In guten Zeiten sollten Überschüsse auf Konten angelegt werden, die in Krisenzeiten verwendet würden. «Also nicht mehr ausgeben als wir einnehmen.»

Bei der Erbschaftsteuerreform, die rückwirkend zum 1. Januar 2007 gelten soll, kämen die Arbeiten gut voran. Firmenerben, die Vermögen im Betrieb lassen, sollten besser bewertet oder nicht besteuert werden. «Das kann ich ihnen aus vollem Herzen und mit vollem Geist versprechen. Das wird so kommen», sagte Merkel. Zuvor hatte der Präsident des Bundesverbands der Deutschen Industrie (BDI), Jürgen Thumann, die Kanzlerin wegen der zähen Verhandlungen kritisiert: «Beenden Sie diese Hängepartie!»

Thumann warnte vor der Einführung flächendeckender Mindestlöhne. «Die Politik flirtet mit dem Mindestlohn und ist dabei, eine gefährliche Liebschaft einzugehen. Der Mindestlohn gefährdet bestehende Arbeitsplätze und verhindert neue», sagte Thumann. Merkel lobte den Mittelstand als Herzstück der deutschen Wirtschaft. Gerade Familienbetriebe seien leistungsbereit, weitsichtig und flexibel. Rund 70 Prozent der neuen Arbeitsplätze und 80 Prozent der Lehrstellen seien im Mittelstand zu finden.

Besucherrückgang bei Computermesse Systems

München (dpa) - Mit einem kräftigen Besucherrückgang ist die Computermesse Systems am Freitag in München zu Ende gegangen. Zu der Schau rund um Computer, Software und Telekommunikation und einem begleitenden Kongress seien lediglich rund 42 000 Besucher gekommen, nach mehr als 53 000 Besuchern im Vorjahr.

Das teilte die Messe München am Freitag mit. Ursprünglich hatten die Veranstalter trotz der Verkürzung von bisher fünf auf vier Messetage auf ein Besucherplus gehofft. «Wir haben uns offensichtlich ein zu ehrgeiziges Ziel gesteckt», sagte Klaus Dittrich von der Messe München.

Die Entwicklung werde man nun sorgfältig analysieren. An ihrer Positionierung als «Arbeitszimmer der Branche» werde die auf den Mittelstand fokussierte Systems aber festhalten. Den Besucherrückgang führten die Veranstalter auch auf den Lokführer-Streik bei der Bahn zurück, der die Anreise erschwerte. Bei der Systems zeigten in diesem Jahr 1198 Aussteller aus 28 Ländern neue Produkte und Dienstleistungen, das waren 62 weniger als im Vorjahr. Mit dem Messeverlauf, der Zahl der Geschäftskontakte an den Ständen und der Qualität der Gespräche hätten sich die Aussteller hochzufrieden gezeigt.

03. Juni 2009 Nachrichten Seite 3

Löws klare Vorgabe: Mit Sieg in EM-Endrunde

Dublin (dpa) - Ein Punkt reicht schon, doch Joachim Löw will am liebsten mit dem 50. Sieg der Fußball-Nationalmannschaft in einem EM- Qualifikationsspiel das Ticket für die Schweiz und Österreich buchen.

Mit der Landung der Sondermaschine LH 5010 im 15 Grad warmen und bewölkten Dublin stiegen Vorfreude und Zuversicht vor dem Duell mit den heimstarken Iren im Lager der deutschen Fußball-Nationalelf noch einmal kräftig an. «Es macht Spaß in Irland zu spielen. Die Iren haben begeisterungsfähige, aber faire Fans. Und die Gastgeber werden alles versuchen, um uns zu ärgern. Da gibt es viel Arbeit», erklärte Abwehrchef Christoph Metzelder. In Turbulenzen, wie kurzfristig wegen starker Nordwinde auf dem zweistündigen Flug von Berlin nach Dublin, will das DFB-Team auf dem Spielfeld nicht kommen.

In der erwarteten heißen Atmosphäre im über 80 000 Zuschauer fassenden «Croke Park» wird auch Jens Lehmann nach seiner Verletzungspause besonders gefordert sein. «Ich habe zwar ein paar Mal trainiert, und das ging ganz gut. Aber ein Spiel ist eine andere Belastung», sagte Lehmann, der zudem durch seine undurchsichtige Position beim FC Arsenal unter Zugzwang steht. Bundestrainer Joachim Löw hatte ein Gespräch angekündigt, wenn der 37-Jährige nicht bis zum Dezember seinen Stammplatz bei den Londonern zurückerobern würde. Lehmann selbst wehrte sich in mehreren Interviews gegen die Darstellung eines «Ultimatums»: Das Entscheidende an der ganzen Sache sei doch das Wenn. «Ich bin aber absolut davon überzeugt, dass ich bei Arsenal wieder spielen werde.»

In Reihe 1 der DFB-Chartermaschine grübelte Löw über die letzten offenen Fragen bei seiner Aufstellung: Simon Rolfes und Piotr Trochowski streiten um den letzten Platz im Mittelfeld, Lukas Podolski und Mario Gomez um die zweite Stürmer- Position neben Kevin Kuranyi. «Wir wollen unbedingt gewinnen», unterstrich Löw nochmals in Dublin. Bei den letzten Einweisungen im Hotel «Four Seasons» konnte er gleich auf eine Handvoll Gründe verweisen, dass es mit dem Jubiläum des «halben Hunderters» in der EM-Ausscheidung klappt.

Die letzte Pleite in der Europameisterschafts- Qualifikation liegt schon neun Jahre zurück, in Irland verlor eine DFB-Auswahl zuletzt vor 51 Jahren. Der Bundestrainer selbst ist in Pflichtspielen noch unbezwungen. Doch die wichtigsten Argumente liefert das aktuelle Team. «Wir haben eine hohe Qualität in unserer Mannschaft, die sich immer mehr mit unserer Philosophie identifiziert», sagte Löw trotz der acht Ausfälle von Kapitän Michael Ballack über Top-Torjäger Miroslav Klose bis hin zur Stammkraft Philipp Lahm. «Die Mannschaft hat bewiesen, dass sie in der Breite stärker besetzt ist», ergänzte Metzelders Abwehr-Kollege Per Mertesacker optimistisch.

Mit der Viererkette Arne Friedrich, Mertesacker, Metzelder und Marcell Jansen, Rückkehrer Torsten Frings, Antreiber Bastian Schweinsteiger und dem gereiften Kuranyi will Deutschland schon vor dem Heimspiel in München gegen Tschechien als erste von insgesamt 50 in der Qualifikation gestarteten Nationen das EM-Ticket buchen - und damit zugleich den ersten großen Zahltag nach der WM im eigenen Land perfekt machen. Maximal 150 000 Euro kann ein Spieler für das Erreichen der EM- Endrunde kassieren, pro Berufung für eine Qualifikations-Partie gibt es pro Akteur 12 500 Euro. Für Löw ist der materielle Anreiz zweitrangig, der 47-Jährige sieht seine Kandidaten ohnehin hoch motiviert: «Es gab nach der WM noch keinen mentalen Einbruch.»

Löw weiß aber auch um die große Motivation der Gastgeber, die zu Hause im laufenden Wettbewerb von vier Spielen drei gewannen und nur beim 1:1 gegen Tschechien ein Gegentor kassierten. Für das Team von Chefcoach Steve Staunton, in der Tabelle derzeit mit 14 Punkten hinter Deutschland (22) und Tschechien (20) nur auf Rang drei, ist es die letzte Möglichkeit. «Sie wollen gegen uns gewinnen und hoffen dann darauf, dass wir die Tschechen schlagen. Dann hätten die Iren am letzten Spieltag noch eine Chance», erklärte der Bundestrainer.

«Gefahr ist immer mit dabei, wenn man sich ein bisschen entspannt», sagte Teamchef Oliver Bierhoff und verwies auf die eigentlich komfortable Situation, gleich vor vier Matchbällen zu stehen. Doch der Charakter des derzeitigen Nationalteams würde die Gefahr minimieren: «Jeder hat einfach Spaß und Interesse, die beste Leistung zu bringen und die Leute immer wieder zu überraschen», betonte Bierhoff.

Jährlich sterben eine halbe Million Schwangere

London/Genf/New York (dpa) - Mehr als eine halbe Million Frauen sind nach UN-Angaben im Jahr 2005 weltweit während der Schwangerschaft oder bei der Geburt gestorben. Die Müttersterblichkeit sinke nur um weniger als ein Prozent pro Jahr, teilte die Weltgesundheitsorganisation (WHO) am Donnerstag in Genf mit.

Die Zahl der Abtreibungen nahm unterdessen von 1995 bis 2003 von 46 auf 42 Millionen ab. Das bedeute, dass zuletzt noch 29 von 1000 Frauen im Alter zwischen 15 und 44 Jahren eine Abtreibung vornahmen, erläuterte das Guttmacher Institut in New York seine Daten. 1995 hatten noch 35 von 1000 Frauen eine Schwangerschaft vorzeitig abbrechen lassen, berichtet das Institut zusammen mit der Weltgesundheitsorganisation (WHO) in dem britischen Fachjournal «The Lancet» (Bd. 370, S. 1338).

Jährlich 70 000 Frauen bezahlten den Eingriff wegen unhygienischer und unprofessioneller Umstände mit dem Leben. Von weltweit 20 Millionen Risiko-Abtreibungen im Jahr 2003 seien 97 Prozent in armen Ländern vorgenommen worden, heißt es in dem Bericht weiter. Den stärksten Rückgang fanden die Autoren in Osteuropa. Wegen der dramatischen Veränderungen dort ging die Abtreibungsrate in Europa von 48 auf 28 pro 1000 gebärfähige Frauen zurück.

536 000 Frauen starben nach WHO-Angaben im Jahr 2005 bei Schwangerschaft oder Geburt. 1990 waren es noch 40 000 mehr. Sinke die Müttersterblichkeit weiter so langsam, könne das UN-Ziel, sie von 1990 bis 2015 um drei Viertel zu reduzieren, nicht erreicht werden, betonte die Organisation. Dazu sei eine Senkung von 5,5 Prozent in den Jahren 1990 bis 2015 nötig.

99 Prozent der Sterbefälle gab es dem Report zufolge in den Entwicklungsländern. Dort starben 2005 pro 100 000 Lebendgeburten 450 Frauen. In den Industrieländern waren es mit 9 pro 100 000 wesentlich weniger, in Deutschland sogar nur etwa 4 Frauen.

Der leichte weltweite Rückgang ist auf die Entwicklung in Ländern mit mittlerem Einkommen zurückzuführen. Länder mit der höchsten Sterberate haben laut WHO in den vergangenen Jahren keinen Fortschritt gemacht. An dem Bericht arbeiteten auch das Kinderhilfswerk UNICEF, der Bevölkerungsfonds UNFPA und die Weltbank mit. Auch er wird im britischen Medizinjournal «The Lancet» (Bd. 370, S. 1311) veröffentlicht.

Im Jahr 2000 hatten die Staats- und Regierungschefs der UN- Mitglieder beim Millennium-Gipfel in New York acht Ziele vereinbart, darunter auch die Senkung der Kindersterblichkeit und die Halbierung der Zahl der Hungernden. Um das Millenniumsziel bei der Müttersterblichkeit zu erreichen, müsse die Gesundheitsversorgung von Frauen verbessert werden, insbesondere der Zugang zu Familienplanung und Verhütung, betonte die WHO. Nötig seien zudem die bessere Ausbildung von Frauen und deren Gleichberechtigung.

«Vorsicht, Schusswaffen!» - Zähe Wortgefechte um Spitzfindigkeiten

Stuttgart (dpa) - Am Stuttgarter Schauspielhaus war am Donnerstag Abend noch einmal der 7. August 1975, 14.35 Uhr: In Stuttgart-Stammheim wird die Hauptverhandlung gegen die Anführer der RAF fortgesetzt.

«Wegen Mordes und anderem» stehen Andreas Baader, Gudrun Ensslin, Jan-Carl Raspe und Ulrike Meinhof vor Gericht. Doch am 24. Verhandlungstag geht es nicht um Mord, dafür um viel anderes. Die Juristen streiten sich akribisch, ob die Verteidiger vor dem Gerichtssaal «durchsucht» oder «körperlich durchsucht» werden. Es geht um Formalitäten, Nebensächliches - und um Macht. Nur verhalten und träge applaudierte das Publikum.

Denn mehr als 30 Jahre später ist das Theater um Spitzfindigkeiten auf der Bühne für die Zuschauer kaum leichter zu ertragen als damals. Regisseur Hans-Werner Kroesinger hat für die Projektwochen «Endstation Stammheim» am Stuttgarter Schauspiel Tonbandmitschnitte von drei Verhandlungstagen inszeniert ohne viel Inszenierung. Ellenlange Wortgefechte hat er auf eine Stunde eingestampft, Schauspieler tragen die Passagen vor. Leichter verdaulich werden sie dadurch nicht.

An grauen Tischen auf grauen Stühlen vor grauen Aktenordnern sitzen die Juristen und diskutieren über das Schamgefühl der Verteidiger bei der täglichen Durchsuchung, über die Verhandlungsfähigkeit der Angeklagten und über die Paragrafen 102, 103 und 81a der StPO. «Es ist unglaublich anstrengend, der Sprache zu folgen», sagte Dramaturg Christian Holtzhauer. Weiter erschwert wird es dem Zuschauer durch die scheinbar wahllose Verteilung der Rollen. Rechtsanwalt Otto Schily etwa wird nicht von einem Schauspieler gesprochen, sondern von allen abwechselnd. «Wir wollten nicht, dass sich jemand mit den Figuren identifiziert», sagte Holtzhauer.

Aus dem Grund kommen auch die Angeklagten nicht zu Wort. Holtzhauer: «Wir wollten nicht DEN Baader auf die Bühne bringen. Das wäre, wie wenn Bruno Ganz Hitler spielt. Das ist uninteressant.» Um die Angeklagten geht es nur am Rande - etwa in der Diskussion um Baaders Untergewicht («Spielt der Hungerstreik eine ursächliche Rolle für den reduzierten Zustand der Angeklagten?»). Ein Hörbuch sollte es auch nicht werden, deswegen habe man die Tonbänder nicht eingespielt.

Stattdessen haben der Regisseur und sein Ensemble die Sprache ins Zentrum der Aufführung gerückt und die Strategien der Juristen, sich verbal Autorität zu verschaffen und Macht zu demonstrieren. Das Wort wird erteilt, ergriffen, entzogen, der Gegner wird unterbrochen, übertönt, ignoriert. «Der Gerichtssaal ist wie eine Bühne, auf der man sich präsentiert», sagte Holtzhauer.

Regisseur Kroesinger - prominenter Vertreter des dokumentarischen Theaters - will ohne den «Umweg» einer Dramatisierung die Quellen selbst zur Grundlage der Inszenierung machen. Für «Vorsicht, Schusswaffen!» recherchierte er im Staatsarchiv Ludwigsburg, wo die Bänder seit kurzem der Öffentlichkeit zugänglich sind.

«Die Faktizität der Ereignisse ist den Konventionen von Bühnentexten haushoch überlegen», schrieb er im September im Monatsheft des baden- württembergischen Landesarchivs. In Stammheim wurde erst fünf Monate nach Prozessbeginn mit der Beweisaufnahme begonnen. Für den Zuschauer ist das «Possenspiel mit rechtsstaatlichen Grundsätzen» (O-Ton aus dem Prozess) nach etwas über einer Stunde vorbei. Zum Glück.

Friedensnobelpreis geht an Al Gore und den UN-Klimarat

Oslo (dpa) - Friedensnobelpreis für den Kampf gegen die drohende Klimakatastrophe: Die mit zehn Millionen schwedischen Kronen (1,1 Millionen Euro) dotierte Auszeichnung geht in diesem Jahr an den früheren US-Vizepräsidenten Al Gore und den UN-Klimarat.

Mit dieser Vergabe wolle das Nobelkomitee «zu größerer Aufmerksamkeit» für den Klimaschutz beitragen, hieß es am Freitag bei der Bekanntgabe in Oslo. Die Entscheidung des Nobelkomitees in Oslo wurde weltweit fast einhellig begrüßt. Gore (59) zeigte sich «zutiefst geehrt». Der Inder Rajendra Pachauri (67), der Chef des UN-Klimarats IPCC, sagte: «Ich bin völlig überwältigt.»

«Wir wollen durch unseren Preis nicht zuletzt mehr Aufmerksamkeit für das Thema Klimakatastrophe schaffen», sagte Komiteechef Ole Danbolt Mjøs. Der Klimawandel sei eine «wirkliche Bedrohung für die gesamte Menschheit». In der Begründung heißt es: «Handeln ist notwendig, bevor der Klimawandel für Menschen außer Kontrolle gerät.»

Auch Bundeskanzlerin Angela Merkel mahnte: «Die Gefahren des Klimawandels gehören zu den größten Herausforderungen für die Menschheit im 21. Jahrhundert.» Außenminister Frank-Walter Steinmeier erhofft sich von der Würdigung «weiteren Schwung» für die internationalen Klimaverhandlungen. UN-Generalsekretär Ban Ki Moon zeigte sich «sehr erfreut». Das wegen seiner Umweltpolitik von Gore häufig kritisierte Weiße Haus reagierte ebenfalls positiv. «Natürlich freuen wir uns», sagte Sprecher Tony Fratto.

Mjøs würdigte Gore als einen «seit Jahren weltweit führenden Umweltpolitiker», der wie kein anderer als «Einzelner zur Schaffung eines breiten Bewusstseins für die Maßnahmen beigetragen (hat), die nun nötig sind». Das Werk «Eine unbequeme Wahrheit» von Gore über die Gefahren des Klimawandels erhielt dieses Jahr einen Oscar als bester Dokumentarfilm. Der UN-Klimarat hat nach Überzeugung der norwegischen Juroren mit seinen wissenschaftlichen Berichten seit 20 Jahren entscheidend zu einem breiten Konsens über das Ausmaß der Klimabedrohung beigetragen. So hatte der Rat in seinem diesjährigen Report die Verantwortung des Menschen für die Erderwärmung so klar wie nie zuvor aufgezeigt.

Gore sagte in einer Erklärung: «Wir haben es mit einer wirklichen globalen Bedrohung zu tun. Die Klimakrise ist kein politisches Thema, sondern eine moralische und geistige Herausforderung für die gesamte Menschheit.» Die Auszeichnung sei umso bedeutungsvoller, als er sie mit dem UN-Klimarat teile, sagte Gore.

IPCC-Chef Pachauri sagte: «Dieser Preis geht an die internationale UN-Gemeinschaft und alle Staaten, die uns unterstützen.» Er habe vor der Entscheidung aus Oslo selbst öffentlich gesagt, dass Gore ihn wie kein anderer verdient habe.

UN-Generalsekretär Ban erklärte, das außergewöhnliche Engagement Gores sei ein Beispiel für die wichtige Rolle, die jeder Einzelne bei der Bewältigung weltweiter Probleme spielen könne. EU-Kommissionspräsident José Manuel Barroso gratulierte dem US- Politiker: «Ihre Beiträge zur Vorbeugung des Klimawandels haben das Bewusstsein in der ganzen Welt erhöht.»

Deutsche Umweltschützer und Politiker sahen in der Entscheidung einhellig ein Signal für stärkere Klimaschutz-Anstrengungen. Die wissenschaftlichen Arbeiten des IPCC zeigten, «dass wir jetzt handeln können und müssen», sagte die Kanzlerin. Bundespräsident Horst Köhler sagte, die profunde wissenschaftliche Analyse des Weltklimarates und das unermüdliche Engagement Gores hätten «maßgeblich dazu beigetragen, dass die Gefahren des Klimawandels nicht mehr allein in Expertenzirkeln diskutiert werden, sondern zum Gegenstand einer weltweiten öffentlichen Debatte geworden sind».

Das Nobelkomitee hatte schon 2004 mit der Vergabe an die Kenianerin Wangari Maathai Engagement für den Umweltschutz ausgezeichnet. Maathai hatte in ihrem Land gegen die Abholzung von Wäldern gekämpft. Zum Zusammenhang zwischen Klimawandel und Friedenssicherung meinte das Osloer Komitee, die Erderwärmung vermindere auch die «Sicherheit der Menschheit».

Nach Angaben des US-Senders CNN will Gore sein Preisgeld der von ihm gegründeten Organisation Alliance for Climate Protection zur Verfügung stellen. Sie setzt sich für Maßnahmen zur Bekämpfung der Klimakrise ein. Gore ist auch der Vorsitzende der Organisation.

Der Nobelpreis wird traditionell am 10. Dezember in Oslo überreicht. Im vergangenen Jahr erhielten der Bankier Mohammed Yunus und dessen Grameen-Bank die Auszeichnung für ihre Mikrokredite an arme Menschen.

Friedensnobelpreis an Gore und UN-Klimarat

Oslo (dpa) - Der diesjährige Friedensnobelpreis geht an den früheren US-Vizepräsidenten Al Gore und den UN-Klimarat. Das teilte das Nobelkomitee in Oslo am Freitag mit.

Gore und die Einrichtung der Vereinten Nationen werden für ihren Einsatz zur weltweiten Mobilisierung gegen eine drohende Klimakatastrophe ausgezeichnet. Der Friedensnobelpreis ist mit umgerechnet 1,1 Millionen Euro dotiert. Er wird traditionell am 10. Dezember in Oslo überreicht.

Analyse: Globaler Klimawandel wichtiger als US-Wahlkampf

Oslo (dpa) - Das norwegische Nobelkomitee hat nach Insiderinformationen lange gestritten, ob es den Friedensnobelpreis wirklich ein Jahr vor den US-Präsidentschaftswahlen einem prominenten Demokraten zuerkennen soll.

Als dann am Freitag in Oslo die Entscheidung für Ex-Vizepräsident Al Gore als Preisträger zusammen mit dem UN-Klimarat verkündet war, tat Komiteechef Ole Danbolt Mjøs Fragen nach etwaiger Wahlkampfhilfe für einen möglichen Kandidaten Gore demonstrativ ab: «Es interessiert uns herzlich wenig, was unsere Preisträger später aus dem Preis machen. Uns geht es um ihre vorherige Leistung für den Klimaschutz.»

Mjøs vermied sorgsam jeden Hinweis auf den von Klimaschützern immer wieder kritisierten US- Präsidenten George W. Bush. Aber dass die Vergabe an den Demokraten Gore indirekt auch als schallende Ohrfeige für den republikanischen Präsidenten gedacht ist, galt in Oslo als ausgemachte Sache.

Bei der Schilderung des globalen Klimaproblems griffen die drei Frauen und zwei Männer der Jury dann auch zu ganz großen Vokabeln, als sie ihre Entscheidung begründeten. Der Klimawandel sei eine «wirkliche Bedrohung für die gesamte Menschheit», sagte Mjøs, ein freundlicher Chefarzt im Pensionsalter aus der Polarstadt Tromsø. Dort sind die Folgen des Klimawandels mit immer schneller wegschmelzenden Eisbergen, einem bedrohten Bestand an Eisbären und schon jetzt schwer fassbaren Änderungen des Klimamusters für alle Einwohner eine beunruhigende Realität.

