3rd International Workshop on PErsonalization in eGOVernment and Smart Cities (PEGOV): Smart Services for Smart Territories

Nikolaos Loutas[1], Fedelucio Narducci[5], Adegboyega Ojo[3], Matteo Palmonari[2], Cécile Paris[4], and Giovanni Semeraro[5]

[1] PwC, Belgium
nikolaos.loutas@pwc.be
[2] Department of Informatics, Systems and Communication
University of Milano-Bicocca, Italy
palmonari@disco.unimib.it
[3] Insight Centre for Data Analytics, National University of Ireland, Galway, Ireland
adegboyega.ojo@deri.org
[4] CSIRO, Australia
Cecile.Paris@csiro.au
[5] University of Bari "Aldo Moro", Italy
name.surname@uniba.it

1 Preface

User modeling and personalization have been playing an important role in the development of intelligent systems, whereby these systems adapt their behavior based on knowledge about users. Such knowledge can include users characteristics, interests and preferences, as well as locations or past behaviors. While personalization has been extensively studied and employed in domains characterized by the digital-object consumption (e-commerce, news, music, video recommendations, etc.), personalization in eGovernment applications is still in its infancy.

e-Government (e-Gov) has transformed interactions between governments, citizens and other stakeholders in the society. Public services and public sector information can now be delivered electronically through Web portals and mobile apps (e.g., see Palmonari et al., 2008, Loutas et al., 2011, Lee et al., 2011, Narducci et al., 2014). In this new context, citizens are the intended users of public services, thus innovative solutions better tailored to citizens' needs can facilitate access to e-Gov services and reduce the red tape that often characterizes the provisioning of public services (e.g., Bianco et al., 2013, Bista et al., 2013, Castelli et al., 2014). It can also enable more targeted information to be delivered to citizens (e.g., Colineau et al., 2013), or helps to overcome the language barrier for accessing to public services in different countries (Narducci et al., 2013). Finally, governments have also started to look at ways to better engage with citizens, both for service delivery and for policy making (e.g., Lee et al., 2011). As a result of these initiatives, providing personalized services, often grouped in life-events and business episodes, is a real possibility now for governments.

Another interesting development is the recent push towards more openness of public sector information, with an emphasis on opening up government data (Ojo et al., 2015), which presents new application areas and opportunities for personalization. This trend has specifically created the need for personalized access to Open Government Data predominantly by means of visualizations and faceted browsers. It has also given rise to opportunities for improved decision making (e.g., Lee et al., 2011), as well as recommendation and personalization of e-Gov services (e.g., Loutas et al., 2011, Baldassarre et al., 2013).

This introduces new challenges for personalization models. On the one hand, personalization can lead to better services and more relevant information. This is seen as desirable, by both the public and governments, as it can improve service delivery (e.g., Colineau et al., 2013, Iaquinta et al., 2013, Penadés et al., 2014, Torsello et al., 2014, Vicente-López et al., 2014) and participation to decision-making processes (e.g., Ardissono et al., 2013, Ardissono et al., 2014) . On the other hand, there are potentially ethical (including privacy) issues related to the fact that citizens might be in a dependence relationship with governments (e.g., Paris et al., 2013), and automatic user profiling might be considered big brother and not desirable.

Personalization in the e-Gov domain is still fairly novel at least in production systems, potentially because of the difficulties to obtain some of the information required for personalization, because of privacy, confidentiality, ethical and potentially trust reasons.

The main goal of this workshop is to stimulate the discussion around problems, challenges and research directions about personalization in e-Gov. This workshop builds on two previous PEGOV workshops at UMAP (2013 and 2014). This year, we extended the scope of PEGOV to Smart Cities, as these also provide new opportunities and new challenges (e.g., Villena-Román et al., 2014, Ojo et al., 2015). Smart Cities can have access to very detailed data about the citizens, e.g., using urban sensing devices, which can support new personalization models.

Improving the quality of both life and services in the city have high relevance in many research fields such as Social Sciences, Psychology, Education, Medicine, and Computer Science. For these reasons Smart Cities are becoming a very interesting topic for different conferences belonging to the ICT area. Different aspects are generally analyzed. For example, the Semantic Smart Cities was the subject of several recent workshops (the Semantic Cities workshop at AAAI 2012 and IJCAI 2013, the Semantic Smart City workshop at WIMS 2013, and the Smart Semantic Cities workshop at AI*IA 2014). Other events analyzed aspect such as the designing of Web Applications for Smart Cities (AW4CITY at WWW2015), or the Web Data Science at the service of Smart Cities (Web Data Science meets Smart Cities at WWW 2015). However, the personalization aspect for designing, implementing and delivering personalized services for new citizen-centered Smart Cities and Territories is not yet properly investigated in the literature.

The original topics of interest listed on the call for paper for the workshop included:

- Motivations, benefits, and issues of personalization in e-Gov and Smart Cities
- Approaches for the personalization of inclusive, personal and interactive services to citizens
- User and context awareness in personalization of services to the citizens
- Multilingual services to citizens
- Adaptation, personalization and recommendation models and goals in city services
- User, group and family modeling in e-Gov and Smart Cities
- Mining of user behavior, opinion mining, and sentiment analysis in e-Gov and Smart Citizens
- Gamification and Crowdsourcing for mining citizens profiles and opinions
- Services for personalized access to (Linked) Open Government Data
- Persistence, removal, and update of citizen profiles
- Semantic techniques for user profiling and personalization in e-Gov and Smart Cities
- Ethical issues, including privacy, in e-Gov and Smart Cities
- Usability of services to citizens
- Evaluation of personalized services in e-Gov and Smart Cities
- Applications of personalization methods in e-Gov and Smart Cities
- Communities and social networks in participatory e-Gov and Smart Cities
- Citizen-centered service design and modelling
- E-health and Smart Health

We accepted two short papers and one long paper. Each submission was reviewed by at least two PC members (none of the chairs has been involved in the review process). We are also pleased to have an invited speaker, Dr Edward Curry, the leader of the Green and Sustainable IT research group at Digital Enterprise Research Institute (DERI) in Ireland.

One paper proposes a personalization model for e-government, a personalized extended government model, to simplify and improve the effectiveness of e-government services. Another paper proposes a methodology for personalized cultural information. Finally, the third paper presents an environment for constraint-based recommender systems that could be use in e-government, for example as an online advisory service for citizens.

Edward Curry's invited talk is about *Open Data Innovation in Smart Cities: Challenges and Trends*. Open Data initiatives are increasingly considered as defining elements of emerging smart cities. However, few studies have attempted to provide a better understanding of the nature of this convergence and the impact on both domains. The talk examines the challenges and trends with open data initiatives using a socio-technical perspective of smart cities. The talk presents findings from a detailed study of 18 open data initiatives across five smart cities to identify emerging best practice. Three distinct waves of open data innovation for smart cities are discussed. The talk details the specific impacts of open data innovation on the different smart cities domains, governance of the cities, and the nature of datasets available in the open data ecosystem within smart cities.

We hope the workshop will stimulate discussion around problems, challenges and research directions about personalization in governments and Smart Cities, with a specific focus on the design of personalized citizen-centered services and the challenges that must be addressed.

Some questions that motivate this workshop and that we hope we will be discussed during the workshop:

1. Can personalization methods support the design of services and applications, which better adapt to the different roles of citizens and companies?
2. Which user characteristics (demographic, cultural, family, etc.) can influence the design and delivery of personalized services for Smart Cities and Territories?
3. How can citizens be involved in the design of adaptive service platforms in different domains (e-gov, e-health, public services, etc.)?
4. Are the general techniques adopted for user modeling and profiling in different domains exploitable for modeling the citizen characteristics?
5. What services can be useful for a patient-empowered Smart Health?
6. How privacy and ethically issues affect the feasibility of effective personalization methods in the Smart Environments?
7. Can semantic models and ontologies support the representation of prototypical users in order to identify categories of citizens based on different characteristics?
8. How can service personalization decrease the costs for public administrations, increasing at the same time the value delivered to the citizen?
9. Would personalization methods be favorably accepted and desired by citizens?
10. How can ethical issues (big brother) and privacy influence the trust in personalized services?

This is an exciting field full of opportunities.