Vor allem wegen der grundlegenden globalen Bedeutung des Themas Klimawandel dürften die Bedenken gegen eine vielleicht nicht ganz «unparteiisch» wirkende Entscheidung für Gore am Ende weniger schwer gewogen haben. Dass das norwegische Komitee mit dem UN-Klimarat IPCC nun schon zum achten Mal eine Einrichtung der Vereinten Nationen ausgezeichnet, galt wohl auch intern als wenig umstritten.

Aber Entscheidungen wie für den IPCC oder vor zwei Jahren für die Internationale Atomenergiebehörde IAEA und ihren ägyptischen Chef Mohammed el Baradei wecken eben auch relativ wenig öffentliche Aufmerksamkeit. Da spielt Al Gore für die internationalen Medien doch in einer ganz anderen Klasse. Für seine «Fans» ist der 59-Jährige durch seinen Oscar im Februar für den Dokumentarfilm «Eine unbequeme Wahrheit» sogar in die Riege der Hollywoodstars aufgestiegen.

«Wir wollen durch unseren Preis nicht zuletzt mehr Aufmerksamkeit für das Thema Klimakatastrophe schaffen», sagte Mjøs vor Journalisten. Diese Aufmerksamkeit garantiert Gore, jedenfalls gemessen am Medieninteresse, mehr als etwa die ebenfalls als Favoritin gehandelte kanadische Umweltschützerin Sheila Watt-Cloutier aus der arktischen Inuit- Urbevölkerung.

Preisträger und Preisträgerinnen aus dieser Gruppe relativ unbekannter Vorkämpfer für Menschenrechte und Umweltschutz hatten die norwegischen Juroren in den letzten Jahren immer wieder ausgezeichnet. Etwa im letzten Jahr den Bankier Mohammed Yunus aus Bangladesch, 2004 die kenianische Umweltschützerin Wangari Maathai und im Jahr zuvor die iranische Juristin Schirin Ebadi.

Dieses Jahr ist das Nobelkomitee seiner Linie einer Erweiterung des Friedensbegriffes um Probleme wie Umwelt, soziale Ungleichheit und Menschenrechte treu geblieben. Nicht umsonst hieß es in der Begründung, der Klimawandel sei auch eine Bedrohung für die Sicherheit aller Menschen. Aber mit der Wahl von Gore entschieden sich die Juroren für einen völlig anderen Typ von Preisträger als in den Jahren zuvor. Das Presseaufgebot zur Verleihung in Oslos Rathaus am 10. Dezember dürfte das mit Abstand größte in der Geschichte des Nobelpreises werden.

19. Mai 2009 Nachrichten Seite 5

Karstadt testet Gravis als Multimedia-Partner

Essen (dpa) - Karstadt will zügig über eine Auslagerung seiner Multimedia-Abteilungen an den Computerhändler Gravis entscheiden. Noch in diesem Jahr soll ein erster Test starten.

Nach der sechsmonatigen Pilotphase werde über eine mögliche Ausweitung der Zusammenarbeit entschieden, sagte ein Sprecher der Karstadt- Muttergesellschaft Arcandor am Montag in Essen. In jedem Fall sei geplant, die bestehenden Multimedia-Abteilungen auch nach einer möglichen Übernahme durch Gravis in allen Karstadt- Warenhäusern beizubehalten. Vorgesehen für den Testbetrieb seien unter anderem die Karstadt-Filialen in Dortmund und Düsseldorf, hieß es in Unternehmenskreisen.

Erst Anfang des Monats hatte Karstadt angekündigt, dass künftig mehr als jede dritte Buchabteilung in den Warenhäusern geschlossen werden soll. Hintergrund ist eine geplante Übernahme der Buchabteilungen durch die Münchner Finanzholding DBH Buch Handels GmbH & Co. KG mit den Buchhandlungen Hugendubel und Weltbild.

Der Chef der Computerhandelskette Gravis, Archibald Horlitz, bezeichnete die geplante Zusammenarbeit mit Karstadt als «weiteren Meilenstein auf dem Expansionsweg» des Unternehmens. Karstadt setzte mit diesem Schritt seine Strategie konsequent fort, sagte Warenhauschef Peter Wolf.

Mit Weblins im virtuellen Raum das Licht anmachen

Hamburg (dpa) - Auf der Internetseite von eBay hat sich eine illustre Runde versammelt. Ein Astronaut hüpft scheinbar schwerelos auf und ab, ein vierarmiger Alien kratzt sich am Kopf und ein Eisbär sondert eine Sprechblase ab: «Hi. Wo kommt ihr her?»

Hinter diesen Pixel-Figuren am Bildschirmrand verbergen sich Menschen aus Fleisch und Blut: Mit dem kostenlosen Programm Weblin können sich Nutzer auf Internet-Streifzügen von einem Alter Ego begleiten lassen und für andere Surfer sichtbar werden.

Das Konzept stammt von der Hamburger Firma Zweitgeist. Die Überlegung der Macher: Surfer, die sich ähnliche Seiten anschauen, haben auch ähnliche Interessen. Weblin ermöglicht ihnen, unabhängig von Chats oder Diskussionsforen in Kontakt zu treten - wie im richtigen Leben etwa an der Supermarktkasse oder an der Ampel. «Das Internet ist ein virtueller Raum, und wir machen das Licht an», sagt Geschäftsführerin Christine Stumpf, die 2005 mit Heiner Wolf die Firma gegründet hat. Der Bedarf ist da: Laut einer Forsa-Umfrage hat jeder fünfte Deutsche im Internet Bekanntschaften gemacht.

Ob Mensch, Tier oder Außerirdischer: Nutzer wählen ihre virtuelle Identität selbst aus. Alle Avatare können mittels Sprechblasen chatten, fröhlich oder böse gucken, tanzen und winken. Die Gesprächsthemen variieren - je nach Seite. Während auf Dogforum.de Vierbeiner im Mittelpunkt stehen, geht es bei Spiegel.de um die Nachrichten des Tages. Häufig stehen die Figuren auch einfach herum. «Ich hab das Programm nur nebenbei an», schreibt eine Benutzerin.

Um beim Surfen sein virtuelles Alter Ego mitnehmen zu können, müssen Nutzer eine kostenlose Software installieren. Das Programm kontaktiert beim Besuch einer Internet-Seite den Weblin-Server und prüft, welche anderen User bereits dort aktiv sind. Deren animierten Figuren erscheinen dann. «Wir protokollieren nicht, welche Seiten unseren Nutzer sich anschauen», beruhigt Stumpf Datenschutz-Bedenken.

Damit aus dem virtuellen Gewusel handfeste Umsätze werden, plant Zweitgeist die Einspielung von Werbung. Zudem zahlen Premium- Mitglieder für besondere Funktionen und Avatare. «Das wird aber vermutlich ein kleinerer Anteil der User sein», sagt Stumpf. Von 2009 an schreibt das Unternehmen mit Weblin schwarze Zahlen, hofft die Geschäftsführerin.

Vorher will Zweitgeist vor allem die Nutzerzahlen steigern. Bisher liegt sie nach eigenen Angaben im fünfstelligen Bereich, bis zum Jahresende sollen es 100.000 sein. Bei der Verbreitung der Software setzt Zweitgeist auf Partner wie Microsoft. Mittlerweile sind auch eine englische und eine italienische Version der Software erhältlich. Im September lernen die Avatare noch Japanisch.

Dass Weblin für einen kommerziellen Erfolg wachsen muss, meint auch Prof. Armin Rott von der Universität Hamburg: «Soziale Netzwerke wie Weblin brauchen eine kritische Masse.» Erst ab 100 000 Nutzern sei eine solche Plattform für die Werbevermarkter interessant. Ob diese dann tatsächlich Werbung schalten, sei schwierig zu sagen. «Die Vermarkter wünschen sich originelle Werbeformen, schrecken aber manchmal vor der aufwendigen Umsetzung zurück», erklärt der Medienökonom. Insgesamt sei der Markt für Sonderwerbeformen noch in einer Experimentierphase.

Wie Werbung auf Weblin aussehen kann, zeigt eine erste Kampagne von Adidas. Der Sportartikelhersteller lässt bereits alle 90 Minuten auf einigen Seiten sportliche Figuren über den Bildschirm laufen. «Hi everybody», grüßt etwa Joggerin Susan. Die Astronauten und Aliens bekommen kommerzielle Gesellschaft.

Commerzbank scheitert mit Samstagsplänen an Betriebsrat

Berlin (dpa) - Die Commerzbank ist mit ihren Plänen für eine Samstagsöffnung der Filialen vorerst am Widerstand der Arbeitnehmervertreter gescheitert. «In diesem Jahr wird es zu keinen Öffnungen am Samstag mehr kommen», sagte ein Commerzbank-Sprecher der «Welt».

Ende Mai hatte Commerzbank- Privatkundenvorstand Achim Kassow angekündigt, dass im zweiten Halbjahr Kunden an rund 250 Standorten an einem oder mehreren Samstagen Zutritt zu den Filialen haben sollen. Die Bank wolle weitere Erfahrung mit dem in der Branche bislang wenig verbreiteten offenen Samstag sammeln. Bereits Mitte März hatte die Bank an einem bundesweiten Aktionstag die Samstagsöffnung getestet.

Die Pläne scheiterten nun am Widerstand der Arbeitnehmervertreter. «Wir halten die Samstagsöffnung grundsätzlich nicht für erforderlich. Wenn die Bank sie aber unbedingt will, dann müssen neue Stellen geschaffen werden», sagte Gesamtbetriebsratschef Uwe Tschäge der Zeitung. Der Tarifvertrag erlaube bislang eine eingeschränkte Öffnung am Samstag, allerdings nur mit Zustimmung der Betriebsräte. In den kommenden Wochen solle es zu erneuten Gesprächen zwischen Bank und Arbeitnehmervertretern kommen.

Pressestimmen: Von Schnapsideen und guten Vorschlägen

Hamburg (dpa) - In den Leitartikeln der deutschen Zeitungen überwiegt die Zustimmung zu den vorerst gestoppten Plänen der Familienministerin, Jugendliche als Testkäufer einzusetzen.

«Pforzheimer Zeitung»

Alle, die jetzt lauthals Schnapsidee rufen, sind herzlich dazu eingeladen, einen konstruktiven Beitrag zum Thema Jugendschutz zu leisten. Denn zumindest damit hat von der Leyen völlig recht: Das geltende Gesetz ist ein zahnloser Tiger, vor dem sich so lange niemand fürchten muss, wie er nicht zupacken kann. Das Debakel der Familienministerin sollte deshalb nicht das Ende der Debatte, sondern ihr Anfang sein.

«Landeszeitung Lüneburg»

Verschärfte Gesetze entfalten nur Wirkung, wenn man ihre Einhaltung wirksam kontrollieren kann. Das gilt nicht nur auf den Straßen, nachdem man Bußgelder für Raser und Drängler erhöht hat, sondern auch für Jugendschutz-Bestimmungen. So mag die Ministerin gedacht haben, als sie den durchaus diskussionswürdigen Plan zum Einsatz jugendlicher Testkäufer entwickelte. Umso bedauerlicher, dass sich Ursula von der Leyen von geballter Kritik stark beeindrucken und ihr Vorhaben fallen ließ. Kein einziger Kritiker verrät, wie man den schwarzen Schafen auf die Schliche kommen könnte, die das zarte Alter ihrer jungen Kundschaft nicht schert, wenn sie ihr Geld abknöpfen können für Alkohol, Zigaretten, Gewaltvideos oder Ballerspiele. Betreute Testkäufer erzielen beachtliche Wirkung, wie Pilotversuche zeigen. Beim Pilotversuch wird's bleiben, dem Bedenkenträgertum sei Dank.

«Flensburger Tageblatt»

In ihrer jetzt zurück genommenen Forderung stellt Frau von der Leyen auch den Behörden ein Armutszeugnis aus. Offensichtlich vertraut sie eher Kindern als Polizei und Ordnungsämtern. Der Plan, Kinder zu Denunzianten heranzuziehen, hätte auf Dauer ein höchst merkwürdiges Verständnis der Ministerin für gesellschaftliches und politisches Miteinander offenbart.

«Neue Osnabrücker Zeitung»

Weil Ursula von der Leyen ihren Testkäufer-Plan platt der Öffentlichkeit verkaufen wollte, muss sie jetzt kleinlaut den Rückzug antreten. Dabei ist die Idee gar nicht so absurd. In der Empörung über junge Spione geht unter, dass das Ministerium endlich dem Ruf nach mehr Kontrollen folgen wollte. Der eigentliche Skandal ist doch der, dass Minderjährige Schnaps, Zigaretten und Gewaltspiele kaufen können - trotz der Verbote. Es mangelt also am Bewusstsein, die Gesetze einzuhalten. Testkäufer können daran tatsächlich etwas ändern, wie Erfahrungen in den Schweizer Städten Bern und Zürich zeigen. Von Jugendlichen ertappt und prompt von den Behörden ermahnt, kontrollieren Verkäufer bei Folgeversuchen das Kundenalter genauer. So verstanden sind Testkäufe kein Mittel zur Bestrafung, sondern der Vorbeugung.

«Hamburger Abendblatt»

Fest steht: Die Rücknahme des umstrittenen Gesetzentwurfs löst das eigentliche Problem nicht. Die nicht abreißende Kette von Alkohol- und Gewaltexzessen unter Jugendlichen zeigt eindrücklich, dass der Jugendschutz in Deutschland zu wünschen übrig lässt. Die Frage, wie er wirksamer gestaltet werden kann, ohne dass dabei die eigentlich Schutzbedürftigen zum Einsatz kommen, müssen die Teilnehmer des runden Tisches beantworten. Und dabei sollten auch die Kritiker von der Leyens mit guten Vorschlägen aufwarten.

«Berliner Morgenpost»

Wieder einmal soll staatsaufsichtliches Regiment dort greifen, wo die ureigenste Pflicht der Erziehungsberechtigten gefragt ist. Wieder einmal soll der Staat anstelle der Eltern regeln, was im Fall der Gewalt-Computerspiele ohnehin vor allem daheim und nicht beim Händler schief läuft. Überall wo die Kinderzimmer am Internet hängen, gelangt die heiße Ware unter den Augen der Eltern in Kinderhand oder eben nicht. Die Lufthoheit über die Kinderzimmer gebührt nicht dem Familienministerium, sondern den Eltern. Was die verderben, kann auch die fürsorglichste Ministerin nicht richten. Und deshalb ist es sehr zu begrüßen, dass Frau von der Leyen ihr Kinderzimmer- Inspektionsprogramm jetzt zurückgezogen hat.

«Stuttgarter Nachrichten»

Von der Leyen jedenfalls scheut sich nicht, unkonventionelle Ideen ins Spiel zu bringen. Ihren Plan, Jugendliche als Testkäufer loszuschicken, hat sie zwar angesichts massiver Kritik vorerst auf Eis legen müssen. Was aber bieten diejenigen an wirksamen Maßnahmen an, die den Vorstoß als untauglich bezeichnen und oft so tun, als befehle von der Leyen den jungen Hilfspolizisten, die Flaschen an Ort und Stelle auszutrinken?

«Kölnische Rundschau»

Und führe uns nicht in Versuchung: Die vorletzte Bitte aus dem Vaterunser hat nun offenbar auch die Familienministerin erreicht. Sie verzichtet auf ihre Idee, jugendliche Testkäufer loszuschicken, um den Jugendschutz im Handel zu prüfen. Ein runder Tisch" soll über das Thema beraten, und dort wird die Diskussion im Sande verlaufen. Deutlich hörbar hatte die Kanzlerin Diskussionsbedarf angemeldet. Ursula von der Leyen reagierte noch rechtzeitig, bevor sie von höherer Stelle hätte gestoppt werden müssen. Der Einsatz von Lockspitzeln ist ja ohnehin heikel. Das Provozieren von Straftaten birgt die Gefahr, dass etwas "aufgeklärt" wird, was ohne Lockspitzel gar nicht passiert wäre.

«Stuttgarter Zeitung»

Würden Sie Ihre 14-jährige Tochter mit amtlicher Genehmigung in den Supermarkt oder an den Kiosk ums Eck schicken mit dem Auftrag, einen Flachmann oder ein Sixpack zu kaufen? Eben. Genau dies will aber Familienministerin Ursula von der Leyen ermöglichen im Dienste des Jugendschutzes. Sie will jugendliche Spitzel ausschicken, um die Gesetzestreue der Verkäufer zu testen. Wer könnte den Gesetzesbruch besser nachweisen, so die Ministerin, als jene, die zu schützen das Gesetz vorsieht. Aber Jugendliche, die fast noch Kinder sind, eignen sich nicht als Köder. Der Vorschlag muss vom Tisch, und zwar schnell. Jugendliche dürfen nicht für das Fehlen geeigneterer Kontrollmöglichkeiten in Haftung genommen werden.

«Offenbach-Post»

Wenn der Kinderschutzbund das als mit der Würde des Kindes unvereinbar bewertet, hat er völlig Recht. Und das Urteil des SPD-Gesundheitsexperten Karl Lauterbach, die Idee grenzt an Kindesmissbrauch, kann man nur dick und fett unterstreichen. Was mag sich von der Leyen, die eine Baustelle nach der anderen in ihrem Ministerium aufmacht und meint, die Gesellschaft mit vermeintlich modernen Gedanken und Gesetzen beglücken zu müssen, bloß dabei denken? Wahrscheinlich nicht allzu viel. Ihr Sendungsbewusstsein dürfte den Blödsinn in Sachen Kinder-Lockvögel übertüncht haben.

«Heilbronner Stimme»

Wie steht es um das Selbstverständnis einer Gesellschaft, die solche Alkoholexzesse zulässt? Nichts spricht dagegen, dass Jugendliche in Begleitung von erwachsenen Amtspersonen Testkäufe durchführen. 50 000 Euro Strafe sind daher angemessen, solange sich Händler weigern, das Alter der Kunden zu kontrollieren. Beim Verkauf von Killerspielen, Zigaretten und Alkohol wurde zu lange weggeschaut. Es wird Zeit, dass die Öffentlichkeit endlich die Augen öffnet.

Bundesligist Werder Bremen vor Aufnahme in G14

Hamburg (dpa) - Werder Bremen soll im Zuge der Erweiterung der Interessenvereinigung der europäischen Fußball-Großclubs offenbar noch in diesem Jahr in den Kreis der so genannten G14 aufgenommen werden.

Der Bundesligist solle eines von 22 neuen Mitgliedern der Organisation werden, die vor einer umfassenden Reform steht, erklärte G14- Geschäftsführer Thomas Kurth in einem Interview mit dem «Handelsblatt».

«Zu den 18 aktuellen Mitgliedern wird aus den 22 bestplatzierten Ländern in der Uefa-Fünfjahreswertung jeweils ein neuer Verein hinzukommen, und zwar jeweils der über die letzten fünf Jahre in der nationalen Meisterschaft am besten klassierte. In Deutschland ist das zum Beispiel Werder Bremen», sagte Kurth. Die Aufnahme solle noch 2007 erfolgen. «In einem nächsten Schritt werden wir uns dann über Europa hinaus auch auf andere Kontinente erweitern.»

Der 50-jährige Schweizer geht davon aus, dass die «neue unabhängige Club-Organisation künftig die Stellung des UEFA-Club- Forums einnehmen wird». Das Club-Forum, das der europäische Fußballverband als Gegenmodell zur G14 gegründet hat und dem Bayern Münchens Vorstandschef Karl-Heinz Rummenigge vorsitzt, bezeichnete Kurth im «Handelsblatt» als «Alibi-Übung».

Analyse: Bahn-Angebot bringt keinen Durchbruch

Berlin/Frankfurt (dpa) - Das sei nun schon das fünfte Angebot an die Lokführergewerkschaft GDL, stellte die Bahnmanagerin fest. Personalchefin Margret Suckale wollte in Berlin keinen Zweifel aufkommen lassen, dass sich das Unternehmen abermals bewegt habe, um doch noch zu einer Tarifeinigung zu gelangen und weitere Streiktage zu vermeiden.

Es gebe in Deutschland «wohl kein anderes Angebot», das den Arbeitnehmern so viel biete. «Jetzt muss die GDL an den Verhandlungstisch kommen.» Doch damit wird es vorerst nichts.

Auf den ersten Blick schien die Bahn die Kernforderung der GDL erfüllt zu haben. Deren streitbarer Vorsitzender Manfred Schell verlangt seit Beginn der Auseinandersetzung im Frühjahr einen «eigenständigen» Tarifvertrag. Er soll den Lokführern außer einem kräftigen Lohnaufschlag in dieser Tarifrunde in Zukunft zu einer stärkere Stellung innerhalb der Belegschaft verhelfen. Die Bahn sprach nun von einem «eigenen» Tarifvertrag, den sie für die Lokführer abzuschließen bereit sei.

«Einen eigenen Tarifvertrag haben wir schon seit 20 Jahren», konterte die Gewerkschaft Deutscher Lokomotivführer (GDL). Doch statt das Angebot aus dem Berliner Bahntower als neue Mogelpackung gleich rundweg abzulehnen und zu neuen Streiks aufzurufen, will die GDL erst noch klären, ob die Bahn vielleicht einfach nur ein Wort verwechselt hat. Mitte der Woche soll es dazu ein «klärendes Gespräch» geben. Dabei handelt es sich ausdrücklich nicht um Tarifverhandlungen, für die das neue Angebot eigentlich die Basis schaffen sollte. Erst nach dem Gespräch werde entschieden, ob die GDL in Verhandlungen mit der Bahn eintrete - oder zu neuen Streiks aufrufe.

Tatsächlich hatte der Bahnkonzern schon als Ergebnis eines Moderatorenverfahrens Ende August einen «rechtlich eigenständigen» Tarifvertrag für die Lokführer akzeptiert. Weil aber die parallele Absprache mit den konkurrierenden Gewerkschaften Transnet und GDBA scheiterte, wurden noch nicht einmal Verhandlungen über diesen Vertrag aufgenommen. Erst ein Spitzengespräch am vergangenen Donnerstag mit Bahn-Aufsichtsratschef Werner Müller brachte wieder Bewegung in den verfahrenen Tarifkonflikt und nährte die Hoffnung auf ein baldiges Ende oder wenigstens eine Streikpause bis Ende Oktober.

Umso größer war die Enttäuschung: «Wir erkennen in diesem Angebot im Grunde nichts Neues», sagte der GDL-Vorsitzende Schell. Der einzige Unterschied bestehe in der Auszahlung von 1400 Euro für 104 bereits geleistete Überstunden. Das sei «in jedem Fall unzureichend», sagte Schell. Die GDL verlangt eine deutliche Einkommensverbesserung für das Fahrpersonal und eine Verkürzung der wöchentlichen Arbeitszeit von 41 auf 40 Stunden. Dass es für Mehrarbeit auch mehr Geld gibt, hält Schell eigentlich für eine Selbstverständlichkeit.