2 Workshop Chairs

Nikolaos Loutas, PwC, Belgium.
Nikolaos is manager at PwC's Technology Consulting practice, involved mainly in projects on interoperability of trans-European ICT solutions, data and software products. Nikolaos specialises in semantic aspects of interoperability, through the application of Semantic Web technologies and Linked Data. He has deep insights into open semantic standards, such as the Asset Description Metadata Schema, the e-Government Core Vocabularies and the DCAT Application Profile for data portals in Europe. Nikolaos is currently driving the Open Data Support project of DG CONNECT, which aims at facilitating the access of citizens and business to Open Government Data published by governments across Europe. Before joining PwC, Nikolaos had been working for leading EU research centers.

He has published more than 55 papers and reports in the field of Semantic Web in international journals, conferences and books.

Fedelucio Narducci, SWAP Research Group, University of Bari Aldo Moro, Italy
Fedelucio Narducci is research assistant at University of Bari Aldo Moro, Department of Computer Science. and member of the SWAP (Semantic Web Access and Personalization) research group. His primary research interests lie in the areas of machine learning, content-based recommender systems, user modeling, and personalization. From April 2012 he is working for the SMART (Services & Meta-services for smART eGovernment) project whose goal is to define models, methodologies, languages for planning, production and delivery of services characterized by optimal social value, value of use, and value of exchange. He served as Co-chair of Pegov 2013. Fedelucio was reviewer and co-reviewer for international conferences and journals on the topics of recommender system, user modeling and personalization. He is also author of several papers in international conferences and journals.

Adegboyega Ojo, INSIGHT Center for Data Analytics, National University of Ireland, Galway
Adegboyega Ojo is a Research Fellow and leads the E-Government Group at The INSIGHT Center for Data Analytics, National University of Ireland, Galway; Republic of Ireland. His research focuses on how to drive innovations in government organizations through the applications of Semantic Web, Linked Open Data and Collaboration technologies. His current portfolio of research and development projects is funded under the Seventh Framework Programme of the European Commission. Before his current role, he worked as Academic Program Officer, Research Fellow and Post-doctoral Fellow at the Center for Electronic Governance, United Nations University ? International Institute for Software Technology (UNU). At UNU, his work benefitted several governments including Macao, Korea, Mongolia, Colombia, Cameroon and Nigeria. He has published widely in the areas of Strategies, Architecture and Standards, e-Participation, Open Governance and Open Data. He obtained his PhD at the University of Lagos, Nigeria (1998), where he was appointed Senior Lecturer and Associate Professor in Computer Science in 2003 and 2012 respectively. He is also Adjunct Lecturer at the National University of Ireland, Galway.

Matteo Palmonari, University of Milano-Bicocca, Italy
Matteo Palmonari is an assistant professor in the Department of Informatics, Systems and Communication at the University of Milan-Bicocca. His research interests include semantic matchmaking, information quality, knowledge representation, and ontologies for the semantic web; several of his research have been applied to service modeling, service matchmaking and e-Government applications. He has been a visiting postdoc and a visiting assistant professor with the ADVIS Laboratory, University of Illinois at Chicago. He has published more than 40 papers in international journals and conferences.

Cécile Paris, CSIRO, Computational Informatics, Australia
Dr Cécile Paris is a Science Leader at CSIRO, Sydney, Australia, leading a re-

search group on Knowledge Discovery and Management. Dr Paris also holds Adjunct Professorships at Macquarie University (Sydney) and the ANU (Australian National University, Canberra, Australia). Dr Paris received her B.A. degree in Computer Science from The University of California at Berkeley, USA, and her Masters and PhD degrees from Columbia University, New York, USA. Her PhD was one User Modeling and Natural Language Generation. Her main research interests lie in the areas of personalized information delivery and language technology. She has been involved in e-Government for over 5 years, and her current work includes tailored delivery for Public Administration, online communities and social media in the context of e-Government. Dr Paris co-organised the workshop on Government and Citizen Engagement at the Communities and Technology conference in 2011. In 2011, she was an invited speaker at the 2nd (Australian) Public Officer Digital Media Forum, and at the 7th Annual AIMIA Digital Summit (AIMIA is the Australian Interactive Media Association). She was a keynote speaker at the (Australian) Emergency Management New and Emerging Technologies Forum in October 2013 and at the National Medicine Symposium in May 2014. Dr Paris has authored over 250 referred technical articles at international journals and conferences. She is currently the chair of CHISIG, the Computer Human Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia.