Viel wird nun von dem klärenden Gespräch mit dem Bahn-Vorstand abhängen. GDL-Chef Schell wird selbst nicht daran teilnehmen: Er fährt für drei Wochen zur Kur an den Bodensee. Schon zweimal hatte er den Gesundheitsurlaub in der Hoffnung auf eine baldige Lösung des Tarifkonflikts schon verschoben. Ein drittes Mal wollte ihn der erholungsbedürftige Gewerkschaftsfunktionär nicht mehr verschieben.

Analyse: Preiserhöhung mit Ansage

Düsseldorf (dpa) - Gerade einen Monat ist es her, dass E.ON-Chef Wulf Bernotat die Verbraucher mit seiner Ansicht geschockt hatte, Strom sei in Deutschland «eigentlich zu billig».

Eine dreiköpfige Normalfamilie koste die Lebensqualität aus der Steckdose «täglich etwa so viel wie eine Kugel Eis pro Kopf», rechnete der Chef des größten deutschen Energiekonzerns vor und löste damit eine Welle der Empörung aus. «Bernotat verhöhnt Verbraucher und energieintensive Unternehmen», polterte beispielsweise Hessens Wirtschaftsminister Alois Rhiel (CDU).

Bernotat sah sich zwar zu einer Entschuldigung für seine Wortwahl gezwungen. An der Grundaussage, Energie werde «perspektivisch teurer», hielt er aber fest. Vier Wochen später wissen die mehr als sieben Millionen E.ON-Kunden, was das zumindest kurzfristig für sie bedeutet. Um bis zu 9,9 Prozent erhöhen die E.ON- Regionalversorger zu Beginn kommenden Jahres die Strompreise. Bis zu 5,80 Euro im Monat muss ein Durchschnittshaushalt mehr zahlen.

Die E.ON-Ankündigung ist der Höhepunkt der jüngsten Preisrunde beim Strom. Nach dem Wegfall der staatlichen Strompreisgenehmigung zum 1. Juli hatten zunächst nur einige Versorger an der Preisschraube gedreht. Jetzt kommt es zu teilweise kräftigen Aufschlägen. Der Branchen-Zweite RWE erhöht bei einigen Versorgungstöchtern die Preise um bis zu 9,8 Prozent. Bei der EWE in Oldenburg müssen sich die Kunden schon zum 1. November auf eine Preisanhebung um 14,2 Prozent einstellen.

Den Länder-Wirtschaftsministerien, die früher ungerechtfertigte Preiserhöhungen kürzen konnten, sind inzwischen weitgehend die Hände gebunden. Nordrhein-Westfalens Wirtschaftsministerin Christa Thoben (CDU) kündigte zwar eine kartellrechtliche Überprüfung «mit sehr strengem Maßstab» an, doch das Ergebnis steht in den Sternen. «Große Stromkonzerne wie E.ON machen in Deutschland was sie wollen», kommentierte die Grünen-Fraktionsvize im Bundestag, Bärbel Höhn, fast resigniert die Entwicklung.

Den Politikern blieb am Montag nur der Appell an die Verbraucher, teuren Stromversorgern «schnellstmöglich» die rote Karte zu zeigen und zu preiswerteren Anbietern zu wechseln. Nur so sei «mehr Wettbewerb auf den Energiemärkten» zu erreichen, glaubt Thoben. Spielraum zum Wechseln gebe es genug. In Nordrhein-Westfalen hat das Ministerium unlängst eine Preisdifferenz von 200 Euro zwischen dem teuersten und dem preiswertesten Anbieter festgestellt.

Dass die Stromkunden inzwischen bereit sind, ihrem gewohnten Versorger zu kündigen, bekam Vattenfall Europe im Sommer zu spüren. Nach einer Preiserhöhung in Hamburg und Berlin verlor das Unternehmen, das auch wegen seiner Informationspolitik bei den Pannen im Kernkraftbereich in der Kritik stand, dort mehrere zehntausend Kunden. Für den Jahresbeginn 2008 plant Vattenfall nach eigenen Angaben aktuell keine Aufschläge.

Figure 8: Generated Layouts for Different Style Instances

and the small portion of it which is searched with the algorithms.

In addition, we conducted further experiments based on the objective function for the page score, with different types of optimization algorithms (the Grouping Genetic Algorithm by Singh et al (Singh and Baghel, 2007), Simulated Annealing and Greedy Algorithms similar to Julstrom (Julstrom, 2005)) on a corpus of news articles (with images) from DPA[6]. In these experiments we varied the parameters of the scoring formula as well as the settings of the style guide. The results that were obtained (cf. Figure 8) are promising and show the basic correctness of the approach and its applicability for different instances of the style.

6. Conclusion and Outlook

The general approach with a minimum set of aesthetic constraints was shown to be valid already, as Strecker et al showed in (Strecker and Hennig, 2009). The presented extension of the objective function to include further aesthetic measures and personal preferences results in even better structured and organized results which go far beyond the results of any existing scientific prototype or existing system. While the current evaluation is based on personal observation and the achieved scores for the different terms of the objective function, future evaluation will include user studies to ensure both the quality of the generated layout and the adherence to the personalization options.

The presented system allows for the flexible and dynamic generation of personalized newspaper without manual intervention, thus closing the gap between manual editing of newspaper layouts and personalized information consumption.

Further research could be performed in several directions. For example, additional options for each article layout could be defined, which would finally lead to an even more individual and authentic reading experience. Another option is the inclusion of additional aesthetic and content measures, e.g. the automatic grouping of articles belonging to similar topics or the combination of similar articles into a single one, of which a basic version was explored by my colleague and me in (Hennig and Strecker, 2008). An important direction is a further selection and the optimization of the algorithms used based on the properties of the search space and data set which we discovered during our research.

7. References

J. E. Beasley. 2004. A population heuristic for constrained two-dimensional non-guillotine cutting. *European Journal of Operational Research*, 156(3):601–627.

Max Bense. 1965a. *Aesthetica. Einfhrung in die neue Aesthetik.* Agis-Verlag, Baden-Baden, Germany.

Max Bense. 1965b. *Projekte generativer Ästhetik.* edition rot, Stuttgart.

Max Bense. 1969. *kleine abstrakte ästhetik.* edition rot, Stuttgart.

Marc Bernard and François Jacquenet Lirsia. 1998. Discovering rules to design newspapers: An inductive constraint logic programming approach. *Applied Artificial Intelligence*, 12:547–567(21).

George David Birkhoff. 1933. *Aesthetic Measure.* Harvard University Press, Cambridge, MA, USA.

Morten Buhr. 1996. Newspaper layout aesthetics judged by artificial neural networks.

David Collier. 1990. *Collier's Rules for Desktop Design and Typography.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Michael R. Garey and David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freeman, San Francisco.

Gary Greenfield. 2005. On the origins of the term ”computational aesthetics”. In *Computational Aesthetics*, pages 9–12. Eurographics Association.

Steven J. Harrington, J. Fernando Naveda, Rhys Price Jones, Paul Roetling, and Nishant Thakkar. 2004. Aesthetic measures for automated document layout. In *Proceedings of DocEng '04*, pages 109–111. ACM Press.

Leonhard Hennig and Thomas Strecker. 2008. Tailoring text for automatic layouting of newspaper pages. In *Proc. of ICPR '08*, pages 1–4, Dec.

Amanda Hiley and Bryant A. Julstrom. 2006. The quadratic multiple knapsack problem and three heuristic approaches to it. In *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 547–552, New York, NY, USA. ACM.

InterTi GmbH. 2010. niiu, die individualisierte tageszeitung. http://www.niiu.de/. last retrieved on 02/01/2010.

Charles Jacobs, Wilmot Li, Evan Schrier, David Bargeron, and David Salesin. 2003. Adaptive grid-based document layout. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, pages 838–847, New York, NY, USA. ACM.

Charles Jacobs, Wil Li, Evan Schrier, David Bargeron, and David Salesin. 2004. Adaptive document layout. *Commun. ACM*, 47(8):60–66.

Bryant A. Julstrom. 2005. Greedy, genetic, and greedy genetic algorithms for the quadratic knapsack problem. In *Proceedings of GECCO '05*, pages 607–614. ACM Press.

Hans Kellerer, Ulrich Pferschy, and David Pisinger. 2004. *Knapsack Problems.* Springer, Berlin.

Xiaofan Lin. 2006. Active layout engine: Algorithms and applications in variable data printing. *Computer-Aided Design*, 38(5):444 – 456.

Jonas Martinsson. 2009. Feedjournal, the newspaper you always wanted. http://www.feedjournal.com/. last retrieved on 02/17/2009.

David Chek Ling Ngo and John G. Byrne. 1998. Aesthetic measures for screen design. In *OZCHI '98: Proceedings of the Australasian Conference on Computer Human Interaction*, page 64, Washington, DC, USA. IEEE Computer Society.

David Chek Ling Ngo, Lian Seng Teo, and John G. Byrne. 2003. Modeling interface aesthetics. *Information Sciences*, 152:25–46.

[6] Deutsche Presseagentur

David Chek Ling Ngo. 1994. *VISIT: visitor information system implementation tool*. Ph.D. thesis, Trinity College Dublin, Ireland.

Rolf F. Rehe. 1985. *Typography and design for newspapers*. IFRA, Germany.

Sartaj Sahni. 1975. Approximate algorithms for the 0/1 knapsack problem. *Journal of the ACM*.

Tugba Saraç and Aydin Sipahioglu. 2007. A genetic algorithm for the quadratic multiple knapsack problem. In *BVAI*, volume 4729 of *Lecture Notes in Computer Science*, pages 490–498. Springer.

Evan Schrier, Mira Dontcheva, Charles Jacobs, Geraldine Wade, and David Salesin. 2008. Adaptive layout for dynamically aggregated documents. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 99–108, New York, NY, USA. ACM.

Andrew Lee Sears. 1993. Layout appropriateness: A metric for evaluating user interface widget layout. *IEEE Trans. Softw. Eng.*, 19(7):707–719.

Alok Singh and Anurag Singh Baghel. 2007. A new grouping genetic algorithm for the quadratic multiple knapsack problem. In *EvoCOP*, volume 4446 of *Lecture Notes in Computer Science*, pages 210–218. Springer.

Thomas Strecker and Leonhard Hennig. 2009. Automatic layouting of personalized newspaper pages. In *Proc. of OR '08*, pages 469–474, Berlin, Heidelberg. Springer-Verlag.

Dennis J. Streveler and Anthony I. Wasserman. 1984. Quantitative measures of the spatial properties of screen designs. In Brian Shackel, editor, *INTERACT 84 - 1st IFIP International Conference on Human-Computer Interaction*, pages 81–89, London, UK.

Syntops GmbH. 2009. Personalnews. http://syntops.de/index.php?id=13&L=2. last retrieved on 02/23/2009.

Tabbloid. 2009. Tabbloid, turn your favorite feeds into a personal magazine. http://www.tabbloid.com/. last retrieved on 02/17/2009.

Thomas S. Tullis. 1997. Screen design. In Marting G. Helander, Thomas K. Landauer, and Prasad V. Prabhu, editors, *Handbook of Human-Computer Interaction (Second Edition)*, pages 503–531. North-Holland, Amsterdam, second edition edition.

Gerhard Wäscher, Heike Haußner, and Holger Schumann. 2007. An improved typology of cutting and packing problems. *European Journal of Operational Research*, 127(3):1109–1130, December.

Personalized Learning to Rank Applied to Search Engines

Christian Scheel, Ernesto William De Luca, Sahin Albayrak,
Sascha Narr, Orlando Macone, Dennis Egert, Matthias Hohner

Technische Universität Berlin,D-10587, Germany
christian.scheel@dai-labor.de, ernesto.deluca@dai-labor.de, sahin.albayrak@dai-labor.de,
sascha.narr@dai-labor.de, macone@cs.tu-berlin.de, dennis.hu.egert@campus.tu-berlin.de, matthias.hohner@tu-berlin.de

Abstract

Learning to rank is a successful attempt of bringing machine learning and information retrieval together. With learning to rank it is possible to learn ranking functions based on user preferences. User preferences for instance depend on user background, features of results, relations to other entities and the occurrence of the searched entities in presented abstracts. The reasons why there are only some applications utilizing learning to rank for personalization can be found in the extended query response time and general additional resource needs. These resource needs come from the use of machine learning and the need to learn and use trained user models. Experiments on standard benchmark data help showing that learning to rank approaches perform well, but currently it is not possible to show how much feedback is needed for an improvement or if personalization is possible. Hence the minimal number of training data for creating a ranking function is not known. We show that keeping the training data as small as possible minimizes the resource needs and even enables the possibility of training personalized ranking functions. In this work we apply learning to rank to an existing search engine and evaluate the conditions and effects of learning personal preferences. We evaluate how much implicit feedback is needed and use this knowledge to reduce the computational requirements, enabling learning to rank based personalization.

1. Introduction

The success of the link analysis algorithm PageRank (Page et al., 1999) shows that preferences in information retrieval are not completely content-driven. There are ranking functions fusing several of such quality and relevance opinions. There are even some ranking algorithms trying to detect your favorite soccer team, your location and other favorites by analyzing past searches. But these ranking algorithms do not consider users individual preferences to compute a rank for a document.

There are several aspects of results which can influence preference decisions, yet user-dependent aspects are not usually being taken into consideration when presenting search results. It is very intuitive to rank results by the following preference statement: “I like to have search terms occurring in the presented title and in the presented text snippet. Also it is always good to have at last one headline in the result including the search terms.”. Very often the factors influencing a preference decision are unconscious and not known to the user, but can be detected by learning to rank.

1.1. Relevance Feedback

One way to achieve such personalization is collecting relevance feedback. Users can be asked for explicit feedback about their preferences, creating detailed user profiles. However, users may not be willing or able to explicitly provide such information and would feel bothered by such an inquiry.

One solution for avoiding this problem is to analyze the click logs of the user as proposed in (Joachims, 2002; Radlinski and Joachims, 2005). The proposed preference rules applied to implicit feedback have proven to result in representative user preferences.

1.2. Learning to Rank

In information retrieval learning to rank algorithms attempt to sort search results according to their individual relevance to the query. There are several learning to rank algorithms proposed each year (Herbrich et al., 2000; Freund et al., 2003; Tsai et al., 2007; Cao et al., 2007; Liu et al., 2007; Guiver and Snelson, 2008).

Learning to rank algorithms analyze user preferences creating a user model for predicting preference decisions of unseen data. With learning to rank it is possible to learn a user model for the general user, individual users or user groups. However, to authors’ knowledge, there are no evaluation results for personalized learning to rank in the literature.

Based on evaluations on benchmark data, learning to rank has a very good retrieval performance. Unfortunately there are drawbacks resulting from the machine learning background of the learning to rank algorithms. The training time directly depends on the number of features and the number of training instances. Hence on given datasets performance evaluations take up to several hours. Fortunately the search scenario drawn by given datasets is not exactly the scenario of a running search engine (Zhang et al., 2009).

Nevertheless there are requirements for search engines like query response time, which have to be respected. Therefore there is a need of selecting a proper learning to rank approach, the most influential features and the most distinguished user preferences. In other words, for fast and reliable model creation it is necessary to collect and use as little, but meaningful data as possible.

1.3. Contribution

The best solution for seamless learning to rank at query time is to limit the time needed to rank a given set of results. In contrast to data sets, features of these results have to be retrieved first, before any learning to rank approach can make decisions based on them. To improve response time when using reranking, ideally to real-time, we focus on efficient ways for preprocessing features.

To demonstrate our approach, we apply RankSVM (Herbrich et al., 2000), showing that the chosen learning to rank approach is not the main bottleneck in practice.

In this paper we describe algorithms and methods applied to integrate learning to rank into search engines. We especially focus on tolerable response times and personalization issues. The integration is done exemplarily on the search engine *PIA* allowing it to provide personalized search results.

Experiments with thousands of queries show that a learning to rank approach can be used in real-world search engines. In Sec. 5. it is evaluated how effectively learning to rank can be used to provide personalized result lists. Furthermore it is evaluated how much implicit feedback is necessary to successfully apply learning to rank.

2. Learning to Rank on Corpora

In classic information retrieval results (for instance documents as used in our work) are represented by word vectors. For learning to rank results are represented by vectors whose elements are opinions about features and relations of these results. The objective of learning to rank is to learn a ranking function which explicitly learns preferences instead of using functions like TF-IDF or BM25.

Because information retrieval performance is the main benefit of learning to rank we summarize and discuss the features of the most important learning to rank benchmark dataset to conclude features for a real world application.

Most of the current learning to rank algorithms were developed and evaluated on the *LE*arning *TO R*ank datasets in LETOR (Liu et al., 2007). In different versions of LETOR different datasets were uniformly transformed to be suitable for the problem of learning to rank.

Each search result in LETOR is represented as a set of features which can be used to learn to rank these results. The information of which result should be ranked higher is derived from user given labels. Mostly these labels are 0 for not relevant and 1 for relevant results with respect to a given query. There is one dataset in LETOR with three levels of relevance, OHSUMED (Liu et al., 2007).

2.1. Result Features

When using LETOR datasets the data is already preprocessed and ready to be used in experiments, hence researchers do not have prepare data sets and evaluation results are comparable to each other.

Features in the LETOR datasets are classified into low-level, high-level, hyperlink and hybrid features. Low-level features for instance are “term frequency of URL” or “document length of body”. “BM25 of title” or “LMIR of anchor” belong to the high-level features. “PageRank” and “Topical HITS authority” are part of the hyperlink features, but there also are hybrid features like “Hyperlink base score propagation: weighted in-link”.

2.2. Implicit Search Task

In LETOR the following search scenario is drawn (exemplary for the included dataset TD2003 of LETOR 3.0): When a query is made, up to 1,000 results are returned. From these results feature vectors have to be extracted at query time. These feature vectors are needed for the ranking algorithms which have to be trained (offline) and used at query time to decide the optimal ranking for these results. Learning to rank is used to learn the general, non-personalized ranking function when all users have provided relevance decisions for the results of 30 queries (3/5 of number of queries in TD2003). The given relevance feedback is used to create a learning to rank model which is used to rerank each further result list.

2.3. Computational Performance

Extracting feature vectors for 1,000 documents, as described in Sec. 2.2., takes a lot of time. Although there is no detailed information given how much time is needed, it is clearly not feasable in real time. Additionally, building a learning to rank model or rather training the given queries takes a lot of resources, but the provided benchmark learning to rank baselines are not connected to training times. It is possible to build the model in the background or on external servers. Own experiments have shown, that the learning to rank approach RankSVM (Herbrich et al., 2000), based on the Support Vector Machine algorithm), takes the most training time and ListNet (Cao et al., 2007) is the fastest approach among those tested.

3. Applying Learning to Rank

Before integrating learning to rank into a search engine, the learning to rank phases have to be identified and analyzed with respect to given quality and response time restrictions. We propose to extend the learning to rank process by the phases presented in Fig. 1. These phases include recording clicks in a log, deducing the preferences, determining the online and offline features, training the learning to rank model and reranking a given result list accordingly.

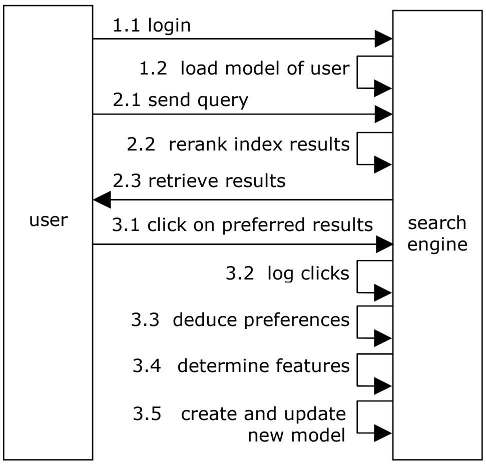


Figure 1: Interaction with a learning to rank search engine including the five phases of the learning to rank process: logging, preference retrieval, feature retrieval, training and reranking. If there is no model yet or the user does not log in, the results coming from the index are passed.

For the phases of logging, preference retrieval and reranking we follow and extend the proposed approaches of

(Joachims, 2002; Joachims and Radlinski, 2007) to apply learning to rank by respecting search engine requirements.

3.1. Logging

A method to compile user's preferences is the retrieval of relevance feedback. In general, it is easier to obtain implicit feedback rather than explicit feedback from a user since the user's benefit of providing feedback on search results is not clear to him. Instead of asking the user for explicit feedback, such as asking the user to rate the presented results, implicit feedback like clicks on results can be used.

The proposed logging in (Joachims, 2002) includes information about who clicked on which result for which search. This data is called clickthrough data and is retrieved by a log server for later preference retrieval. Logging can be done in parallel and does not affect the search engine's response time.

3.2. Preferences

Before learning preferences, the logged data has to be transformed to preference relations. Therefore the rules to generate preference relations given in (Radlinski and Joachims, 2005) were applied. With these rules and the clickthrough data, a set of preferences for each user can be deduced. These preferences will later form the training data for the learning to rank algorithm.

3.3. Features

There are two use cases where results have to be transformed to a feature vectors. The transformation is needed first, for training preference pairs and second for using a trained model to rerank results at query time.

Before preference pairs can be trained, the results have to be expressed as feature vectors containing relevance and quality opinions of several information retrieval strategies. Retrieving these opinions can be very time consuming. Fortunately, the transformation and the learning can often be done offline and do not have to be performed at query time. Additionally, by limiting the number of features, learning to rank algorithms usually need less training time.

The processing time of the transformation of potential results to feature vectors depends on the type of features (information retrieval strategies) used. Generally, features can be divided into two categories: Quality-based features and query-dependent features. Quality-based features can be processed offline, because they depend on general features like *number of pages* or *PageRank*. Query-dependent features are the scores coming from retrieval strategies which for instance address the *term frequency of the query terms* in different parts of a result. These features can only be processed online, at query time.

In order to compute a feature vector $\vec{f}$ that describes a given result document, properties of results have to be used, which are expressible as real numbers. The feature vector consists of two sub-vectors:

- $\vec{g}$, containing real values for quality-based features.
- $\vec{h}$, containing real values for query-dependent features.

3.3.1. Quality-based features

Quality-based features are features of a result that can be computed independently of the search query. They describe result properties such as word counts, age or PageRank, or in the case of documents number of authors, citations or file size. The static nature of these features makes it ideal for them to be aggregated and stored at idle server times, allowing them to be accessed much faster at real-time during a reranking procedure.

3.3.2. Query-dependent features

Query-dependent features are features of a result that are dependent on the search query. Therefore, they have to be computed during query time. The query-dependent feature vector $\vec{h}$ itself is divided up into two groups:

- Statistical query-dependent features
  Vector $\vec{j}$, which contains feature values for features that are based on statistics over search terms.
- Query reformulation query-dependent features
  Vector $\vec{k}$, which contains values for features that are computed through slight reformulation of the search query. The main idea is to look in adjacent directions in search space with respect to the query phrase. Using that approach makes it possible to automatically weight certain words from the search phrase lower or higher and to omit one or more words. Additionally, words that are spelled similar are detected by fuzzy search. Few documents are usually returned if one word of the search phrase is misspelled, but they should have higher values for query-dependent features that are based on similar words.

Due to the fact that the feature values depend on the search query, they have to be computed very fast, ideally in realtime, as opposed to the quality-based feature values that can be computed at fixed time points.

3.3.3. The Feature vector

The feature vector $\vec{f}$ for one result is composed of the quality-based feature vector $\vec{g}$ and the query-dependent feature vector $\vec{h}$, which is subdivided into two vectors, statistical query-dependent features $\vec{j}$ and query reformulation query-dependent features $\vec{k}$:

$$\vec{f} = \left\{ \vec{g}(\vec{j}\vec{k}) \right\} \tag{1}$$

3.4. Training

Representing search results as feature vectors f, we can deduce preference pairs. A preference consists of two results of one search. A result that have been clicked by the user and one that have not been clicked.

The training data T is defined as

$$T = \left\{ (\vec{t_k}, n_k), (\vec{t_k}^{-1}, {n_k}^{-1}) \right\} \tag{2}$$

where $\vec{t_k} = \vec{f_i} - \vec{f_j}$ and the document d relation $(d_i \succ d_j) \rightarrow n_k = 1$. However only having $(\vec{t_k}, n_k)$ will generate training data for one class only, namely for the class "preferred" (clicked search results). To successfully train learning to rank algorithms, data from the class "not preferred" (search results that have not been clicked) is needed, too, since reranking also means to rank non-prefered results lower. $(\vec{t_k}^{-1}, {n_k}^{-1})$ represents data from the other class, where $\vec{t_k}^{-1} = \vec{f_j} - \vec{f_i}$ and $(d_i \succ d_j) \rightarrow n_k = -1$.

The training phase results in a trained model which represents a ranking function. If a single user has provided enough feedback to build a training data set, then the learning to rank algorithm can create a personalized model. For later use, this model should be persistable.

In general, creating a learning to rank model via training can be done at any time since only collected clickthrough data is needed to compute the model. For users' best search experience, result lists should become more personalized as more feedback is given. Hence, each provided feedback should start the training phase (computed in parallel). In the optimal case, the resulting model can be used for ranking the results of the next search.

3.5. Ranking

As soon as a user sends a query to the search engine the normal index based search is performed. The resulting list is then reranked using the latest model retrieved from the training phase (resp. deserialized model from login, see Fig. 1).

For the first n entries in the result list, feature vectors are generated. To generate the feature vectors the query-dependent features are determined and the result's quality-based features are retrieved from preprocessed data. Note that these feature vectors have to be structured in the same way as the vectors in the training phase.

The model is used to rank these feature vectors and the results have to be ranked accordingly.

When the model is used to reform a result list for a given search query, the top n (for example 30) results of the result list are ranked according to the model's ranking function. In order to do this, the feature vectors for the top n results have to be computed at query time, which increases the response time of the search engine.

4. Settings and Methods

In this section we describe the settings of our experiments. All evaluations are based on the search engine *PIA* which was extended to rank results by learning to rank using a SVM. The measurements were performed on a standard system with 1.3 GHz single-core Intel processor with 2 GB RAM.

4.1. The Search Engine *PIA*

Applying learning to rank to a search engine can be done in a completely unintrusive manner. Everything needed for reranking are result sets and connected clickthrough data. Hence it is possible to superimpose a reranking module communicating with the search engine. The only restriction is that the results can be represented as uniform features.

Because the offline retrieval of quality-dependent features can only be done when knowing the objects which are potential results, we exemplary extended an existing search engine where the belonging content base is fixed.

Following the five phases, this section describes the implementation of learning to rank algorithms into the search engine *PIA*. *PIA* is a retrieval architecture developed by *DAI-Labor* that allows search in scientific publications. For evaluation we applied learning to rank to a *PIA* implementation where the document base is limited to ACM[1] documents. *PIA* includes a user management system which allows to collect not only general, but personalized user feedback.

Training and Reranking For applying learning to rank to *PIA* we have chosen the support vector machine approach(Herbrich et al., 2000). Using the SMOClassifier from the WEKA toolkit(Witten and Frank, 2005). The SVM was trained with the training data of each registered user individually.

In contrast to the usual approach of ranking by the distance to the class separating hyperplane, we used the model to determine the preferences of all results compared to each other. The results were then ranked by the total number of times they were preferred over any other result; in other words the reranking is based on predicted preference popularity and the most preferred result is listed on top.

We have chosen the first 30 results for reranking and left the order of the following results untouched. We did not apply any feature selection strategy.

Preferences For extracting preferences the two most intuitive rules were applied; namely "Click $\succ$ Skip Above" and "Click First $\succ$ No-Click Second" as proposed in (Radlinski and Joachims, 2005). When clicking a result, the "Click $\succ$ Skip Above" results in preferences where the clicked result is preferred over all preceding results in the result list which were not clicked. In "Click First $\succ$ No-Click Second" the notion is that if a user clicks on the top result and not the second ranked result, the top ranked result is preferred over the second ranked result.

Selected Features For the retrieval of query-dependent features, the Apache Lucene indexing architecture[2] is used to build small and fast temporary indexes for the top n documents. Doing this, this retrieval only has to be performed in a small RAM index instead of crawling through the main, hard-disk based index. Query-dependent features are received from these indexes by doing query reformulations like *terms as phrase* (see Appendix for a complete list). The resulting scores for each of the indexed documents are the query-dependent feature values.

Having a small index instead of a complete one influences some feature values. Nevertheless, this procedure is reasonable since we are only interested in query-dependent feature values for documents that appear in our initial search results.

For the retrieval of quality-based features we preprocessed the belonging values. Static meta-data like *number of pages* is added to the indexed documents, changing values like *PageRank* are inserted into another index which is updated when necessary.

Using an additional index is of course not helpful for fast feature retrieval. On the other hand documents in the main index do not have to be re-indexed, when quality-based

[1] `http://portal.acm.org`

[2] `http://lucene.apache.org/`

features change. For user-dependent future extensions like *user has seen result* such an index becomes necessary nevertheless.

Hence, to receive all necessary features for a list of results, these results are indexed to retrieve query-dependent features. Some quality-based features are taken from document fields itself and some by querying another index.

4.2. User model

For testing the architecture we set up a user model to generate users which are able to click search result links automatically, from here on referred to as clickers. These clickers act independently, without *PIA* noticing that these clicks do not come from real users. Each generated user has been modeled to have different preferences which have direct impact on their clicking behavior. In this section we describe how these users are modeled and how clicked documents lead to preferences for every clicker.

Each time a user clicks on a result in a result list, preferences are expressed. Assuming that these preferences are somehow expressible with a set of given features belonging to each result, it is possible to learn these preferences and rank results accordingly.

Using these features it is possible to generate clickers that model a typical *PIA* user. With regard to *PIA* and its scientific services background, feature based preferences are the following

clicker (1) At least one searched phrase has to be in the title. If there are search terms which do not belong to phrases, they are treated as phrase.
clicker (2) At least one search term has to be in the title, even if the search term is part of a phrase.
clicker (3) There have to be at least three authors.
clicker (4) Long summaries are preferred.
clicker (5) At least one searched phrase has to be in the summary.
clicker (6) By a chance of 0.5, a result is clicked or not.
clicker (7) Combined clicker of (2) AND (4) AND (5).

Because these preferences are feature based, rules can be written to express these preferences. If a user u prefers a result r then the probability that r is clicked is one:

$$p_u(r) = \begin{cases} 1, \; if\ user\ prefers\ result \\ 0, \; else \end{cases} \tag{3}$$

Treating *PIA* as a black box (see Fig. 2), each generated clicker performs a login (or registration if necessary) and starts querying *PIA* for results. For each query *PIA* returns a result list and the generated clicker starts clicking on preferred results.

4.3. Query Set

For the evaluation we defined a fixed list of 1000 queries. These queries were extracted from author given keywords of randomly chosen ACM documents. If a keyword included a space, it was transformed to a phrase query. Per document one or two randomly chosen keywords were transformed to an OR-query. The resulting queries look like: *“student experiments” teaching*. Each generated clicker used the same list of queries. We did not model query chains (Radlinski and Joachims, 2005) and each query is connected to a different information need.

4.4. Preference Precision

Because it is not of interest whether a result is relevant or not, standard information retrieval performance measures can not be taken into account.

Hence, the quality of the rankings is computed by the mean average preference precision, where the preference precision is 1 if a result is preferred or 0 if a result is not preferred (see Equ. 3). The computation of this measure follows the computation of the mean average precision. Note that the preference precision depends on the user and hence it is possible to evaluate a (personalized) result list with different preference precisions to conclude if the architecture learned to rank by the individual preferences.

5. Evaluation

For underlining the quality of the proposed approach, we investigate the resulting degree of personalization. Furthermore we evaluate the minimal amount of necessary feedback and the additional time to achieve such personalization.

5.1. Personalized Preference Learning

The first evaluation investigates whether personalized learning to rank can be applied successfully.

For this purpose clickers had to go through a set of 900 queries and click on preferred results. These clicks were recorded and used as training data for the RankSVM. After retrieving preference relations from these logs, the

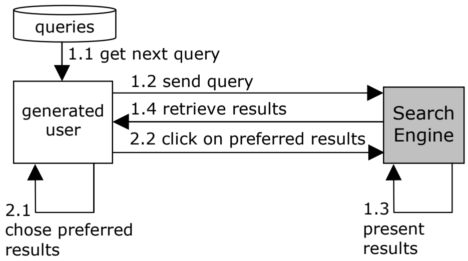


Figure 2: The generated user retrieves results from *PIA* and clicks on preferred results.

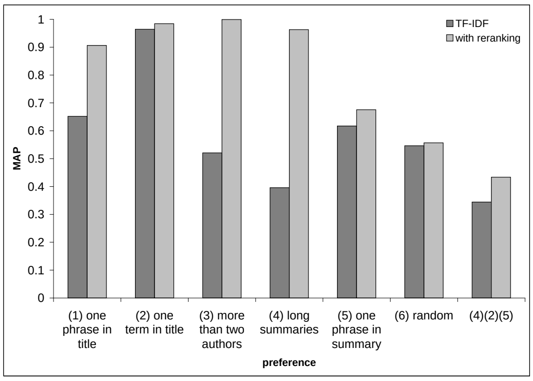


Figure 3: MAP values of results for selected preference clickers. The TF-IDF performance is based on the result lists before reranking was applied. Note that the precision does not depend on relevance but on the respective preference driven clicks.

RankSVM was trained and a personalized model was generated.

Then the clickers went through a second set of 100 queries to measure the degree of personalization of reranked results. These queries served to evaluate the difference between the original TF-IDF based ranking and reranked results.

The results of this evaluation are shown in Fig. 3. The bars in a darker shade represent the mean average precision of the results which were not reranked and are solely based on TF-IDF. Those bars in a lighter shade show the MAP of the results after the model was trained and the results were reranked.

It is quite clear to see that the MAP values with reranking exceed the MAP values solely based on TF-IDF. The results for the clickers "more than two authors" (**clicker (3)**) and "long summaries" (**clicker (4)**) look particularly promising as it is not possible for TF-IDF to map those preferences of these clickers.

As such reranking with RankSVM has increased the MAP by over 90% and 140%. The increase in MAP based on clickers "one phrase in title" (**clicker (1)**) and "one term in title" (**clicker (2)**) was 39% and 2%. The increase is not as high as those preferences can be mapped relatively well by TF-IDF. It was not possible to increase the MAP of a random clicker. For a more complex and realistic clicker who had multiple preferences the performance could be increased by 26%.

5.2. Amount of Feedback

The second evaluation we conducted was for investigating how much feedback is needed for personalization.

We address the question of how many queries with clicked results are needed before an effect on the reranking of the result list for a new user becomes apparent, and furthermore, after how many of such queries there is no more visible improvement of reranking. To evaluate this, we use sets of 100 randomly selected queries to be clicked by a "realistic" clicker with a complex preference model. **Clicker (7)** will click on any search result which has a non-empty description text, contains at least one phrase of the search term in this description and one word of the search term in its title. After each query, the average precision of the result clicks is calculated and stored, and the learning to rank model is retrained to include the newly obtained preferences. Queries in which no results have been clicked are discarded immediately and a different query string is substituted. After clicking on the results of 100 queries, 100 AP values are obtained, each accommodating the effects of reranking based on the preceding queries. To mitigate the variance of AP values for different search results, the AP values are averaged over 100 of such test runs. As baseline value, the result list before reranking was evaluated, displaying the search engine's native TF-IDF based ranking.

The results of the evaluation where the learning to rank model was retrained after each query can be seen in Fig. 4. This evaluation points to the question of when reranking affects the quality of results. The TF-IDF baseline which results from the evaluation of the result lists before reranking was done is more or less constant at a MAP of 0.4.

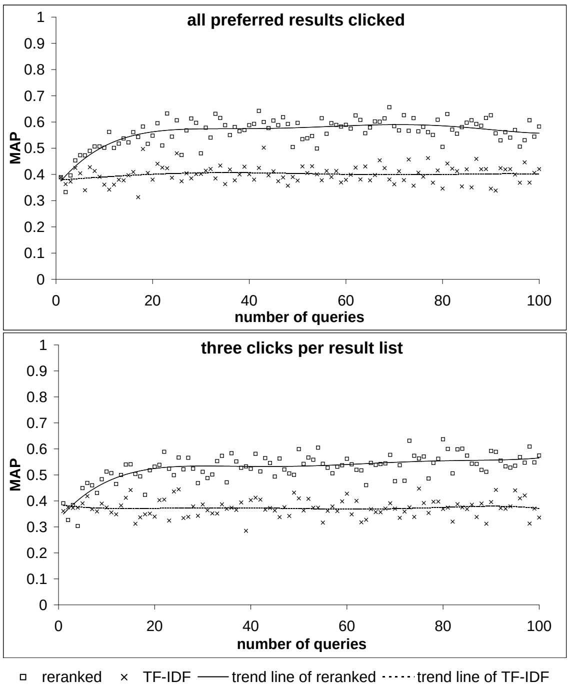


Figure 4: Comparison: MAP per query of TF-IDF and reranked results per query. The clicker clicked on all (top), or respectively on at most three (bottom), preferred results (with long summaries, at least one search term in the title, and at least one searched phrase in the summary). After each query there was a new training of the model. Data points are the mean of 100 test runs and the polynomial trend line has the order 5.

On top of Fig. 4 an evaluation can be seen, where **clicker (7)** clicked on all preferred results. On the bottom the clicker only clicked on the first three preferred results. Nevertheless the presented MAP (mean of 100 test runs) value is based on the preference of each result.

An increasing effect of reranking on the search result lists with every query can be discerned in both evaluations. After about 10 search queries that included result clicks, the MAP of reranked results increased by 0.1 compared to the TF-IDF based ranking presented by the search engine without reranking. This means that personalized reranking of results for a user has a visibly increasing effect on the search results presented to the user, as useful results had been reranked to appear higher up in the list.

For a user which intensively browses 30 result lists, the following reranked results have a MAP of about 0.55 which, compared to the MAP of about 0.4 of the native result ranking, is a 37% increase. Result lists for users which generate only a little bit of implicit feedback have a MAP of 0.5 after 30 queries, an increase of 25%. There is a slight increase in MAP for these users after query 30.

Note that these results do not point to the fact that the reranked results always perform better. There were result list where the TF-IDF ranking had a better AP. Additionally these MAP values strongly depend on **clicker (7)** which represents a user with several preferences.

After 30 queries, no further significant improvement of ranking quality can be observed, though in this test the

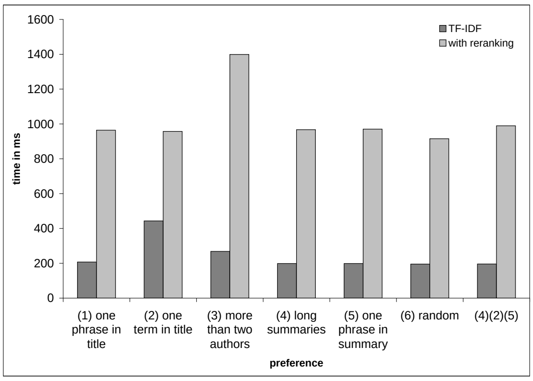


Figure 5: Time comparison between unranked and reranked results. The time for reranking includes the querying of the index, the feature vector retrieval, and the reranking of the first 30 results.

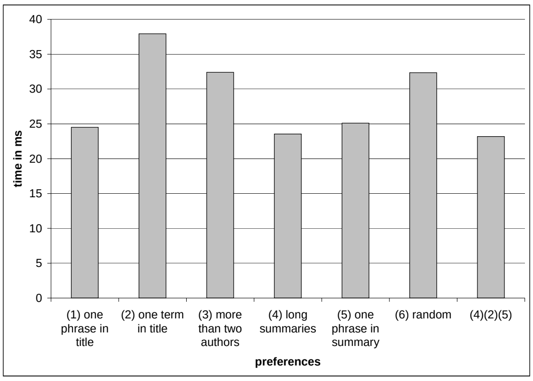


Figure 6: Mean training time for one preference.

user's criteria that determine which results are clicked are invariable. It can be concluded that with an applied learning to rank method, distinctly personalized results for users with a non-random click-behavior can be achieved after only 10 trained queries, in which returned results were clicked by that user.

5.3. Additional Response Time

Learning to rank is not just based on the learning of preference data. This data has to be retrieved first. Result clicks have to be logged, transferred to preferences, and the results have to be converted to feature vectors before they can be trained. Hence there are three parts in a search engine which have to be modified or added to apply learning to rank: logging, learning, and reranking. The additional time needed is logged for evaluation.

As learning to rank involves an additional reranking step, we have measured the extra time needed to do so. In Fig. 5 it is quite clearly visible that the overhead of reranking is significant. The average time increase is at over 350%. In average the user had to wait about 1020ms until the results were presented.

In Fig. 6 the average time for training a single preference is presented. The mean training for a single preference ranges from 24.5ms to 32.2ms and averages in those clickers at around 28.4ms.

6. Conclusion

In this work we applied learning to rank to a real search engine. We investigated whether it is possible to use learning to rank for ranking search results by personal preferences. The evaluations have clearly shown that personalization is possible, even if the personal preferences are complex.

Applying learning to rank to search engines is possible, but unfortunately for the cost of an extended query response time. In average users of our architecture have to wait one second longer if the first 30 results are reranked by their individual preferences.

While ranking by preferences which are based on only one feature is easily be done, it is more difficult with a more complex preference scheme.

We strongly suggest preprocessing quality features of potential results offline and storing these attributes in an index. Such quality information is needed when creating a representing feature vector for each result and querying an index for this information can be done fast. This especially becomes necessary at query time where response time restrictions have to be followed.

A learning to rank search engine including the proposed preference retrieval method automatically scales by becoming better. If a search engine returns perfectly ranked results, it is not possible to extract preferences from clickthrough data and hence no recreation of the used model will be performed.

If there is a point of no real quality improvement when training a learning to rank model with additional preferences, there should be a general limitation to the number of used preferences in such an application. Additionally preferences might change and hence it is necessary nevertheless to dismiss old click logs.

The training of a learning to rank model should be done shortly after the user interaction which led to new preferences. The training time for creating a model for one user never exceeded one minute (on a standard system) and can be done in parallel to the interaction with the user. When the model is created, the old model can be replaced and the user benefits from learning to rank immediately.

7. Future Work

Possible extension to the presented approach is the implementation of explicit feedback retrieval. In addition to the extracted preference pairs from clickthrough data, users will be able to actively express their preference between two results and thus improve their personal ranking performance.

Another function to implement could be the use of query chains as presented in (Radlinski and Joachims, 2005). This would make it possible to exploit the fact that a lot of queries are interdependent, and thus improve the quality of search results for successive queries.

In our work, we use clickers with fixed probabilities of 1 to click on a preferred document and 0 to click on a random document. To produce a more realistic user model, a probabilistic click behavior that causes clickers to skip some of the preferred documents and to select some of the non-preferred documents will be included.

Since users' preferences may change over time, their result click pattern can vary. The existing clickers that combine different preference models can be extended to model this effect. Using these clickers the system can be monitored to reveal how long it takes for a trained model to adapt to changed likings.

In future work, we will introduce additional user-dependent features, like *user has read result* or *result was presented to user in an earlier search*. We are confident that such features will further improve users' search experience.

8. References

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 129–136, New York, NY, USA. ACM.

Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. 2003. An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.*, 4:933–969.

John Guiver and Edward Snelson. 2008. Learning to rank with softrank and gaussian processes. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 259–266, New York, NY, USA. ACM.

R. Herbrich, T. Graepel, and K. Obermayer. 2000. Large margin rank boundaries for ordinal regression. In *Advances in Large Margin Classifiers*, pages 115–132, Cambridge, MA. MIT Press.

Thorsten Joachims and Filip Radlinski. 2007. Search engines that learn from implicit feedback. *Computer*, 40(8):34–40.

Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 133–142.

Tie-Yan Liu, Jun Xu, Tao Qin, Wenying Xiong, and Hang Li. 2007. Letor: Benchmark dataset for research on learning to rank for information retrieval. In *LR4IR 2007, in conjunction with SIGIR 2007*.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

F. Radlinski and T. Joachims. 2005. Query chains: Learning to rank from implicit feedback. In *ACM SIGKDD International Conference On Knowledge Discovery and Data Mining (KDD)*, pages 239–248.

Ming-Feng Tsai, Tie-Yan Liu, Tao Qin, Hsin-Hsi Chen, and Wei-Ying Ma. 2007. Frank: a ranking method with fidelity loss. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 383–390, New York, NY, USA. ACM.

Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd edition edition.

Min Zhang, Da Kuang, Guichun Hua, Yiqun Liu, and Shaoping Ma. 2009. Is learning to rank effective for Web search? In *SIGIR2009 workshop: Learning to Rank for Information Retrieval (LR4IR 2009)*.

APPENDIX

In the following we present a list of features used.