Giovanni Semeraro, SWAP Research Group, University of Bari Aldo Moro, Italy
He is associate professor of computer science at the University of Bari Aldo Moro and leads the Semantic Web Access and Personalization Research Group Antonio Bello. His research interests include AI, recommender systems, user modeling, personalization, intelligent information retrieval, semantic and social computing, the Semantic Web, natural language processing, and machine learning. He received his M.Sc. degree in computer science from the University of Bari. He served as General Co-chair of UMAP 2013, IIR 2013, SemExp 2012, IIR 2012, IIA 2008, AI*IA 2008, SWAP 2007, CILC 2006, and as Program Co-chair of RecSys 2015, IntRS@RecSys 2015 & 2014, DeCAT@UMAP 2015, PeGOV@UMAP 2014, Decisions@RecSys 2013 & 2012, DART 2013, 2012 & 2011, RSmeetDB@DEXA 2013 & 2012, SeRSy@RecSys 2013 & SeRSy@ISWC 2012, DEMRA@UMAP 2011, SPIM@ISWC 2011, EC-Web 2010, SWAP 2010, Web Mining 2.0@ECML/PKDD 2007, ISMIS 2006, WebMine@ECML/PKDD 2006, IEA-AIE 2005. He is co-author of more than 350 papers published in journals, international conferences and workshops.

References

1. Ardissono, L., Voghera, A., & Velluto, M. (2013). Selecting People for Involving Them in Participatory Decision-Making Processes. In PEGOV 2013: Personalization in eGovernment Services and Applications, co-located with UMAP'13 (Vol. 997). CEUR-WS.
2. Ardissono, L., Lucenteforte, M., Savoca, A., & Voghera, A. (2014). Community Mapping for Participatory Decision-Making Processes. In PEGOV 2014: Personal-

ization in eGovernment Services, Data and Applications, co-located with UMAP'14 (Vol. 1181). CEUR-WS.

3. Baldassarre, C., Cremaschi, M. & Palmonari, M. (2013). Bridging the Gap between Citizens and Local Administrations with Knowledge-Based Service Bundle Recommendations. In Morvan, F., Tjoa, A.M., & Wagner, R. (Eds.): 24th International Workshop on Database and Expert Systems Applications, DEXA 2013. IEEE 2013: 157-161
4. Bianco, A., Campegiani, P., Forrester, J., Iaquinta, L., & Torsello, M. A. (2013). SMART project: Industrial and Academic Collaboration for Service Design. In PEGOV 2013: Personalization in eGovernment Services and Applications, co-located with UMAP'13 (Vol. 997). CEUR-WS.
5. Bista, S. K., Nepal, S. & Paris, C. (2013). The Human Touch of Government Services. In PEGOV 2013: Personalization in eGovernment Services and Applications, co-located with UMAP'13 (Vol. 997). CEUR-WS.
6. Castelli M., Comerio M., & Cremaschi M. (2014). Towards the Definition of Value-added Services for Citizens: a New Model for the Description of Public Administration Services. International Journal of Management & Information Technology 4(1):166-173, 2013. ISSN 1460-6720.
7. Colineau, N., Paris, C. & Vander Linden, K. (2013). Automatically Producing Tailored Web Materials for Public Administration. In New Review of HyperMedia and MultiMedia, 9(2): 158-181.
8. Iaquinta, L., Torsello, M. A., Comerio, M., Fanelli, A. M., & Semeraro, G. (2013). User Segmentation in e-Government Services. In PEGOV 2013: Personalization in eGovernment Services and Applications, co-located with UMAP'13 (Vol. 997). CEUR-WS.
9. Lee D., Loutas N., Sanchez Nielsen E., Mogulkoc E., & Lacigova O. (2011). Inform-Consult-Empower: A Three-tiered Approach to eParticipation. In Proceedings of 3rd ePart Conference, Delft, The Netherlands.
10. Loutas N., Peristeras V., & Tarabanis K. (2011). A citizen-centric semantically-enhanced governmental portal. In Electronic Government, An International Journal, vol. 8 (4), pp. 363-384.
11. Narducci F., Palmonari M., & Semeraro G. (2014). Cross-Language Semantic Retrieval and Linking of E-Gov Services. In The Semantic WebISWC 2013. Springer Berlin Heidelberg, 2013, pp.130-145.
12. Narducci F., Palmonari M., & Semeraro G. (2013). A Cross-language Semantic Matching for Discovering Links to E-Gov Services in the LOD Cloud In KNOW@ LOD. 2013, pp. 21-32.
13. McCleary, R. (2007). Ethical Issues in Online Social Work Research. In Journal of Social Work Values & Ethics, Volume 4, Issue 1, pp.6-26.
14. Ojo, A., Curry, E., & Zeleti, F. A. (2015). A Tale of Open Data Innovations in Five Smart Cities. In System Sciences (HICSS), 2015 48th Hawaii International Conference on (pp. 2326-2335). IEEE.
15. Palmonari, M., Viscusi, G. & Batini., C. (2008). A semantic repository approach to improve the government to business relationship. Data Knowl. Eng. 65(3): 485-511 (2008)
16. Paris, C., Colineau, N., Nepal, S., Bista, S. K., & Beschorner, G. (2013). Ethical considerations in an online community: the balancing act. In Ethics and Information Technology. 15(4): 301-316.
17. Penads, M. C., Martí, P., Canós, J. H., & Gmez, A. (2014). Product Line-based customization of e-Government documents. In PEGOV 2014: Personalization in