Quality-based features:

1. Publication-related features
   - *Number of authors*
   - *Year of publication*
   - *Number of words in abstract*
   - *PageRank*
   - *Cite-impact*
   - *Number of references to other publications*
   - *Number of publications of main author*
   - *Number of publications of all authors*
2. Document file-related features
   - *Document page count*
   - *Number of words in publication*
   - *Document file size*

Query-dependent features:

1. Statistical query-dependent features:
   - *Term frequency* (TF), *Inverse document frequency* (IDF), *TF-IDF* Product of TF and IDF.
2. Query reformulation query-dependent features:[3]
   - *First word fuzzy:* Searches with fuzziness constraint for the first word. Three fuzziness levels with values of 0.5, 0.8 or 0.9 lead to three features.
   - *Last word fuzzy:* Searches with fuzziness constraint for the last word. Three fuzziness levels with values of 0.5, 0.8 and 0.9 lead to three features.
   - *First word only:* Searches for the first word of the query phrase.
   - *Fuzzy only first word:* Searches with fuzziness constraint for the first word from the query phrase only. Three fuzziness levels with values of 0.5, 0.8 and 0.9 lead to three features.
   - *Last word only:* Searches for the last word of the query phrase.
   - *Fuzzy only last word:* Searches with fuzziness constraint for the last word from the query phrase only. Three fuzziness levels with values of 0.5, 0.8 and 0.9 lead to three features.
   - *All words fuzzy:* Apply fuzziness constraint for all words from the query phrase. Three fuzziness levels with values of 0.5, 0.8 and 0.9 lead to three features.
   - *Words within range:* Looks whether all words from the query phrase are located within a distance of two, ten or number of words in the query. Leads to three features.
   - *Boosting first word:* Apply higher priority to the first word of the query phrase. Boosts by five or ten leading to two features.
   - *Boosting last word:* Apply higher priority to the last word of the query phrase. Boosts by five or ten leading to two features.
   - *Boosting gradient:* Increases or decreases the priority beginning with the first word to the last word. Leads to two features (increasing or decreasing priority).
   - *Fuzzy gradient:* Increases or decreases the fuzziness beginning with the first word to the last word. Leads to two features (increasing or decreasing fuzziness).
   - *Terms as phrase:* All words from the query phrase are treated as a single phrase.

The complete feature vector consists of 64 single features.

[3] Each query dependent feature is computed for all known document parts, like title or summary separately.

Personalized Multi-Document Summarization using N-Gram Topic Model Fusion

Leonhard Hennig, Sahin Albayrak

DAI Labor, TU Berlin
Berlin, Germany
firstname.lastname@dai-labor.de

Abstract

We consider the problem of probabilistic topic modeling for query-focused multi-document summarization. Rather than modeling topics as distributions over a vocabulary of terms, we extend the probabilistic latent semantic analysis (PLSA) approach with a bigram language model. This allows us to relax the conditional independence assumption between words made by standard topic models. We present a unified topic model which evolves from sentence-term and sentence-bigram co-occurrences in parallel. Sentences and queries are represented as probability distributions over latent topics to compute thematic and query-focused sentence features in the topic space. We find that the inclusion of bigrams improves the descriptive quality of the latent topics, and leads to a substantially reduces the number of latent topics required for representing document content. Experimental results on DUC 2007 data show an improved performance compared to a standard term-based topic model. We further find that our method performs at the level of current state-of-the art summarizers, while being build on a considerably simpler model than previous topic modeling approaches to summarization.

1. Introduction

Automatically producing summaries from a set of input documents is an extensively studied problem in IR and NLP (Jones, 2007). In this paper, we investigate the problem of multi-document summarization, where the task is to "synthesize from a set of related documents a well-organized, fluent answer to a complex question"[1]. In particular, we focus on generating an extractive summary by selecting relevant sentences from a document cluster (Goldstein et al., 2000). The condensation of information from different sources into an informative summary is an increasingly important task, since it helps to reduce information overload.

A major challenge of identifying relevant information is to model document content. A document will generally contain a variety of information centered around a main topic, and covering different aspects (subtopics) of this main theme (Barzilay and Lee, 2004). Human summaries also tend to cover different aspects of the original source text to increase the informative content of the summary. In addition, in query-focused multi-document summarization tasks, the user query often explicitly requests information about different aspects of the main theme of the document cluster (see Table 1). An ideal summary should therefore aim to include information for each of the "subquestions" of the complex user query.

Various summarization approaches have exploited observable features based on the identification of topics (or thematic foci) to construct summaries. Often, such features rely on the identification of important keywords (Yih et al., 2007; Nenkova et al., 2006), or on the creation of term-based topic signatures (Lin and Hovy, 2000; Conroy et al., 2007). However, it is well known that term matching has severe drawbacks due to the ambivalence of words and to differences in word usage across authors (Manning and Schütze, 2001). This is especially important for automatic summarization, as summaries produced by humans may differ significantly in terms of word usage. (Lin and Hovy, 2003b).

Topic models such as Latent Semantic Indexing (LSI) (Deerwester et al., 1990), Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 1999), and Latent Dirichlet Allocation (LDA) (Blei et al., 2003) provide a means to overcome the problem of term matching, and furthermore allow for the modeling of inter- and intradocument statistical structure. These models introduce hidden variables as the origin of the observed term co-occurrences. Whereas LSI is a mapping of the original term-document vector space onto a linear subspace based on singular value decomposition, PLSA and LDA model documents as a distribution of mixture components, where each mixture component is a multinomial distribution over words. The mixture components can be interpreted as "topics", and the corresponding word distribution characterizes the semantics of the topic.

This reduced description will generally capture some aspects of synonymy and polysemy, since words with similar meanings tend to occur in similar contexts. Furthermore, semantically similar words are clustered based on the assumption that the co-occurrence of terms signals semantic relatedness. However, words are considered independent given the topics, resulting in the standard bag-of-words assumption (Blei et al., 2003). N-Gram language models (Ponte and Croft, 1998) allow us to relax this assumption in order to capture multi-word concepts, where word order plays a critical role (Wang et al., 2007).

1.1. Our contribution

Our approach extends the standard topic modeling approach such that the topic model is estimated from the term co-occurrence as well as bigram co-occurrence observations in parallel. The integration of a bigram language model allows us to represent the mixture compo-

[1] DUC summarization track task description, `http://www.nist.gov/tac`

Table 1: A complex user query from DUC 2006.

ID	D0631D
Title	Crash of the Air France Concorde
Query	Discuss the Concorde jet, its crash in 2000, and aftermaths of this crash.

nents as multinomial distributions over terms and bigrams, leading to an improved representation of the components. Each document's distribution over the mixture components is re-estimated based on maximizing the likelihood of the data given both the term co-occurrence and the bigram co-occurrence distributions.

Furthermore, the integration of the bigram language model allows us to relax the (conditional) independence assumption made by the standard topic model, since bigrams encode syntactic dependencies between consecutive terms. Even though one can consider a bigram simply to be a co-occurrence of two terms, and as such captured well enough by a standard topic model, our assumption is that bigram co-occurrence patterns will reinforce the observed term co-occurrence patterns. We show that this results in more descriptive latent topics, and considerably reduces the number of latent topics required for a good model.

We apply the modified topic model in query-focused multi-document summarization, and model sentences and queries in this novel latent topic space. This allows us to compute thematic and query-focused sentence similarity features for extractive summarization.

The rest of this paper is structured as follows: We start with an overview of related work in Section 2.. In Section 3. we describe our approach for integrating a language model into the PLSA algorithm. Next, in Section 4., we give details of our summarization system, the sentence-level features we use, and of our sentence ranking and selection approach. In Section 5., we describe and analyze the data sets we use to verify the assumptions of our approach, and we present experimental results. Finally, Section 6. concludes the paper.

2. Related work

Probabilistic topic models for the representation of document content have also been explored by Barzilay and Lee (Barzilay and Lee, 2004). They use Hidden Markov Models (HMM) to model topics and topic change in text, albeit only for generic single-document summarization. The model assumes that a topic is formed by clustering sentences based on vector space similarity, and bigram distribution patterns are learned from these topical clusters. Each sentence is assigned to exactly one topic cluster, corresponding to a HMM state. Documents are modeled as sequences of topics, representing the typical discourse structuring of texts in specific domains. In contrast, our approach models each sentence as a distribution over multiple topics, and also models queries in the latent topic space for query-focused multi-document summarization.

More related to our method is the approach of Daumé and Marcu (Daumé and Marcu, 2006), who utilize a model similar in style to LDA. However, the latent classes are chosen to capture general language background vocabulary, document- and query-specific vocabularies. Each sentence is modeled as a distribution over these three mixture components, e.g. consisting of 60% query information, 30% background document information, and 10% general English (Daumé and Marcu, 2006). Topical information is not considered, and neither are the subtopics contained in a document.

The method proposed by Haghighi and Vanderwende takes up this approach, but constructs summaries by optimizing the KL-divergence between the summary topic distribution and the topic distribution of the source document set (Haghighi and Vanderwende, 2009). Subtopics are modeled by introducing a hierarchical LDA process. Instead of drawing words only from a generic "content" distribution they allow for either generic or topic-specific word distributions for each sentence. However, for each sentence only one distribution is selected, and all content words of that sentence are drawn from this distribution. Topic-specific distributions are ordered sequentially over sentences similar to the approach of Barzilay and Lee. The proposed approach does not allow for query-focused summarization.

In previous work, we showed that a term-sentence co-occurrence based PLSA model can be effectively used for query-focused multi-document summarization (Hennig, 2009). The proposed model outperformed existing systems on DUC 2006 data, and performed comparable to state-of-the-art summarization systems on the DUC 2007 dataset.

All of the above methods consider either unigram or bigram distributions for representing topics, but not the combination of both. In our approach, we combine unigram and bigram observations to create topic representations that consist of multinomial distributions over both unigrams and bigrams.

In the area of topic modeling, Wallach proposed an approach to relax the bag-of-words assumption in (Wallach, 2006). The LDA model she discusses incorporates, in a fashion similar to typical n-gram language models, the conditional probability of a word at position t given the word at position $t-1$, such that $p(w_t) = p(w_t|w_{t-1})$. Each topic is represented as a set of W distributions – contrasting with the single word distribution per topic typically used – where W is the size of the vocabulary. Each of the W word distributions per topic is conditioned on the context of a previous word w_{t-1}. The number of parameters to be estimated is hence $WT(W-1)$, whereas our model has $(W-1)T(B-1)$ free parameters (B is the number of distinct bigrams).

3. Topic and Language Model Combination using PLSA

For simplicity, we utilize and adapt the PLSA algorithm to test the validity of our approach, but for all purposes this can be considered equivalent to using a more complex topic model such as LDA.

PLSA is a latent variable model for co-occurrence data that associates an unobserved class variable $z \in \mathcal{Z} = \{z_1, \ldots, z_k\}$ with each observation (d, w), where word $w \in \mathcal{W} = \{w_1, \ldots, w_i\}$ occurs in document $d \in \mathcal{D} = \{d_1, \ldots, d_j\}$. Documents and words are assumed independent given the topic variable Z. The probability that a word occurs in a document can be calculated by summing over all latent variables Z:

$$P(w_i|d_j) \quad = \quad \sum_k P(w_i|z_k)P(z_k|d_j). \qquad (1)$$

Similarly, we can associate each observation (d, b) of a bigram $b = (ww')$, where bigram $b \in \mathcal{B} = \{b_1, \ldots, b_l\}$ occurs in document d, with the same unobserved class variable z. We assume the same hidden topics of the term-sentence co-occurrences (d, w) as the origin of the bigram-sentence co-occurrence observations (d, b):

$$P(b_l|d_j) \quad = \quad \sum_k P(b_l|z_k)P(z_k|d_j). \qquad (2)$$

Notice that both decompositions share the same document-specific mixing proportions $P(z_k|d_j)$. This couples the conditional probabilities for terms and bigrams: each "topic" has some probability $P(b_l|z_k)$ of generating bigram b_l as well as some probability $P(w_i|z_k)$ of generating an occurrence of term w_i. The advantage of this joint modeling approach is that it integrates term and bigram information in a principled manner. This coupling allows the model to take evidence about bigram co-occurrences into account when making predictions about terms and vice versa. Following the procedure in Cohn and Hofmann (Cohn and Hofmann, 2000), we can now combine both models based on the common factor $P(z|d)$ by maximizing the log-likelihood function

$$\begin{aligned} L \quad = \quad & \sum_j \Bigg[\alpha \sum_i n(d_j, w_i) \log P(w_i|d_j) \\ & + (1-\alpha) \sum_l n(d_j, b_l) \log P(b_l|d_j) \Bigg] \qquad (3) \end{aligned}$$

where α is a predefined weight for the influence of each two-mode model. Using the Expectation-Maximization (EM) algorithm we then perform maximum likelihood parameter estimation for the aspect model. During the expectation (E) step we first calculate the posterior probabilities:

$$\begin{aligned} P(z_k|w_i, d_j) \quad &= \quad \frac{P(w_i|z_k)P(z_k|d_j)}{P(w_i|d_j)} \qquad (4) \\ P(z_k|b_l, d_j) \quad &= \quad \frac{P(b_l|z_k)P(z_k|d_j)}{P(b_l|d_j)}, \qquad (5) \end{aligned}$$

and then re-estimate parameters in the maximization (M) step as follows:

$$\begin{aligned} P(w_i|z_k) \quad &= \quad \sum_j \frac{n(w_i, d_j)}{\sum_{i'} n(w_{i'}, d_j)} P(z_k|w_i, d_j) \quad (6) \\ P(b_l|z_k) \quad &= \quad \sum_j \frac{n(b_l, d_j)}{\sum_{l'} n(b_{l'}, d_j)} P(z_k|b_l, d_j) \quad (7) \end{aligned}$$

The class-conditional distributions are recomputed in the M-step as

$$\begin{aligned} p(z_k|d_j) \qquad & \propto \alpha \textstyle\sum_i \frac{n(w_i,d_j)}{\sum_{i'} n(w_{i'},d_j)} P(z_k|w_i, d_j) \\ & + (1-\alpha) \textstyle\sum_l \frac{n(b_l,d_j)}{\sum_{l'} n(b_{l'},d_j)} P(z_k|b_l, d_j) \quad (8) \end{aligned}$$

Based on the iterative computation of the above E and M steps, the EM algorithm monotonically increases the likelihood of the combined model on the observed data. Using the α parameter, our new model can be easily reduced to a term co-occurrence based model by setting $\alpha = 1.0$.

4. Topic-based summarization

Our approach for producing a summary consists of three steps: First, we represent sentences and queries in the latent topic space of the combined PLSA model by estimating their mixing proportions $P(z|d)$. We then compute several sentence-level features based on the similarity of sentence and query distributions over latent topics. Finally, we combine individual feature scores linearly into an overall sentence score to create a ranking, which we use to extract sentences for the summary. We follow a greedy approach for selecting sentences, and penalize candidate sentences based on their similarity to the partial summary. These steps are described in detail below.

4.1. Data Set

We conduct our analysis and evaluate our model based on the multi-document summarization data sets provided by DUC[2]. Specifically, we use the DUC 2007 data set for evaluation. The data set consists of 45 document clusters, with each cluster containing 25 news articles related to the same general topic. Participants are asked to generate summaries of at most 250 words for each cluster. For each cluster, a title and a narrative describing a user's information need are provided. The narrative (query) is usually composed of a set of questions or a multi-sentence task description.

4.2. Sentence representation in the latent topic space

Given a corpus $\mathcal{D}$ of topic-related documents, we perform sentence splitting on each document using the NLTK toolkit[3]. Each sentence is represented as a bag-of-words $\mathbf{w} = (w_1, \ldots, w_m)$. We remove stop words for the unigram model, and apply stemming using Porter's stemmer (Porter, 1980). We create a term-sentence matrix TS containing all sentences of the corpus, where each entry $TS(i, j)$ is given by the frequency of term i in sentence j, and a bigram-sentence matrix BS, where each entry $BS(l, j)$ is given by the frequency of bigram l in sentence j. We then train the combined PLSA model on the matrices TS and BS.

After the model has been trained, it provides a representation of the sentences as probability distributions $P(z|s)$ over the latent topics Z, and we arrive at a representation of sentences as a vector in the "topic space":

$$\mathbf{s} = (p(z_1|s), p(z_2|s), \ldots, p(z_K|s)), \qquad (9)$$

[2] http://www.nist.gov/tac

[3] http://www.nltk.org

where $p(z_k|s)$ is the conditional probability of topic k given the sentence s.

In order to produce a query-focused summary, we also need to represent the query in the latent topic space. This is achieved by folding the query into the trained model. The folding is performed by EM iterations, where the factors $P(w|z)$ and $P(b|z)$ are kept fixed, and only the mixing proportions $P(z|q)$ are adapted in each M-step (Hofmann, 1999). We fold the title and the query of the document clusters, the document titles, and document and cluster vectors into the trained PLSA model. Query vectors are preprocessed in the same way as training sentences, except that no sentence splitting is performed. Document and document cluster term vectors are computed by aggregating sentence term vectors.

4.3. Computing query- and topic-focused sentence features

Since we are interested in producing a summary that covers the main topics of a document set and is also focused on satisfying a user's information need, specified by a query, we create sentence-level features that attempt to capture these different aspects in the form of per-sentence scores. We then combine the feature scores to arrive at an overall sentence score. Each feature is defined as a similarity $r(S,Q)$ of a sentence topic distribution $S = P(z|s)$ compared to a "query" topic distribution $Q = P(z|q)$:

- $r(S, CT)$ - similarity to the cluster title
- $r(S, N)$ - similarity to the cluster narrative (query)
- $r(S, T)$ - similarity to the document title
- $r(S, D)$ - similarity to the document centroid
- $r(S, C)$ - similarity to the cluster centroid

Since measures for comparing two probability distributions are typically defined as divergences, not similarities, we invert the computed divergence. In our approach, we employ the Jensen-Shannon (JS) divergence, but a variety of other similarity measures can be utilized towards this end. The JS-divergence is a symmetrized and smoothed version of the Kullback-Leibler divergence:

$$\begin{aligned} r_{JS}(S,Q) &= 1 - \Bigg[\frac{1}{2}D_{KL}(S||M) \\ &\quad + \frac{1}{2}D_{KL}(Q||M)\Bigg], \end{aligned} \tag{10}$$

where $M = 1/2(S+Q)$.

As the training of a PLSA model using the EM algorithm with random initialization converges on a local maximum of the likelihood of the observed data, different initializations will result in different locally optimal models. As Brants et al. (Brants et al., 2002) have shown, the effect of different initializations can be reduced by generating several PLSA models, then computing features according to the different models, and finally averaging the feature values. We have implemented this model averaging using 5 iterations of training the PLSA model.

4.4. Sentence scoring

The system described so far assigns a vector of feature values to each sentence. The overall score of a sentence consisting of the features $(r_1, \ldots, r_P)$ is then defined as:

$$score(s) = \sum_p w_p r_p, \tag{11}$$

where w_p is a feature-specific weight. We optimized the features weights on the DUC 2006 data set, which is similar to our evaluation data set. We initialized all feature weights to a default value of 1, and then optimized one feature weight at a time while keeping the others fixed. The most dominant features in our experiments are the sentence-narrative similarity $r(S, N)$ and the sentence-document similarity $r(S, D)$, which confirms previous research. Sentences are ranked by this score, and the highest-scoring sentences are selected for the summary.

We model redundancy similar to the maximum marginal relevance framework (MMR) (Carbonell and Goldstein, 1998). MMR is a greedy approach that iteratively selects the best-scoring sentence for the summary, and then updates sentence scores by computing a penalty based on the similarity of each sentence with the current summary:

$$score_{mmr}(s) = \lambda(score(s)) - (1-\lambda)r(S, SUM), \tag{12}$$

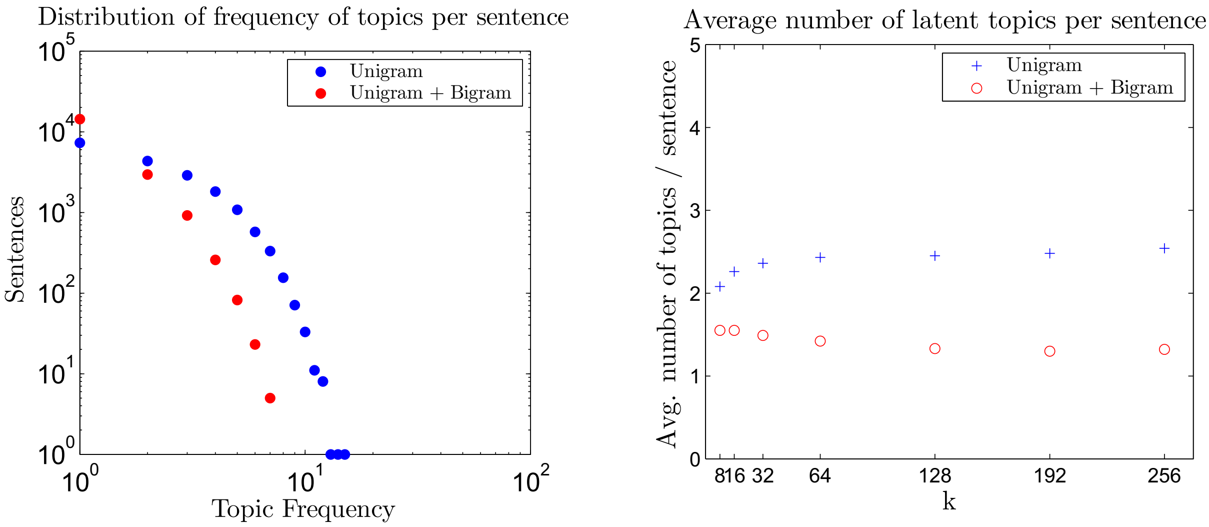
where the score of sentence s is scaled to $[0, 1]$ and $r(S, SUM)$ is the cosine similarity of the sentence and the summary centroid vector, which is based on the averaged distribution over topics of sentences selected for the summary. We optimized λ on DUC 2006 data, with the best value $\lambda = 0.4$ used in our experimental evaluation.

4.5. Topic distribution over sentences

It is well known that documents cover multiple subtopics related to the main theme of the document (Barzilay and Lee, 2004). Standard topic models such as LDA therefore represent a document as a distribution over a set of latent topics. In our approach, we extend this notion and treat each sentence as a document, thus assuming that a sentence covers one or more topics of the document set. For example, a sentence of a news article related to a meeting of government leaders may provide information on the people who have met as well as on the location of the meeting. Our intuition is that the number of topics that a sentence covers should be rather low, but greater than one.

Figure 1(a) shows the distribution of the number of topics per sentence for a PLSA model based on terms only and for the PLSA model combining unigrams and bigrams. We only consider topics with a probability greater than some small value ϵ ($\epsilon > 0.01$). We see that the distributions follow a power law: There are very many sentences which are assigned a single dominant topic, and very few sentences which are assigned many topics. We note that the combined model assigns less topics to a sentence than the term-based model.

From Figure 1(b) we see that the average number of topics assigned to a sentence is relatively robust to varying the value of k (the free parameter specifying the number of latent topics for the PLSA algorithm). Even for $k <= 16$,



(a) DUC 2007: Topic distribution

(b) DUC 2007: Average number of topics

Figure 1: (a) Distribution of number of topics per sentence ($p(z|d) > 0.01$) for a $k = 128$ factor decomposition of the DUC 2007 document sets, using terms only or the combined model; and (b) Average number of topics per sentence ($p(z|d) > 0.01$) for different values of k, using terms only or the combined model

where k is actually smaller than the number of input documents, on average more than one topic is assigned to a sentence. This confirms our intuition that sentences may cover multiple subtopics. Again we see that the combined model on average assigns less topics to a sentence, which suggests that the descriptive quality of the topics better fits the available data.