eGovernment Services, Data and Applications, co-located with UMAP'14 (Vol. 1181). CEUR-WS.
18. Peristeras, V., Goudos, S. K., Loutas, N., & Tarabanis, K. (2009). Ontology-based search for eGovernment services using citizen profile information. Journal of Web Engineering, 8(3), 245.
19. Torsello, M. A., Iaquinta, L., & Comerio, M. (2014). A Fuzzy Model for Service Value Assessment. In PEGOV 2014: Personalization in eGovernment Services, Data and Applications, co-located with UMAP'14 (Vol. 1181). CEUR-WS.
20. Vicente-López, E., de Campos, L. M., Fernndez-Luna, J. M., & Huete, J. F. (2014). Personalization of parliamentary document retrieval using different user profiles. In PEGOV 2014: Personalization in eGovernment Services, Data and Applications, co-located with UMAP'14 (Vol. 1181). CEUR-WS: 28-37.
21. Villena-Román, J., Luna-Cobos, A., & González-Cristóbal, J. C. (2014). Tweet-Alert: Semantic Analytics in Social Networks for Citizen Opinion Mining in the City of the Future. In PEGOV 2014: Personalization in eGovernment Services, Data and Applications, co-located with UMAP'14 (Vol. 1181). CEUR-WS: 7-11.

A Wiki-based Environment for Constraint-based Recommender Systems Applied in the E-Government Domain

Stefan Reiterer[1], Alexander Felfernig[1], Michael Jeran[1], Martin Stettinger[1], Manfred Wundara[2], and Wolfgang Eixelsberger[3]

[1] Institute for Software Technology, A-8010 Graz, Inffeldgasse 16b
firstname.lastname@ist.tugraz.at
[2] City of Villach, A-9500 Villach, Rathausplatz 1
manfred.wundara@villach.at
[3] Carinthia University of Applied Sciences, A-9524 Villach, Europastraße 4
w.eixelsberger@fh-kaernten.at

Abstract. Constraint-based recommenders support customers in identifying relevant items from complex item assortments. In this paper we present WeeVis, a constraint-based environment that can be applied in different scenarios in the e-government domain. WeeVis supports collaborative knowledge acquisition for recommender applications in a MediaWiki-based context. This paper shows how Wiki pages can be extended with recommender applications and how the environment uses intelligent mechanisms to support users in identifying the optimal solutions to their needs. An evaluation shows a performance overview with different knowledge bases.

1 Introduction

Constraint-based recommender applications help users navigating in complex product and service assortments like digital cameras, computers, financial services and municipality services. The calculation of the recommendations is based on a knowledge base of explicitly defined rules. The engineering of the rules for recommender knowledge bases (for constraint-based recommenders) is typically done by knowledge engineers, mostly computer scientists [1]. For building high quality knowledge bases there are domain experts involved who serve the knowledge engineers with deep domain knowledge [1]. Graphical knowledge engineering interfaces like [2] improved the maintainability and accessability and moved the field one step further.