5. Experiments

For the evaluation of our system, we use the data set from the multi-document summarization task in DUC 2007. For all our evaluations, we use ROUGE metrics[4]. ROUGE metrics are recall-oriented and based on n-gram overlap. ROUGE-1 has been shown to correlate well with human judgements (Lin and Hovy, 2003a). In addition, we also report the performance on ROUGE-2 (bigram overlap) and ROUGE-SU4 (skip bigram) metrics.

5.1. Results

We present the results of our system in Table 2. We compare our results to the best peer (*peer 15*) and to a *Lead* sentence baseline system. The *Lead* system uses the first n sentences from the most recent news article in the document cluster to create a summary. In the table, system *PLSA* uses a standard term co-occurrence based model, and system *PLSA-F* combines term and bigram co-occurrences, based on the best value for parameter $\alpha = 0.6$. The *PLSA-F* system outperforms the standard PLSA model on ROUGE-1, ROUGE-2 and ROUGE-SU4 scores, although the improvements are not significant. More interestingly, the *PLSA-F* achieves its best score using only $k = 32$ latent classes, compared to $k = 256$ for the *PLSA* system. This suggests that the information supplied by the bigram co-occurrence observations indeed reinforces the term co-occurrence observations, such that the model can better represent the different latent topics contained in the document cluster.

Table 2: DUC-07: ROUGE recall scores for best number of latent topics k. The *PLSA* system uses term co-occurrences only, the *PLSA-F* combines term and bigram co-occurrence information, with $\alpha = 0.6$. The *PLSA-F* variant outperforms the best participating system (peer 15) on ROUGE-1.

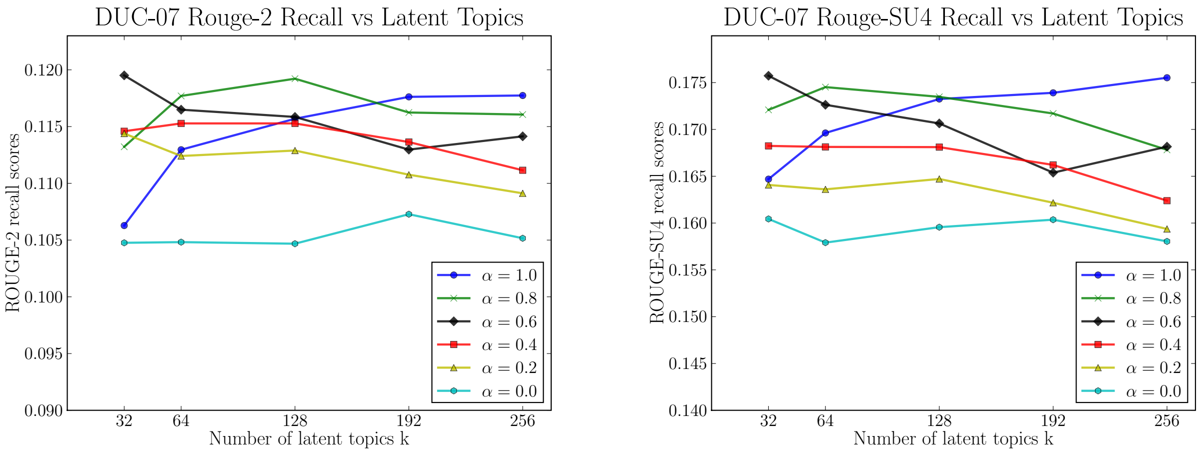
System	k	Rouge-1	Rouge-2	Rouge-SU4
peer 15	-	0.44508	0.12448	0.17711
PLSA-F	32	**0.45400**	0.11951	0.17573
PLSA	256	0.44885	0.11774	0.17552
Lead	-	0.31250	0.06039	0.10507

Our combined approach outperforms *peer 15* on ROUGE-1 recall, and is not significantly worse on ROUGE-SU4 recall. For ROUGE-2, our system's performance is only slightly lower than the 95%-confidence interval of the top system's performance (0.11961–0.12925). The results of our system are also comparable to the topic modeling approach of Haghighi and Vanderwende (Haghighi and Vanderwende, 2009), who report a ROUGE-2 score of 0.118 for a model based on bigram distributions, but are significantly better than the 0.097 they report for an unigram-based model.

5.2. System variations

To verify the experimental observation that the combined model allows for a better representation of the latent topics, we conducted a series of experiments varying the number of latent classes and the weight of the parameter α. The results of these experiments are shown in Figure 2. We have

[4] ROUGE version 1.5.5, with arguments -n 4 -w 1.2 -m -2 4 -u -c 95 -r 1000 -f A -p 0.5 -t 0



(a) DUC 2007 Rouge-2

(b) DUC 2007 Rouge-SU4

Figure 2: Summarization performance on DUC 2007 data in terms of ROUGE-2 (a) and ROUGE-SU4 (b) recall for different values of latent topics k and parameter α.

omitted results for $k < 32$, where none of the models can cope with the complexity of the data. We also do not show results for $k > 256$, since the performance of all models either stabilizes with respect to their performance at $k = 256$, or the models start to overfit, resulting in lower ROUGE scores. We observe that the models combining term and bigram co-occurrence information outperform the models based only on term co-occurrence ($\alpha = 1.0$) respectively bigram co-occurrence ($\alpha = 0.0$) for small numbers of latent classes k. As k increases, the performance of the combined models decreases, or exhibits only small variations (e.g. $\alpha = 0.6$ for $k = 256$). This suggests that the quality of the learned latent topics is starting to decrease, as the algorithm creates topics with idiosyncratic word combinations (Steyvers and Griffiths, 2006). The performance of the term-based model, however, increases until $k = 256$, reaching a maximum ROUGE-2 recall of 0.11776, before also overfitting (not shown here).

Our observations therefore indicate that the information obtained from the combined model allows for a more descriptive representation of the latent topics contained in the document collection.The most interesting observation shown in Figure 2 is that adding bigram-sentence co-occurrence observations to a standard PLSA model can substantially improve ROUGE-2 scores and significantly reduce the number of latent classes required for a good model. The effect is less pronounced for ROUGE-SU4 scores, but still recognizable. All combined models outperform the term and bigram baseline models on ROUGE-2 for $k = 32$ latent classes.

We further note that the term-based model ($\alpha = 1.0$) consistently outperforms the bigram-based model ($\alpha = 0.0$), indicating that bigram co-occurrence information alone captures less of the topical relations that exist in the document collection.

We also find that the effect of varying the parameter α follows an expectable pattern: For $\alpha = 0.8$, the term-based model dominates the latent topic model, and the ROUGE-2 score curve follows that of the model with $\alpha = 1.0$ (for $k <= 128$). The experimentally optimal value of $\alpha = 0.6$ weights term and bigram co-occurrences almost equally, with ROUGE-2 scores similar for $\alpha = 0.4$. For lower values of α, the summarization performance of the model decreases substantially, ROUGE-SU4 scores are constantly lower than for the other models.

6. Conclusion

We introduced a novel approach to query-focused multi-document summarization that combines term and bigram co-occurrence observations into a single probabilistic latent topic model. The integration of a bigram language model into a standard topic model results in a system that outperforms models which are based on term respectively bigram co-occurrence observations only. Furthermore, it requires fewer latent classes for optimal summarization performance.

We observe that the distribution of topic frequencies across sentences follows a power law. On average, sentences are assigned more than two latent topics for a standard topic model, but only between one and two topics for our combined model. This suggests that the combined model results in a better representation of the underlying subtopics of a document set. We also find that the average number of topics assigned to a sentence is relatively robust with respect to variations in the number of latent classes.

Our results are among the best reported on the DUC-2007 multi-document summarization tasks for ROUGE-1, ROUGE-2 and ROUGE-SU4 scores. We have achieved these excellent results with a system that utilizes a considerably simpler model than previous topic modeling approaches to multi-document summarization.

In future work, we plan to implement our approach using LDA instead of PLSA to address shortcomings of PLSA such as overfitting and the lack of generative modeling at the document level.

7. References

Regina Barzilay and Lillian Lee. 2004. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proc. of HLT-NAACL*.

David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. 2003. Latent dirichlet allocation. *J. of Machine Learning Research*, 3:2003.

Thorsten Brants, Francine Chen, and Ioannis Tsochantaridis. 2002. Topic-based document segmentation with probabilistic latent semantic analysis. In *Proc. of CIKM*, pages 211–218.

Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proc. of SIGIR '98*, pages 335–336.

David Cohn and Thomas Hofmann. 2000. The missing link - a probabilistic model of document content and hypertext connectivity. In *Neural Information Processing Systems 13*, pages 430–436.

J. M. Conroy, J. D. Schlesinger, and D.P. Leary. 2007. CLASSY 2007 at DUC 2007. In *Proc. of DUC 2007*.

Hal Daumé, III and Daniel Marcu. 2006. Bayesian query-focused summarization. In *Proc. Int. Conf. on Computational Linguistics (ACL)*, pages 305–312.

Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *J. of the American Society for Information Science*, 41:391–407.

Jade Goldstein, Vibhu Mittal, Jaime Carbonell, and Mark Kantrowitz. 2000. Multi-document summarization by sentence extraction. In *NAACL-ANLP 2000 Workshop on Automatic summarization*, pages 40–48.

Aria Haghighi and Lucy Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proc. of NAACL-HLT*.

Leonhard Hennig. 2009. Topic-based multi-document summarization with probabilistic latent semantic analysis. In *Recent Advances in Natural Language Processing, RANLP 2009*.

Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proc. of SIGIR '99*, pages 50–57.

Karen Spärck Jones. 2007. Automatic summarising: The state of the art. *Inf. Process. Manage.*, 43(6):1449–1481.

Chin-Yew Lin and Eduard Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proc. Int. Conf. on Computational Linguistics (ACL)*, pages 495–501.

Chin-Yew Lin and Eduard Hovy. 2003a. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proc. of NAACL-HLT 2003*, pages 71–78.

Chin-Yew Lin and Eduard Hovy. 2003b. The potential and limitations of automatic sentence extraction for summarization. In *Proc. of the HLT-NAACL 2003 Workshop on Text Summarization*, pages 73–80.

Christopher D. Manning and Hinrich Schütze. 2001. *Foundations of Statistical Natural Language Processing*. MIT Press.

Ani Nenkova, Lucy Vanderwende, and Kathleen McKeown. 2006. A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization. In *Proc. of SIGIR '06*, pages 573–580.

Jay M. Ponte and W. Bruce Croft. 1998. A language modeling approach to information retrieval. In *Proc. of SIGIR '98*, pages 275–281.

M. F. Porter. 1980. An algorithm for suffix stripping. *Program*, 14(3):130–137.

Mark Steyvers and Tom Griffiths. 2006. Probabilistic topic models. In S. Dennis T. Landauer, Mcnamara and W. Kintsch, editors, *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum.

Hanna M. Wallach. 2006. Topic modeling: beyond bag-of-words. pages 977–984, Pittsburgh, Pennsylvania. ACM.

Xuerui Wang, Andrew McCallum, and Xing Wei. 2007. Topical n-grams: Phrase and topic discovery, with an application to information retrieval. In *Proc. of ICDM '07*, pages 697–702.

W. Yih, J. Goodman, L. Vanderwende, and H. Suzuki. 2007. Multi-document summarization by maximizing informative content-words. In *Proc. of IJCAI 2007*, pages 1776–1782.

Population and Enrichment of Event Ontology using Twitter

Shashi Narayan, Srdjan Prodanovic, Mohammad Fazleh Elahi, Zoë Bogart
University of Malta
E-mail: shashi.narayan@gmail.com, tarzanka@gmail.com, mohammadfazlehelahi@gmail.com, zbogart@gmail.com

Abstract

In this paper we present an approach towards creating and enriching an event ontology from social media data gathered from Twitter. Such an ontology is useful in providing accurate, up-to-date information in response to user queries from around the globe. The dynamic nature of Twitter and the broad coverage of user tweets allow us to consolidate information about a wide range of events from conferences to concerts, festivals, and so on. To enrich the ontology we have established a methodology for learning new types of events, so that, in parallel with population of the ontology, completely automatic enrichment is performed as well, fully.

1. Introduction

Ontology can be described as an explicit, formal specification of a domain of interest that should be expressed in a machine-readable way (Gruber, 1994). It should be restricted to a given domain of interest and therefore model concepts and relations that are relevant to a particular task or application domain. During the last few years, there has been a growing interest in ontologies due to their ability to explicitly describe data in a common and structured way.

Currently, information about events are available either through local websites (that publish information only on local events) or through the event publisher websites[1]. Common feature of both kinds is that the information published there are either manually inserted or purchased from a third party supplier. As a result of the latter, data about events are often incomplete, when it comes to lower profile events. End users are therefore often faced with the pointless search of the event information about those events that are not that well known.

The goal of this paper is to populate an ontology with the most recent events extracted from social media. Because of this, we needed to choose a social media format which contains information about currently occurring events. We chose *Twitter* as an online social network based on a micro blogging platform used by millions of people around the world.

Data obtained from Twitter are then parsed and mined for various features which are later placed in the event ontology. In the process, data obtained are also analyzed for the purpose of ontology enrichment. Enhanced ontology allows having finer grained and more sophisticated queries about events.

2. Related work

There are several research work where the main objective is to identify events from news documents (e.g., newswire, radio broadcast) (Allan et al, 1998; Kumaran & Allan, 2004; Yang et al, 1998). Our task of event extraction is similar to those event detection tasks. But these traditional event detection tasks aim to discover and cluster events found in textual news articles which are in the form of standard text in nature. Therefore, most state-of-the-art event detection approaches leverage between NLP tools such as named-entity extraction and POS tagging (Makkonen et al, 2004; Zhang et al, 2007). However, these approaches have been developed mainly for standard written text. Event detection has not been developed for unstructured text organized in the form of social media like Twitter where the tweets (basic text message unit of Twitter) are noisy, abbreviated, and condensed in nature.

Twitter has evolved from such a simple text message-based communication system to a powerful platform for distributing and consuming all types of information. The dataset of available tweets is extremely large and dynamic with a great diversity of topics and large number of users from all around the globe (Buzzgain, 2010). Twitter's content gives a wide range of events from conferences to concerts, festivals, and so on. Java et al (2007) found many users report latest news or comment about current events on Twitter. Twitter is no longer just a system to provide simple status updates of personal information; Twitter plays a significant role in time sensitive information seeking and decision making in less urgent and everyday activities (Brooks & Churchill, 2010; Sutton et al, 2008).

3. Methodology

3.1 Tweet retrieval

The main purpose of this module is to retrieve all tweets containing information about events (hereby referred to as *event tweets*), and for that purpose we used the Twitter API. We found that a search of Twitter by keywords proved to be the most efficient retrieval method. For a keyword-based search, a basic set of 28 event types was compiled. For the purpose of improving retrieval process, we have extended base set with the synonyms from WordNet using NLTK[2]. When retrieved, tweets were filtered on a simple criterion – the existence of a time and/or a date in the tweet, based on reasoning that an event cannot be an event if this component is not present.

[1] http://eventful.com/, http://eventguide.com/

[2] http://www.nltk.org/

3.2 Extraction of event features

The base ontology consisted of five properties, the features of event: *Name*, *Date and/or Time*, *Location*, *Type*, and *URL*. For date and time we used the existing ontology Time[3] (Hobbs & Pan, 2004), adopted by W3C. The NLP methods of feature extraction from event tweets are described below:

For *Date and Time* extraction, regular expressions were used. This date/time information was then stored in Time ontology (Hobbs & Pan, 2004) mentioned earlier. A link (*URL*) is a very common feature found in twitter events, and we included this information to allow users to access a detailed description of event. For *Location and Name*, a set of various linguistic cues such as pre-modifiers, parts of speech, and word capitalization was used to identify candidates for locations or names; these candidates were then compared against the Alexandria Digital Library Gazetteer (1999) to find and validate proper locations which were then placed in our own version of the Gazetteer for future reference to improve the overall efficiency of the system. We used the expanded set of keywords to identify candidates for different event types (*Type*). These type words were then matched to the types already present in ontology, and if found, they were then instantiated as those types for the specific event being examined. The set of types can be iteratively expanded and it can serve as basis for ontology enrichment process.

These events were then checked against and mapped into our ontology, according to their type. We present an example of a good event tweet below:

5th Annual Winter Wildlands Alliance (*Name*) Film Festival (*Type*) in Seattle **(***Location***)** – 1/22/20108 (*Date*) http://bit.ly/4syypF (*URL*).

In the final step, results were taken directly from our established format of tagged events and were mapped into the OWL file by converting them from XML to RDF format.

4. Results

The average number of retrieved tweets was 200 per hour. From every request made towards the Twitter API, at least two (out of fifteen returned) candidate event tweets were retrieved after filtering. The dataset for our experiment was created using Twitter API for a whole day. Out of a total of 470 filtered tweets, 413 were found to be event tweets, so our ontology contained 413 instances. Out of 413 found event tweets, 386 were correctly classified event tweets. Rest 27 was wrongly classified non-event tweets. These results yield overall precision of 95.76% and recall of 93.46%, which gives F-measure of 94.59. Note that evaluation presented here is done on already filtered tweets based on date existence criteria as already previously described.

Despite the success of our processing techniques, certain number of event tweets were either wrongly classified or missed out. This is due to the fact that there some irregularities in the data that made extraction of event features more difficult, namely misspellings (due to the chat-like structure of tweets) and structural irregularity (missing parameters of events like name or type of event). These irregularities could be smoothed and reduced to a minimum by improving NLP techniques for event feature extraction, which is one of our objectives for the future development of the system.

5. Conclusion and future work

Initial results suggest that this is a promising method for building an event ontology, as one can easily acquire a lot of information about events from Twitter. The choice of Twitter as the source for building and populating the event ontology is a sensible one, because of the large volume and dynamic nature of data available.

We believe that in employing ontology, our system helps introduce the necessary structure needed for efficient exploitation of these easily accessible data which are currently being underutilized. Storing these data in ontology has the following advantages over storage in a traditional database:

- Data availability: instead of keeping the data stored in a database and retrieving them via methods specific to that database, in an ontology, these data are available in the common language platform and so it is much simpler to retrieve them.
- Inference power: by employing the inference mechanisms of ontology languages, lists of specific types of events can easily be obtained for users (Walton, 2007).

As far as the domain of future work is concerned various following improvements of the system could be explored:

- Ontology enrichment: by considering the words in context around the event type and their frequency of appearance new subtypes/sub-concepts could be introduced. For instance, if the phrase “press conference” appeared frequently in event tweets, we would consider it to be a new sub-concept of the concept “conference”. Along with ontology enrichment process the inference power of ontology would increase as well
- Recommendation system: by using the data that Twitter has on its users (geo-targeting, possibly age and gender information), specific events could be recommended to them. Also if the user is following large number of twitter feeds, there is a possibility that he might miss an event, as stated in (Bernstein et al, 2010). Specifically, a priority could be given to the tweets that come from the Twitters that user is following, as oppose to tweets that come from other feeds.

In sum, we believe that the methodology described in this paper possesses great potential for extracting useful information from the source, which is free, global, and up-to-date. Mapping these data into an ontology makes their use of them easier and more efficient, due to the underlying expressive power of ontology languages.

[3] http://www.w3.org/TR/owl-time/

6. References

Alexandria Digital Library Gazetteer (1999). *Map and Imagery Lab, Davidson Library, University of California, Santa Barbara. Copyright UC Regents*

Allan, J., Papka, R., Lavrenko, V. (1998). On-line new event detection and tracking. In *Proceedings of the 21st ACM International Conference on Research and Development in Information Retrieval, SIGIR 1998, Melbourne, pp. 37-45*

Bernstein, M., Kairam, S., Suh, B., Hong, L., Chi, E.H. (2010). A Torrent of Tweets: Managing Information Overload in Online Social Streams. *ACM CHI 2010, Atlanta, Georgia, USA*

Brooks, A.L., Churchill, E. (2010). Tune In, Tweet on, and Twit out: Information snacking on Twitter, *ACM CHI 2010, Atlanta, Georgia, USA*

Buzzgain (2010). How many tweets does it take to be a trending topic on Twitter?. *http://news.buzzgain.com/how-many-tweets-does-it-take-to-be-a-trending-topic-on-twitter/*

Gruber, T. (1994). Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human and Computer Studies, Vol. 43, Issues 5-6, November 1995, pp. 907-928*

Hobbs, J. R., Pan, F. (2004). An Ontology of Time for the Semantic Web. *ACM Transactions on Asian Language Processing (TALIP): Special issue on Temporal Information Processing, Vol. 3, No. 1, March 2004, pp. 66-85*

Java, A., Song, X., Finin, T., Tseng, B. (2007). Why we twitter: understanding micro blogging usage and communities. *International Conference on Knowledge Discovery and Data Mining, ACM New York, USA, pp. 56-65*

Kumaran, G., Allan, J. (2004). Text classification and named entities for new event detection. In *Proceedings of the 27th ACM International Conference on Research and Development in Information Retrieval, SIGIR 2004, pp. 297-304*

Makkonen, J., Ahonen-Myka, H., Salmenkivi, M. (2004). Simple semantics in topic detection and tracking. *Information Retrieval, Vol. 7, Issues 3-4, September-December 2004, pp. 347-368, 2004*

Sutton, J., Palen, L., Shklovski, I. (2008). Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires, *Proceedings of the 5th International ISCRAM Conference – Washington, DC, USA, May 2008*

Walton C (2007). Agency and the Semantic Web (Chapter 5: Reasoning On the Web) *Oxford University Press*

Yang, Y., Pierce, T., Carbonell, J. (1998). A study on retrospective and on-line event detection. In *Proceedings of the 21st ACM International Conference on Research and Development in Information Retrieval, SIGIR 1998, Melbourne, pp. 28-36*

.Zhang, K., Zi, J., Wu, L.G. (2007). New event detection based on indexing-tree and named entity. In *Proceedings of the 30th ACM International Conference on Research and Development in Information Retrieval, SIGIR 2007, Amsterdam, pp. 215-222*

Semantic TV Engine: An IPTV Enabler for Personalized Recommendations

Juri Glass, Stefan Marx, Torsten Schmidt, Fikret Sivrikaya

DAI-Labor
TU-Berlin, Ernst-Reuter-Platz 7, D-10587 Berlin
{juri.glass|stefan.marx|torsten.schmidt|fikret.sivrikaya}@dai-labor.de

Abstract

With today's availability of an increasing amount of media content delivered by IPTV systems, the need for personalized content preparation arises. In order to organize media content tailored to the users' needs, we are faced with challenges regarding data representation, integration and user profiling issues. Our approach is to use semantic technologies to build a unified data integration layer based on well-known ontologies and to apply a recommender system on top of the data plane that provides personalized recommendations considering the user's context. In this paper we introduce a high level overview of an enabler component for IPTV infrastructures, which combines semantic data management and recommendation functionalities. As a proof of concept we have developed an IPTV application that interacts with the new enabler.