Other recommendation approaches like collaborative filtering use information about the rating behavior of other users to identify recommendations [3, 4]. Content-based filtering [5] exploits features of items for the determination of recommendations. Compared to these approaches, constraint-based recommenders are more applicable for complex products and services due to their explicit knowledge representation.

In the line of Wikipedia[4] where users build and maintain Wiki pages collaboratively we introduce WeeVis[5]. WeeVis is a MediaWiki[6] based environment that exploits the properties of MediaWiki and enables community based development and maintenance of knowledge bases for constraint-based recommenders. WeeVis is freely available as a platform and successfully applied by four Austrian universities (in lectures about recommender systems), in the financial services domain and in e-government .

In the e-government domain officials as well as the community residents can take numerous advantages of knowledge-based recommenders:

- WeeVis can be used as an online advisory service for citizens for example for documents that are necessary to apply for a private construction project. The online recommendation of necessary documents in advance to on-site appointments can lead to a time reduction for community residents and community officials.
- WeeVis can be used for modeling internal processes like the signing of travel applications for example a community official wants to visit a conference, based on different parameters like the conference type, or if it's abroad or in the domestic area, different officials have to sign the travel request. In WeeVis the appropriate rules for such internal processes can be mapped and especially for new employees WeeVis recommenders can provide substantial assistance.
- WeeVis can be used as an information platform for example with integrated knowledge-based recommenders for community residents e.g. to identify the optimal waste disposal strategy for a household (this example is used as a running example in this paper, see Section 2). Instead of providing plain text information, like common municipality web pages, the knowledge representation as a recommender provides an easier way for community members to identify the optimal solution for their situation.

A recommender development environment for single users is introduced in [2]. This work is based on a Java platform and focuses on constraint-based recommender applications for online selling. Compared to [2], WeeVis provides a wiki-based user interface that allows user communities to develop recommender applications collaboratively. Instead of an incremental dialog, where the user answers one question after the other, like [2], WeeVis provides an integrated interface where the user is free to answer questions in any order.

The WeeVis interface also provides intelligent mechanisms for an instant presentation of alternative solutions in situations where it is not possible to find a solution for a given set of user (customer) requirements, i.e., the requirements are inconsistent with the recommendation knowledge base and the user is in the need for repair proposals to find a way out from the *no solution could be found dilemma*. Model-based diagnosis [6] can be applied for the identification

[4] www.wikipedia.org
[5] www.weevis.org
[6] www.mediawiki.org

of faulty constraints in a given set of customer requirements. In this context efficient divide-and-conquer based algorithms [7] can be applied to the diagnosis and repair of inconsistent requirements. The environment supports the user with integrated model-based diagnosis techniques [6,8]. A first approach to a conflict-directed search for hitting sets in inconsistent CSP definitions was introduced by [9]. With regard to diagnosis techniques, WeeVis is based on more efficient techniques that make the environment applicable in interactive settings [8, 10].

A Semantic Wiki-based approach to knowledge acquisition for collaborative ontology development is introduced, for example, in [11]. Compared to [11], WeeVis is based on a recommendation domain specific knowledge representation (in contrast to ontology representation languages) which makes the definition of domain knowledge more accessible also for domain experts.

The remainder of this paper is organized as follows. In Section 2 we present an overview of the recommendation environment WeeVis and it's application in the e-government domain. In Section 3 we present results of a performance evaluation that illustrates the performance of the integrated diagnosis technologies. With Section 4 we conclude the paper.

2 WeeVis Overview

Since WeeVis is based on the MediaWiki platform, it can be installed on freely available web servers. On the website *www.weevis.org* a selection of different WeeVis recommenders is publicly available. For internal processes WeeVis can be deployed in the local intranet. Standard wiki pages can be complemented easily by recommender knowledge bases. Currently, WeeVis calculates recommendations based on previously entered requirements. If the requirements would result in a *no solution could be found message* WeeVis calculates alternative solutions based on diagnoses (see Section 2.4). In line with the Wiki idea, WeeVis provides the ability to build knowledge bases collaboratively, a valuable feature in e-government domain, because depending on the community department multiple people are responsible for data management and administration. Furthermore, WeeVis exploits the basic functionalities provided by MediaWiki and allows rapid prototyping processes where the result of a change can immediately be seen by simply switching from the *edit mode* to the corresponding *read mode*. This approach allows an easy understanding of the WeeVis tags and also of the semantics of the provided WeeVis language.

2.1 WeeVis User Interface

Since WeeVis is a MediaWiki-based environment the user interface relies on the common Wiki principle of the *read* mode (see Figure 1) for executing a recommender and the *write* mode (see Figure 2) for defining a recommender knowledge base. The development and maintenance of a knowledge base is supported a textual fashion with a syntax that is similar to the standard Wiki syntax (see Figure

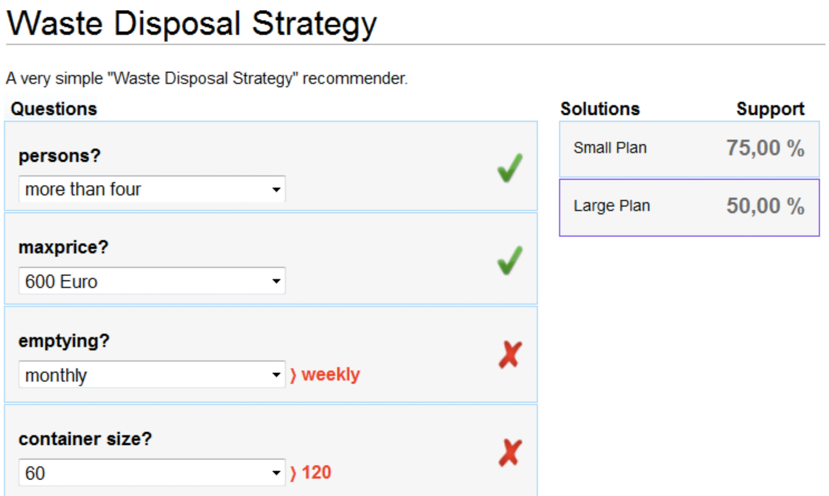

Fig. 1. *Waste Disposal Strategy* a simple recommender knowledge base from the *e-government* domain (WeeVis *read* mode).

2). In the following we will present the concepts integrated in the WeeVis environment on the basis of a working example from the e-government domain. More specifically we present a recommender that supports households in identifying their optimal waste disposal strategy. In this recommendation scenario, a user has to specify his/her requirements regarding, for example, the number of persons living in the household or how frequently the containers should be emptied. A corresponding WeeVis user interface is depicted in Figure 1. Requirements are specified on the left hand side and the corresponding recommendations for the optimal waste disposal plan are displayed in the right hand side.

For each solution, a so-called support score is determined. If a solution fulfills all requirements, this score is 100%, otherwise it is lower and, when clicking on the score value, a corresponding repair action is displayed on the left-hand side (see Figure 1). Due to the automated alternative determination, WeeVis is always able to present a solution and users are never ending up in the *no solution could be found* dilemma (see Figure 1).

An example of the definition of a (simplified) e-government recommender knowledge base is depicted in Figure 2. The definition of a recommender knowledge base is supported in a textual fashion on the basis of a syntax similar to MediaWiki. Basic syntactical elements provided in WeeVis will be introduced in the next subsection.

2.2 WeeVis Syntax

A WeeVis recommender consists of three necessary aspects, the definition of questions and possible answers, items and their properties, and constraints (see Figure 2).