1. Introduction

Today's IPTV systems allow the delivery of all kinds of multimedia content. The ever-increasing amount of available multimedia resources requires filter and recommendation mechanisms to assist the user in content selection. Widely researched and used approaches are content-based, collaborative or hybrid filter and recommendation methods. Content-based methods determine the similarity between program metadata and user preferences, whereas collaborative methods are based on correlations between users and their consumption behaviour. Hybrid approaches combine both methods to minimize the drawbacks of each of both methods.

The use of recommenders for personalized TV environments is common practice and the utilization of semantic techniques for the underlying data representation, on which the recommenders are working, is fairly adopted. AVATAR is an automatic content recommender based on Semantic Web technologies to enable hybrid recommendations. This content recommender approach utilizes a hierarchical TV content classification based on a self-defined ontology to infer new knowledge, in order to match user profiles with TV content. Here, similarity of content is derived from nearness in the defined classification hierarchy and the discovery of relations between contents (Fernandez et al., 2006). However, in the user profile only the related content items are stored. To further enhance recommendations more context from the user can be incorporated. SenSee, a framework for personalized access to TV content, improves recommendations by the evaluation of user context attributes such as location, time and audience (Aroyo et al., 2007). For SenSee heterogenous meta-data sources (TV-Anytime, XML-TV, IMDB movie meta-data) are transformed into TV-Anytime representation to enable uniform data access for their recommender. iFancy, an application based on SenSee, uses in contrast to AVATAR well-known ontologies for time, geography and linguistic concepts, in which TV Anytime data is mapped (Akkermans et al., 2006).

As seen above TV content describing meta-data is provided by different sources in different formats, so we are facing the challenge to represent all data needed by a recommender in a uniform format. An issue that is not considered by most TV recommendation research issues is the integration of a recommender and the related data management functions in existing, standardized IPTV infrastructures. So we identified this as a second challenge.

In this paper we present the *Semantic TV Engine*, an enabler for personalized recommendations which uses semantic technologies to build a data integration layer that combines content metadata with user related information like interests, attention and interaction data. The use of a unified representation and well known ontologies allows us to easily extend the data plane with almost any information source. On top of the semantic data layer we set up a recommender system that is used to calculate the most appropriate content tailored to the user's needs. We show how the *Semantic TV Engine* could be integrated in the standardized IPTV infrastructure developed by the ITU-T.

For testing and demonstration purposes, a user application has been developed, which allows the user to watch TV streams, to view a TV program guide, to commit ratings, or to select and define TV program preferences such as genres and favorites. Using these data as input, the application makes use of the Semantic TV Engine to estimate a personalized *MyTV channel*, which consists in a sequence of recommended program items.

The remainder of the paper is organized as follows: In Section 2 we introduce our approach to build a semantic data organization enabler, whereas in Section 3 we describe its individual components in more detail. In Section 4 we discuss the representation of data and metadata and an overview of the ontologies we use. Section 5 outlines the integration of *Semantic TV Engine* as an enabler in a standardized IPTV architecture and its use by a demonstrator application. Section 6 concludes the paper with a brief outlook.

2. Semantic TV Engine Architecture

The Semantic TV Engine Architecture consists of four main components; the Data Retrieval and Transcription component, the Semantic Store, the Universal Recommender, and the Semantic Services Exposure (Figure 1).

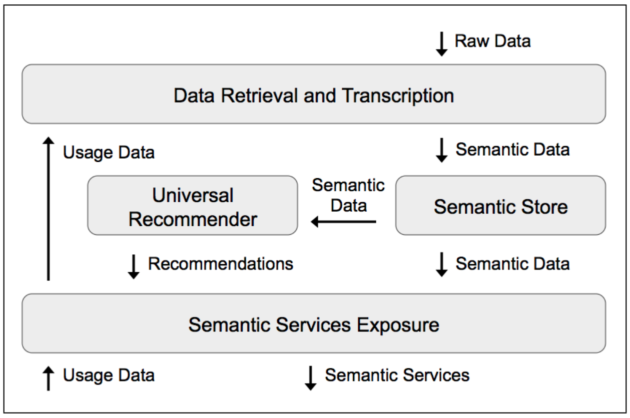


Figure 1: Semantic engine architecture

The *Data Retrieval and Transcription* component deals with the retrieval and transformation of unstructured or semi-structured data into a semantic format. All transformed data is stored in the *Semantic Store*, which is a storage entity for the management of semantic data graphs. The third component is the *Universal Recommender*, which uses the data stored in the semantic store to calculate recommendations delivered by the Semantic Services Exposure component. The *Semantic Services Exposure* component is providing the enabler interface to multiple client applications allowing the retrieval of metadata, such as EPG, the request for recommendations, and MyTV channel composition. The Semantic Services Exposure also provides interfaces to applications for collecting user and usage data through the *Data Retrieval and Transcription* component, which in turn is stored in the semantic store linked to content metadata. Next we present a more detailed overview of each component within the Semantic TV Engine.

2.1. Data Retrieval and Transcription

In the general case, datasets can be modeled as a set of entities connected by relationships, whereas complex datasets contain multiple relationship types. This is especially true when several datasets are combined. As a result such datasets form a semantic network, connecting several entity types through typed relationships. To model a semantic network, the semantic TV engine takes unstructured and semi-structured data and transforms it to a linked data graph represented in RDF (Resource Description Framework). Therefore content metadata (e.g. TV Anytime, proprietary channel lists, etc.) is gathered via Web Service requests against a content provider middleware, whereas usage data is gathered through REST interfaces provided by the semantic TV Engine itself. The retrieved usage data and metadata are parsed and mapped to entity and relationship types defined by well known ontologies. To implement the mapping we have analyzed the data and selected appropriate ontologies beforehand. By its modular design, the Data Retrieval and Transcription component is easily extendable by new retrieval adapters, mapping modules and ontologies. With the completion of the semantic modeling process, the component stores the data in the semantic store.

2.2. Semantic Store

Semantic Store is the component responsible for storing all the information retrieved from external metadata sources and user interactions. By linking both sets of information via ontologies together, a hybrid recommender combining collaborative and content-based approaches is empowered. The Semantic Store uses the open-source semantic web framework Jena for organizing the data. Jena can be used as an abstraction layer on top of a triple or quad store. A semantic triple consists of three parts: subject, predicate, and object, where the subject is an identifiable node in the semantic graph, the predicate is a named directed edge and the object is the node which relation to the subject is described through the predicate. This triple structure reflects the structure of the widely adopted RDF (Resource Description Framework).

The semantic store incorporates two beneficial characteristics regarding data integration:

- Subjects or even subgraphs identified by URIs are automatically updated during a storage transaction
- Relationships that connect entities are automatically added to the semantic model, which allows easy extension of the dataset by new information resources

In extension named graphs, represented as quads where the fourth element denotes the affiliation to a specific graph, are used to cope with the transient nature of metadata in live TV. The named graphs are annotated with a timestamp and named as aging graphs to enable efficient memory management. The access to the Semantic Store is done via SPARQL (SPARQL Protocol and RDF Query Language). However, the actual data representation is described in Section 3.

2.3. Universal Recommender

Universal Recommender (UR) is the component in the described architecture responsible for the generation of personalized recommendations (Kunegis et al., 2009). Functionally, a recommender takes an entity as input (e.g. user or content-id) and outputs a list of ranked entities (e.g. program items).

Designed for the use of semantic datasets that generalizes domain-specific recommenders such as content-based, collaborative, social, bibliographic, lexicographic, hybrid and other recommenders, the UR applies to any dataset that allows a semantic representation. In addition, the UR is capable of working with any number of entity and relationship types and can learn all its parameters without human interaction through advanced machine learning techniques.

We apply the UR to the metadata about the ongoing program as well as the usage data collected from users.

2.4. Semantic Services Exposure

The Semantic Services Exposure component exposes services through a multi-purpose interface to a client application, which can be a user-client or a server-side IPTV function. The services provided by the Semantic Services Exposure component can be grouped according to three different use-cases: Plain metadata retrieval, recommendation requests, and submitting user feedback (including different user actions, such as rating, marking as favorite, and consumption) to the *Semantic TV Engine*.

These interfaces follow the REST (Representational State Transfer) model, whereas the exchanged data is represented as a JSON formatted payload. REST allows the lightweight access needed by user-clients as well as high performance throughput due to the implicit caching possibilities since it is based on HTTP. Furthermore, it suits the set/get nature of the data communication of the presented application.

An example for a request could be $/programs/now$, which may return all the information about the current programs.

3. Data Representation

To enable hybrid recommendations, which combines content-based and collaborative techniques, the integration of heterogeneous data is an important task. The approach for data integration presented in this paper is the exploitation of semantic methods as seen in the Semantic Web.

One major concept for the Semantic Web is the use of reusable ontologies. Ontologies are a formal representation of entities and the relations between them. Ontologies can be used to classify and order entities hierarchically, but are in most cases restricted to specific domain. They can be used to identify similar concepts in different and independent datasets, which is an important task in data integration.

For our context we identified several well known ontologies, which we use to model our data semantically; BBC Programme Ontology (Raimond et al., 2009), FOAF (Brickley and Miller, 2010), Event Ontology (Raimond and Abdallah, 2007a), Timeline (Raimond and Abdallah, 2007b), RDF Review Vocabulary (Ayers and Heath, 2007), Tag Ontology (Ayers et al., 2005) and an ontology of our own, which we call *Program Ontology Extensions*.

Data following the BBC Programme Ontology is central in our data representation since all TV broadcasts and programs are modeled accordingly. Channel and program descriptions gathered from TVA (TV Anytime) program information table is modeled with the ontologies' core entities *Broadcast*, *Service* and *ProgramItem*, whereas the schedule information table is mapped to Entities derived from the Timeline and Event Ontology interlinked with BBC programs. For the modeling of users we use the well known FOAF ontology and link usage data with *foaf:Person* entities representing users. The usage data is modeled following the Review Vocabulary and the Tag Ontology that allow the semantic representation of e.g. ratings and tags. For other parts of the user profile, such as the viewing history and user preferences, we introduce the Program Ontology Extensions. The extension also interlinks information available in TV Anytime with the Program Ontology that is not a priori defined by it (e.g. parental guidance information).

4. Semantic TV Engine Integration in an IPTV Environment

In our IPTV system development we are following an IP Multimedia Subsystem (IMS) based approach, which is announced by several standardization bodies like the ETSI TISPAN, ITU-T and Open IPTV Forum. Here the IPTV service control functions and applications are built on top of an IMS based service control core that makes use of the Session Initiation Protocol (SIP) to provide user profile based service control (TISPAN, 2009) (ITU-T, 2008) (OpenIPTV, 2009). Since we see the Semantic TV Engine as an additional IPTV service enabler, we integrate it in IPTV service providers' IPTV applications layer, on top of the IMS service control core and IPTV service support and control functions. This achieves the smooth interaction with other IMS and IPTV service enablers and applications like IPTV service control and applications, IPTV user profiles, presence server and XDMS based group management.

4.1. Example Integration with the ITU-T IPTV architecture

The following describes the integration especially with the architecture of an IPTV system recommendation of the ITU-T. We foresee the Semantic TV Engine located in the Application Functions layer, between the Content Preparation Functions and the content and meta data sources. The Content Preparation Functions transform all available content (VOD data, EPG, metadata, TV streams) to the required format requested by the IPTV applications and content delivery functions. By accessing the content and meta data sources through the Semantic Services Exposure component, the Content Preparation Functions are able to provide enhanced data to the IPTV applications.

In this integration the Data Retrieval and Transcription component of the Semantic TV Engine collects metadata produced by content providers and service providers to describe multimedia content, as well as metadata produced by metadata clients to describe user preferences or context. The metadata is stored in the Semantic Store and is accessed by or contributed from metadata clients through the Semantic Services Exposure component.

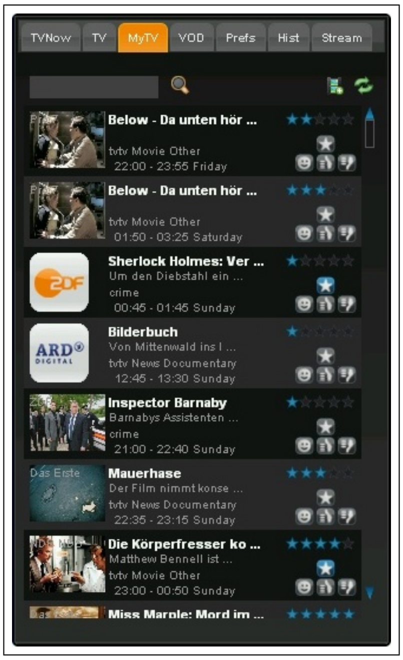
In our current development metadata clients are HTML/Javascript based applications running in a Webbrowser on the IPTV client, accessing the REST interface of the Semantic Services Exposure component.

4.2. Integration with the user application

Our WebTV user application depicted in Figure 2 is a hybrid application consisting of a web application based on ECMA and HTML/CSS for the EPG and recommendation views and a Java based application utilizing SIP and H.264 streaming for live TV stream delivery. The web based approach allows the platform-independent integration of our application as long as a web browser is available.

Beside its other functionalities the user application provides an EPG program, the recommenation base MyTV program and a preferences configuration view. From the EPG program view the user can trigger streaming of the current broadcasts, mark program items as personal favorite, rate program items and recommend program items to other users. In the preferences configuration view the user defines a set of genres he likes.

All these user input actions are tracked by the application and sent via AJAX requests to the REST interface of the Semantic Services Exposure component. However, other ap-



Below - Da unten hör ...
tvtv Movie Other
22:00 - 23:55 Friday

Below - Da unten hör ...
tvtv Movie Other
01:50 - 03:25 Saturday

Sherlock Holmes: Ver ...
Um den Diebstahl ein ...
crime
00:45 - 01:45 Sunday

Bilderbuch
Von Mittenwald ins I ...
tvtv News Documentary
12:45 - 13:30 Sunday

Inspector Barnaby
Barnabys Assistenten ...
crime
21:00 - 22:40 Sunday

Mauerhase
Der Film nimmt konse ...
tvtv News Documentary
22:35 - 23:15 Sunday

Die Körperfresser ko ...
Matthew Bennell ist ...
tvtv Movie Other
23:00 - 00:50 Sunday

Miss Marple: Mord im ...

Figure 2: WebTV application: MyTV recommendations view

proaches are possible; for instance a special SIP application server placed in the SIP signaling path can provide tracked user actions to the Semantic Services Exposure component. EPG data and the MyTV program is asked by the user application from the Semantic Services Exposure component, too.

5. Conclusion and Outlook

For the future we plan to enrich our metadata with data provided via linked open data (LOD). Large amounts of information is publicly available in the so-called LOD-cloud, which is assumed to be very valuable to integrate with our application and the IPTV context. We envision integrating linkedMDB data, a semantic integration of the famous iMDB service providing detailed information about movies, TV series etc. Another valuable integration would be the interlinkage with DBPedia, which is a semantic dataset derived from Wikipedia articles (Kobilarov et al., 2009). In the future we may extend our semantic TV engine by modules that are able to compute owl:sameAs relationships, which expresses equivalence between entities in two different datasets.

Additional work has to be done to exploit semi-structured application data that is already in use throughout standardized IPTV systems operated by e.g. telecommunication providers. We expect a huge potential in the semantic exploitation of communications related information (e.g. presence data, social web applications). Exploiting these information bears large potential to improve personalized recommendations. Furthermore, this kind of data might form an interesting new dataset, which can be contributed to the LOD-cloud as a basis for new intelligent information services. A similar goal is described in (van Aart et al., 2009). An interesting research topic in this direction will be the modeling and representation of real-time information and events in semantic formats. Regarding real-time information, a next step will be the integration of social web services to our application, since this is nearly a mandatory feature of multimedia applications today.

6. References

P. Akkermans, L. Aroyo, and P. Bellekens. 2006. iFanzy: Personalised filtering using semantically enriched TVAnytime content. In *ESWC2006 Proceedings*.

Lora Aroyo, Pieter Bellekens, Martin Björkman, Geert-Jan Houben, Paul Akkermans, and Annelies Kaptein. 2007. Sensee framework for personalized access to tv content. In *EuroITV*, pages 156–165.

D. Ayers and T. Heath. 2007. RDF review vocabulary. http://purl.org/stuff/rev#.

D. Ayers, S. Russell, and R. Newman. 2005. Tag ontology. http://www.holygoat.co.uk/owl/redwood/0.1/tags/.

D. Brickley and L. Miller. 2010. FOAF Vocabulary Specification 0.97. http://xmlns.com/foaf/spec/.

Y. B. Fernandez, J. J. P. Arias, M. L. Nores, A. G. Solla, and M. R. Cabrer. 2006. AVATAR: An improved solution for personalized TV based on semantic inference. *IEEE Transactions on Consumer Electronics*, 52:223–231.

IPTV Focus Group ITU-T. 2008. ITU-T IPTV Focus Group Proceedings. http://www.itu.int/publ/T-PROC.

Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. 2009. Media meets semantic web - how the bbc uses dbpedia and linked data to make connections. In *ESWC*, volume 5554 of *Lecture Notes in Computer Science*, pages 723–737. Springer.

J. Kunegis, A. Said, and W. Umbrath. 2009. The Universal Recommender. http://www.dai-labor.de/index.php?id=1691&pubID=442.

Open Iptv Forum OpenIPTV. 2009. Functional Architecture V2.0. http://www.openiptvforum.org/specifications.html.

Y. Raimond and S. Abdallah. 2007a. The Event Ontology. http://motools.sourceforge.net/event/event.html.

Y. Raimond and S. Abdallah. 2007b. The Timeline Ontology. http://motools.sourceforge.net/timeline/timeline.html.

Y. Raimond, P. Sinclair, N. J. Humfrey, and M. Smethurst. 2009. BBC Programmes Ontology. http://www.bbc.co.uk/ontologies/programmes/2009-09-07.shtml.

ETSI TISPAN. 2009. ETSI TS 182 027 V2.4.1 - TISPAN; IPTV Architecture; IPTV functions supported by the IMS subsystem. http://etsi.org.

C. van Aart, R. Siebes, V. Buser, L. Aroyo, Y. Raimond, D. Brickley, G. Schreiber, M. Minno, L. Miller, D. Palmisano, and M. Mostarda. 2009. The NoTube Beancounter: Aggregating user data for television programme recommendation. In *Social Data on the Web (SDoW2009)*.

Summarization and Visualization of Digital Conversations

Vincenzo Pallotta[1], Rodolfo Delmonte[2], Marita Ailomaa[3]

Department of Computer Science
Webster University
Geneva - Switzerland

Department of Language Science
Università "Ca Foscari"
Venezia - Italy

Artificial Intelligence Lab
EPFL
Lausanne - Switzerland

E-mail: pallotta@webster.ch, delmont@unive.it, marita.ailomaa@gmail.com

Abstract

Digital conversations are all around us: recorded meetings, television debates, instant messaging, blogs, and discussion forums. With this work, we present some solutions for the condensation and distillation of content from digital conversation based on advanced language technology. At the core of this technology we have argumentative analysis, which allow us to produce high-quality text summaries and intuitive graphical visualizations of conversational content enabling easier and faster access to digital conversations.

1. Introduction

Conversations have been central to the Internet since its birth. One of the first Internet applications, IRC (Internet Relay Chat), was of conversational nature. Later, conversational systems have proliferated in various forms. With the advent of Web 2.0, the Internet has become more collaborative and in many situations, new modes of information sharing were based on conversation (e.g. blogs, social media, remote conferencing, wiki).

Communication through conversation is very effective because it releases users from the commitment to engage in the full process of content publishing, which was the original model of the Web. Conversations are situated within a context and users can avoid using a narrative style in their communications. This effectiveness also leads to the fact that individual contributions to the conversation are often impossible to be isolated from their context thus leading to some difficulties in retrieving relevant information from this type of data. For instance, in online reputation management[1], tracking the users' attitude towards a product or a brand in public forums may become very challenging. Simply considering the polarity of individual contributions could lead to misinterpretations of data in cases such as those when a user negatively comments on another user's comment. In such a case, the attitude cannot be simply understood as negative to the main topic but it needs to be understood in its own context, namely as a reply to a (possibly negative) comment to the main topic.

The above is only one of the possible problems that a standard approach to the analysis of Internet conversation might generate. Other problems are the absence of certain keywords that might be used to retrieve passages of conversations that are not uttered by the participants. For instance, if one wants to retrieve passages where someone disagreed with someone else in a forum, the term "disagree" is not likely to be found in the conversation. In these cases, it is essential to enrich the conversational content with metadata so that a retrieval system can find relevant information not just from the content terms.

We provide a solution to the above problems by presenting a new approach to the indexing and retrieval of conversational data. This approach is based on the reasonable assumption that, when engaged in conversations, users follow a flexible but well-defined discourse structure. This structure is very often an *argumentative structure* where participants contribute with specific dialogue acts with a precise argumentative force. This model might not be applicable to every conversation, especially those which do not have a clear purpose. However, we believe that a great deal of digital conversations are purposeful and that users almost always pursue the goal of either seeking agreement or consensus from other participants on their opinions, or trying to convey their arguments in favor or against other participant's opinions.

The article is organized as follows. In section 2 we explore the domain of digital conversations with a special focus on Internet conversations. We advocate the need of new tools to provide more appropriate ways for accessing this type of content. In section 3 we present an approach for the analysis of conversations from the argumentative perspective making the case for a new way of indexing conversational data under specific conditions. In section 4 we present a diagrammatical visualization which provides users with an intuitive global view of digital conversations by highlighting their argumentative structure. Finally, we showcase a summarization tool for digital conversations which produces high-quality memos based on both automatically extracted thematic and argumentative structure. We conclude the article with an assessment of the current state of our research and propose a roadmap for future work.

2. Digital conversations

Digital Conversational Systems (DCSs) are systems that support the creation of digital records of conversations[2]. DCSs can be based on stand-alone or interconnected computer systems (e.g. networked through Internet). We

[1] en.wikipedia.org/wiki/Online_reputation_management.

[2] The term Digital Conversation is also used in marketing to denote new ways of advertising based on dialogue between producers and consumers. See en.wikipedia.org/wiki/Digital_conversation.

review different types of DCSs by outlining what are the essential retrieval requirements for each type. We distinguish two main categories of DCSs: synchronous and asynchronous. Synchronous DCSs are those in which the exchange between participants occurs in real-time and no history of the conversation (before one joins in) is maintained by the system whatsoever. When the user joins an already started conversation there is no means of retrieving what has been said before joining, and when the user leaves the conversation what follows is no longer accessible to the user. In this category we find Instant Messaging and VoIP systems and meeting recording systems. Asynchronous systems, in contrast, allow users to keep track of the whole conversations even when they are not directly participating to them. In this category we find the most advanced Internet collaborative technologies such as discussion forums, blogs, micro-blogs, wikis, and social media.

Conversations from synchronous systems can be recorded and made accessible after the conversation is over similarly to asynchronous systems. However, we need to bear in mind that because of its nature, these conversations might be shorter and more focused than asynchronous ones. Consider for instance the case of a recording of a chat conversation from the moment one user enters the chat to the moment the user leaves it. It is likely that there was an already started conversation going on at the joining time. The user will have to listen to the ongoing conversation for a while in order to figure out the current topic. When the user starts to actively participate in the conversation we can assume that the context is clear enough for the user and we can interpret the utterances only relying on that (limited) context. When looking for an argumentative discourse structure, we can no longer assume that we will find the root of the argument (i.e. the initial topic) in the recorded conversation.

Asynchronous ICSs have a more consistent structure because roots of arguments can be easily identified. However, they might have a mixed narrative and argumentative structure thus making sometimes the analysis more difficult. For example, consider a blog or a social media where there is an initial post or video and a sequence of comments attached to it. Clearly, the initial post provides a wealth of context for the interpretation of the related comments. Comments can relate both to the main topic (established in the post or in the video) and to other comments.

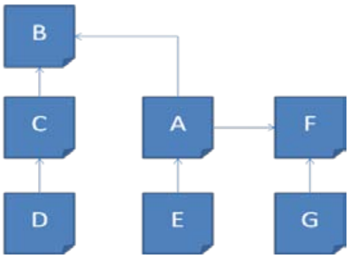


Figure 1: Conversational Structure of Blogs

The case of blogs is also challenging because the conversations are distributed over different sites. Rolling out a blog conversation can be very difficult because comments to the initial post can be contained as individual posts in other blogs. The distribution of conversational content implies that any retrieval technique needs to perform a crawl of the sites where the contributions are stored. This also entails that there may be multiple entry points in the conversations and that one post can contribute to several conversations.

In Figure 1, the post A participates in two conversations (B<A<E and A>F<G), while the post B has started two independent conversations (B<C<D and B<A<E) [3]. Notice that the same author might have posted multiple posts (e.g. A and D), but also that authors are by default unaware of the replies to their post unless they explicitly search for them. In order to avoid this loss of information, blogging platforms have enabled an acknowledge system that informs the author of a post when the post is referenced in another blog post. This mechanism is referred to as *linkback*[4].

Being actually fully fledged websites, blogs offer a wealth of media and can contain any amount of metadata. As a matter of fact, blogs are hosted in content management systems and the way the media are integrated can be sometimes rigid. From the indexing and retrieval perspective, blogs still retain their document connotation. Blog search engines (such as Technorati[5]) index blogs on their textual content. Additionally, relevance ranking is modified and restricted to inter-blogs links. In such a way it would be impossible to retrieve all blog posts that commented (i.e. linked) a specific post. Moreover, inter-blogs comments are not treated separately and in the best case they are simply excluded from the index. While blogs represent a huge portion of Internet conversations they are not treated in an appropriate way and the conversational structure is typically not made explicit for search.

3. Argumentative indexing of conversations

Conversations are a pervasive phenomenon in our digital society. We need to consider appropriate techniques to analyze conversational content from different perspectives beyond classical thematic indexing approaches. We advocate that recognizing the argumentative structure of digital conversations can help in improving the effectiveness of standard retrieval techniques in simple cases and even overcome their limitations in complex cases. To better understand the impact of argumentative analysis we will provide in this section a real example of how argumentative indexing can solve outstanding problems in indexing and retrieval of conversational content. In our approach, we adopt a representation of conversational structure based on argumentation theory (Pallotta 2006). The argumentative structure defines the different patterns of argumentation used by participants in the dialog, as well as their organization and synchronization in the discussion.

[3] We use the notation ">" to indicate the "replies-to" relation: A>B means A replies to B.

[4] Three linkback protocols are commonly in use in different blogging platforms: pingback, trackback and refback.

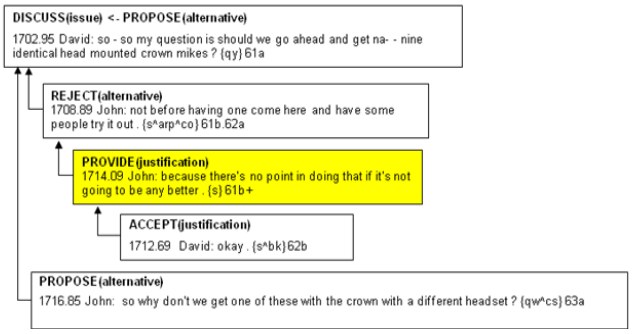
[5] www.technorati.com.

A dialog is decomposed into several argumentative episodes such as issues, proposals, elaborations and positions, each episode being possibly related to specific aggregations of elementary dialog acts. We adopted an argumentative coding scheme, the Meeting Description Schema (Pallotta et al. 2004) where the argumentative structure of a meeting is composed of a set of topic discussion episodes (a discussion about a specific topic). In each topic discussion, there exists a set of issue episodes. An issue is generally a local problem in a larger topic to be discussed and solved. Participants propose alternatives, solutions, opinions, ideas, etc. in order to achieve a satisfactory decision. Meanwhile, participants either express their positions and standpoints through acts of accepting or rejecting proposals, or by asking questions related to the current proposals. Hence, for each issue, there is a corresponding set of proposal episodes (solutions, alternatives, ideas, etc.) that are linked to a certain number of related position episodes (for example a rejection to a proposed alternative in a discussed issue) or questions and answers.

We illustrate the power of this approach by contrasting the limitation of classical term-based indexing for retrieving relevant content of a conversation. Consider the following conversation excerpt from the ICSI meetings corpus (Janin et al. 2003):

> **1702.95 David:** *so – so my question is should we go ahead and get na- - nine identical head mounted crown mikes*?
> **1708.89 John:** *not before having one come here and have some people try it out*.
> **1714.09 John:** *because there's no point in doing that if it's not going to be any better.*
> **1712.69 David:** *okay.*
> **1716.85 John:** *so why don't we get one of these with the crown with a different headset?*

For a query such as: "*Why was the proposal on microphones rejected*?", a classical indexing schema would retrieve the first turn from David by matching the relevant query term "microphone". There is no presence of other query terms such as "reject" or "proposal". Moreover, it is not possible to map the Why-question onto some query term (e.g. reason, motivation, justification, explanation). This makes it impossible to adequately answer the query without any additional metadata that highlight the argumentative role of the participants' contributions in the conversation.



DISCUSS(issue) <- PROPOSE(alternative)
1702.95 David: so - so my question is should we go ahead and get na- - nine identical head mounted crown mikes ? {qy} 61a

REJECT(alternative)
1708.89 John: not before having one come here and have some people try it out . {s^arp^co} 61b.62a

PROVIDE(justification)
1714.09 John: because there's no point in doing that if it's not going to be any better . {s} 61b+

ACCEPT(justification)
1712.69 David: okay . {s^bk}62b

PROPOSE(alternative)
1716.85 John: so why don't we get one of these with the crown with a different headset ? {qw^cs} 63a

Figure 2: Argumentative Structure of a conversation.

In Figure 2, we show the argumentative structure of the conversation excerpt that allows us to correctly answer the query by selecting the third turn. In fact, the Why-question is mapped to a query term which is found as an argumentative index, "justification", for that turn. Of course, finding justification is not enough, and the retrieval algorithm needs to check whether that justification has been provided as a rejection of a "proposal" (or "alternative") made to an issue on the topic of microphones. This can be done by navigating back through the argumentative chain up to the "issue" episode whose content thematically matches the "microphone" query term.

4. Argument extraction

We have developed a system that computes the argumentative structure of conversations. This system makes it possible to perform argumentative indexing as well as visualizing arguments with the method discussed in section 5 and for generating summaries as described in section 6.

The core of our solution for argument extraction is based on adapting and extending GETARUNS (Delmonte. 2007; 2009), a natural language understanding system developed at the University of Venice. Automatic argumentative annotation is carried out by a special module of GETARUNS activated at the very end of the analysis of each conversation, taking as input its complete semantic representation.

To produce argumentative annotation, the system uses the following 21 discourse relations:

statement, narration, adverse, result, cause, motivation, explanation, question, hypothesis, elaboration, permission, inception, circumstance, obligation, evaluation, agreement, contrast, evidence, hypoth, setting, prohibition.

These are then mapped onto five general argumentative labels:

ACCEPT,

REJECT/DISAGREE,

PROPOSE/SUGGEST,

EXPLAIN/JUSTIFY,

REQUEST.

In addition we use the label DISFLUENCY for all those turns that contain fragments which are non-sentences and are not semantically interpretable. Details of the algorithm are available in (Delmonte et al. 2009), which has been evaluated on conversations from the ICSI meeting corpus (Janin et al. 2003) annotated with argumentative structure during the user study carried out by (Pallotta et al. 2007).

On a total of 2304 turns, 2251 have received an argumentative automatic classification, with a Recall of 97.53%. We computed Precision as the ratio between Correct Argumentative Labels/Found Argumentative Labels, which corresponds to 81.26%. The F-score is 88.65%.

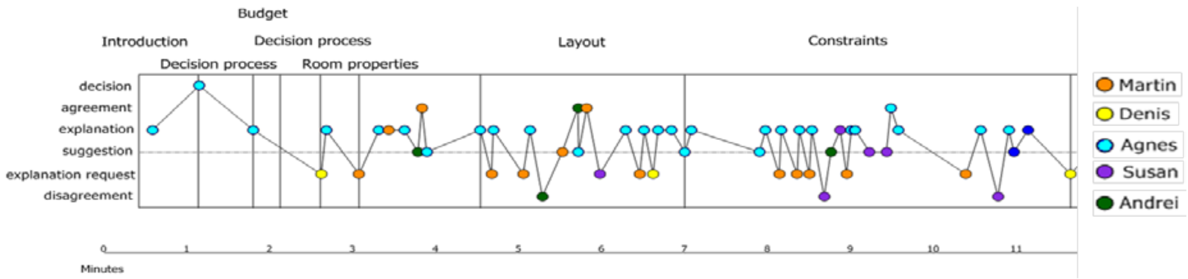


Figure 3: Conversational Graph

5. Visualization of arguments

Argumentative indexing of conversations is relatively novel in the field of information retrieval and visualization, and there are not yet standard metadata schemas for structuring and indexing conversations, or for visualizing their structure (Pallotta et al., 2004; Verbree, 2006). Previous work on visualizing argumentation has mainly been driven by the need for tools to improve argumentation in real-time meetings (Bachler et al., 2003; Fujita et al., 1998; Michaelides et al., 2006; Rienks et al., 2005). Some research has also addressed the use of such visualizations for browsing past conversations, and end user evaluations have been positive (Rienks and Verbree, 2006).

In this paper, we propose a visualization of arguments as *conversation graphs*, to helps users search in digital conversations using argumentative and thematic indexing of conversations. Conversation graphs are diagrams that summarize what topics were discussed, how long they were discussed, which participants were involved in the discussion, and what type of arguments they contributed with (see Figure 3).

An important criterion in the design of the graphs is that the visualization of the argumentative structure has to be intuitive so that users do not need to spend effort on learning the argumentative categories before being able to use them for searching in digital conversations. For this purpose, the graph representation introduces the notion of "positive" and "negative" contributions in discussions. Positive contributions, such as agreements and decisions, are visualized as peaks and negative contributions, such as disagreements, as valleys along the time axis. Suggestions are neutral in polarity and are positioned in the middle.

The goal is that users will use conversation graphs to make more efficient queries about digital conversations by combining topical, argumentative and participant criteria rather than perform pure content-based search. For example if a user wants to find out what objections a participant had about some proposal, the argumentation graph shows that the selected participant disagreed several times during the discussion about that topic. Displaying this information should make it intuitive to search for the relevant meeting episodes by specifying argumentative search criteria rather than simple content-based criteria.

The second aspect of how conversation graphs can be useful in conversational information retrieval is that they can help users to browse the results of their search. When a user opens a conversation transcript and browses through the highlighted sections that correspond to their search criteria, they can compare these highlighted sections with the argumentation points in the graph. By referring to the graph the user can extract information about how many sections of the conversation correspond to their search criteria (in our example, as many as there are disagreements by a participant in the graph). The users may then derive that some, but not necessarily all, of the search results in the transcript are relevant for answering their original question.

First results of user studies have shown that conversation graphs are indeed promising tools both for querying and browsing indexed digital conversations (Ailomaa and Rajman, 2009, Ailomaa, 2009).

6. Abstract Summarization of Conversations

A complementary way to provide users with simplified access to conversational content is by means of summaries. Analyzing and summarizing conversations (or dialogues) is very challenging (Maybury 2007). Many existing summarization techniques are tailored for the narrative genre and can hardly be adapted to the dialogue genre. Moreover, most of the techniques are based on extractive summarization (Zechner 2002; Murray et al. 2005; Garg et al. 2009) that proves to be inadequate for summarizing conversations. In fact, this method has severe limitations due to the intrinsic characteristics of the source data: conversations are not as coherent as ordinary narrative text (such as news or scientific articles) and obtaining a coherent text from conversations is practically impossible using the extractive approach. Moreover, any system that performs extractive summarization must be evaluated against human-annotated test datasets. As pointed out in (Buist et al., 2005), inter-annotator agreement is very low for this type of task, which makes test data nearly useless for evaluation. Intuitively, selecting salient content from conversations is a really difficult task and subjective selection of excerpts leads to fairly different results. In any case, the proposed solutions for extractive summarization of meetings have already reached their qualitative upper bounds as shown by (Riedhammer et al. 2008).

We advocate for abstractive summarization of conversational content. Abstractive summarization of narrative texts is typically based on sentence compression and/or paraphrase (Mani & Maybury 1999). This approach is clearly not appropriate for conversations because turns are already highly compressed. Instead, our abstractive summarization system generates descriptions of the conversation dynamics based on both thematic content and argumentative structure we are able to automatically extract as described in the previous sections.

Our approach differs from that of (Kleinbauer et al. 2007) who provide short abstractive indicative summaries of conversations exclusively based on thematic content.

The output of our system consists instead of several sections, namely describing the conversational settings, participants' number and names, statistics about the interactivity of participants (e.g. the degree of competitiveness), the topics discussed, and the arguments.

Arguments are grouped into episodes bound together by thematic cohesion. All this information is extracted by the system described in section 4. For instance, in more formal meetings we expect to map these episodes to agenda items.

The following is an example of a memo which can be generated with our system from the analysis of ICSI meetings (only turns and speaker with no additional annotations):

MEETING MEMO

GENERAL INFORMATION ON PARTICIPANTS

- The participants to the meeting are 7.
- Participants less actively involved are Ami and Don who only intervened respectively for 38 and 68 turns.

LEVEL OF INTERACTIVITY IN THE DISCUSSION

- The speaker that has held the majority of turns is Adam with a total of 722 turns, followed by Fey with a total of 561.
- The speaker that has undergone the majority of overlaps is Adam followed by Jane.
- The speaker that has done the majority of overlaps is Jane followed by Fey.
- Jane is the participant that has been most competitive.

DISCUSSION TOPICS

The discussion was centered on the following topics:

schemas, action, things and domain.

The main topics have been introduced by the most important speaker of the meeting.

The participant who introduced the main topics in the meeting is: **Adam**.

The most frequent entities in the whole dialogue partly coincide with the best topics, and are the following:

action, schema, things, 'source-path-goal', person, spg, roles, bakery, intention, specific, case, categories, information, idea.

ARGUMENTATIVE CONTENT

The following participants:

Andreas, Dave, Don, Jane, Morgan

expressed their dissent 52 times. However Dave, Andreas and Morgan expressed dissent in a consistently smaller percentage.

The following participants:

Adam, Andreas, Dave, Don, Jane, Morgan

asked questions 55 times.

The remaining 1210 turns expressed positive content by proposing, explaining or raising issues. However Adam, Dave and Andreas suggested and raised new issues in a consistently smaller percentage.

The following participants:

Adam, Andreas, Dave, Don, Jane, Morgan

expressed acceptance 213 times.

EPISODE ISSUE No. 7

In this episode we have the following argumentative exchanges between the following speakers: **Don, Morgan**.

Morgan provides the following explanation:
    [oh, that-s_, good, .]
then he , overlapped by Don, continues:
    [because, we, have, a_lot, of, breath, noises, .]
Don accepts the previous explanation:
    [yep, .]
then he provides the following explanation:
    [test, .]
Morgan continues:
    [in_fact, if, you, listen, to, just, the, channels, of, people, not, talking, it-s_, like, ..., .]
then he , overlapped by Don, disagrees with the previous explanation
    [it-s_, very, disgust, ..., .]
Don, overlapped by Morgan, asks the following question:
    [did, you, see, hannibal, recently, or, something, ?]
Morgan provides the following positive answer:
    [sorry, .]
then he provides the following explanation:
    [exactly, .]
    [it-s_, very, disconcerting, .]
    [okay, .]
…

7. Conclusions

In this article we have presented the core language technology for analyzing digital conversations and producing from their analysis intuitive visualizations and high-quality summaries.

We addressed the issue of capturing the conversational dynamics through the adoption of argumentation theory as the underlying model for making pragmatic analysis of conversations.

We made the case for the importance of such a type of analysis showing how the shortcomings of classical information retrieval techniques can be overcome adopting our approach.

We provided an evaluation of the performance of the current analysis system with respect to the important task of automatically recognizing argumentative force of conversational contributions.

Finally, we presented two applications of our analysis system for the visualization and summarization of conversational data in order to demonstrate the effectiveness of our approach in presenting condensed and distilled conversational data.

7.1 Future Work

We are currently working on a new application that will analyze and summarize (micro)-blog conversations (e.g. Twitter, Google Wave) for online reputation management.

We also expect to start an evaluation campaign for assessing the quality of the abstractive summaries and to investigate how conversational graphs scale in terms of length of conversation and number of participants.

8. References

Ailomaa M. (2009) Answering Questions About Archived, Annotated meetings. PhD thesis N° 4512, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

Ailomaa M. and Rajman M. (2009) Enhancing natural language search in meeting data with visual meeting overviews. In *Proceedings of the 10th Annual Conference of the NZ ACM Special Interest Group on Human-Computer Interaction* (CHINZ 2009), Auckland, New Zealand, 6-7 July.

Bachler M. S., Shum S. J. B., Roure D. C. D., Michaelides D. T., and Page K. R. (2003). Ontological mediation of meeting structure: Argumentation, annotation, and navigation. In *Proceedings of the 1st International Workshop on Hypermedia and the Semantic Web* (HTSW2003), August 26 - 30, 2003, Nottingham, UK.

Buist, A. H., Kraaij W., and Raaijmakers S. (2005). Automatic Summarization of Meeting Data: A feasibility Study. In *proceedings of the 15th CLIN conference*, 2005.

Delmonte R. (2007). *Computational Linguistic Text Processing – Logical Form, Semantic Interpretation, Discourse Relations and Question Answering*, Nova Science Publishers, New York.

Delmonte R. (2009). *Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution*, Nova Science Publishers, New York.

Fujita, K., Nishimoto, K., Sumi, Y., Kunifuji, S., and Mase, K. (1998). Meeting support by visualizing discussion structure and semantics. In *Proceedings of the 2nd International Conference on Knowledge-Based Intelligent Electronic Systems* (KES '98), April 21-23, Adelaide, Australia, volume 1, pages 417-422.

Garg N., Favre B., Reidhammer K., Hakkani-Tür D. (2009). ClusterRank: A Graph Based Method for Meeting Summarization. In *Proceedings of Interspeech 2009*, Brighton, UK.

Janin A., Baron D., Edwards J., Ellis D., Gelbart D., Morgan N., Peskin B., Pfau T., Shriberg E., Stolcke A. and Wooters C. (2003). The ICSI Meeting Corpus. In *Proceedings of IEEE/ICASSP 2003*, 6-10 April 2003, Hong Kong, vol. 1, pp. 364-367.

Kleinbauer T., Becker S., Becker T. (2007). Combining Multiple Information Layers for the Automatic Generation of Indicative Meeting Abstracts. In *Proceedings of the 11th European Workshop on Natural Language Generation* (ENLG07), June 17th-20th, 2007, Dagstuhl, Germany

Mani I. and Maybury M. (eds.) (1999). Advances on Automatic Text Summarization. MIT Press.

Maybury M. (2007). Keynote on Searching Conversational Speech. In *Proceedings workshop on Searching Spontaneous Conversational Speech* part of the ACM SIGIR'07 Conference, 27 July 2007, Amsterdam.

Michaelides D., Buckingham Shum S., Juby, B. Mancini, C. Slack, R. Bachler, M. Procter, R. Daw, M. Rowley, A. Chown, T. De Roure, D. and Hewitt T. (2006). Memetic: Semantic meeting memory. In *Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises* (WETICE 2006), June 26-28, 2006, Manchester, UK, pages 382-387.

Murray G., Renals S. and Carletta J. (2005). Extractive summarization of meeting recordings. In *proceedings of the 9th European Conference on Speech Communication and Technology*, (EUROSPEECH 05) pp. 593-596.

Pallotta V., Ghorbel H., Ballim A., Lisowska A. and Marchand-Maillet S. (2004). *Towards meeting information systems: Meeting knowledge management*. In Proceedings of ICEIS 2005, pp. 464–469, Porto, Portugal.

Pallotta V. (2006), Framing Arguments. In *Proceedings of the International Conference on Argumentation* ISSA, June 2006, Amsterdam, Netherlands.

Pallotta V., Seretan V. and Ailomaa M. (2007). User requirements analysis for Meeting Information Retrieval based on query elicitation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics* (ACL 2007), pages 1008–1015, Prague.

Riedhammer K., Gillick D., Favre B. and Hakkani-Tür D. (2008). Packing the Meeting Summarization Knapsack. In *Proceedings of the 9th International Conference of the ISCA (Interspeech 2008)*, Brisbane, Australia, pages 2434-2437.

Rienks R. and Verbree D. (2006). About the usefulness and learnability of argument-diagrams from real discussions. In *Proceedings of the 3rd International Machine Learning for Multimodal Interaction Workshop* (MLMI 2006), May 1-4, 2006, Bethesda (MD), USA.

Rienks, R., Heylen, D., and van der Weijden, E. (2005). Argument diagramming of meeting conversations. In A. Vinciarelli and J.-M. Odobez, editors, In *Proceedings of the Multimodal Multiparty Meeting Processing Workshop at the 7th International Conference on Multimodal Interfaces* (ICMI 2005) October 3-7, 2005, Trento, Italy, pages 85-92.

Verbree, A. (2006). On the structuring of discussion transcripts based on utterances automatically classified. Master's thesis, University of Twente, The Netherlands.

Zechner K. (2002). Automatic Summarization of Open-Domain Multiparty Dialogues in Diverse Genres. *Computational Linguistics* Volume 28, Number 4, 2002.

Author Index