Editing „Waste Disposal Strategy“

```
A very simple "Waste Disposal Strategy" recommender.
<recommender>
  &PRODUCTS
    {!name!sizep!emptyingp!#pricep!
     |Small Plan|60|monthly|300|
     |Medium Plan|120|monthly|300|
     |Large Plan|120|weekly|400|
     }
  &QUESTIONS
    {|persons?(one to two, three to four, more than four)|
     |maxprice?#(200,800,200, Euro)|
     |emptying?(monthly, weekly)|
     |container size? (120, 60)|}
  &CONSTRAINTS
    {
     |persons? = more than four &INCOMPATIBLEWITH container size? = 60|
     |persons? = more than four &INCOMPATIBLEWITH emptying? = monthly|
     |maxprice? >= #pricep|
     |&IF container size? = 120 &THEN sizep = 120|
     |&IF container size? = 60 &THEN sizep = 60|
     |&IF container size? = 120 &THEN sizep <> 60|
     |&IF emptying? = monthly &THEN emptyingp = monthly|
     |&IF emptying? = weekly &THEN emptyingp = weekly|}
  &TEST
    {|show| persons? = more than four,container size? = 60, emptying? =
monthly|}
</recommender>
```

Fig. 2. The *Waste Disposal Strategy* recommender knowledge base (*view source (edit)* mode).

The definition of an item assortment in WeeVis starts with the *&PRODUCTS* tag (see Figure 2). The first line represents the attributes separated by the exclamation mark. In our example, the item assortment is specified by the *name*, *sizep*, the container size, *emptyingp*, the emptying frequency, and *pricep*, the price of the waste disposal plan. Each of the next lines represents an item with the values related to the attributes, in our example there are three items specified: *Small Plan*, *Medium Plan*, and *Large Plan*.

The second aspect starts with the &QUESTIONS tag. In our example the following user requirements are defined: *persons*, specifies the number of persons living in the household (one to two, three to four, more than four) and *maxprice* specifies the upper limit regarding the price of the waste disposal plan. Furthermore, *emptying* represents the sequence in which the dustbins will be emptied, weekly or monthly, and *container size*, the preferred size of the dust container, 120 or 60.

The third aspect represents the definition of the constraints. Starting with the &CONSTRAINTS tag in WeeVis different types of constraints can be defined. For the first constraint in our example the &INCOMPATIBLE keyword is used to describe incompatible combinations of requirements. The first incompatibility constraint describes an incompatibility between the number of persons in the household (*persons*) and the *container size*. For example, a waste disposal plan with (*container size*) 60 must not be recommended to users who live in a household with more than four persons. Filter constraints describe relationships between requirements and items, for example, $maxprice \geq pricep$, i.e., the price of a waste disposal plan must be equal or below the maximum accepted price.

2.3 Recommender Knowledge Base

A recommendation knowledge base can be represented as a CSP (Constraint Satisfaction Problem) [12] on a formal level. The CSP has two sets of variables V ($V = U \cup P$) and the constraints $C = PROD \cup COMP \cup FILT$ where $u_i \in U$ are variables describing possible user requirements (e.g., *persons*) and $p_i \in P$ are describing item properties (e.g., *emptyingp*). Each variable v_i has a domain d_j of values that can be assigned to the variable (e.g., *one to two*, *three to four* or *more than four* for the variable *persons*). Furthermore, there are three different types of constraints:

- $COMP$ represents incompatibility constraints of the form $\neg X \vee \neg Y$
- $PROD$ the products with their attributes in disjunctive normal form (each product is described as a conjunction of individual product properties)
- $FILT$ the given filter constraints of the form $X \rightarrow Y$

The knowledge base specified in Figure 2 can be transformed into a constraint satisfaction problem where &QUESTIONS represents U, &PRODUCTS represents P and &CONSTRAINTS represents $PROD$, $COMP$, and $FILT$. Based on this knowledge representation WeeVis is able to determine recommendations that take into account a specified set of user requirements. The results collected are represented as unary constraints ($R = \{r_1, r_2, ..., r_k\}$). Finally the determined set of solutions (recommended items) is presented to the user.

2.4 Diagnosis and Repair of Requirements

In situations where requirements $r_i \in R$ (unary constraints defined on variables of U such as $emptying = monthly$) are inconsistent with the constraints in C, we are interested in a subset of these requirements that should be adapted to be able to restore consistency. On a formal level we define a *requirements diagnosis task* and a corresponding *diagnosis* (see Definition 1).

Definition 1 (Requirements Diagnosis Task). Given a set of requirements R and a set of constraints C (the recommendation knowledge base), the requirements diagnosis task is to identify a minimal set Δ of constraints (the diagnosis) that has to be removed from R such that $R - \Delta \cup C$ is consistent.

As an example $R = \{r_1 : persons = morethanfour,\ r_2 : maxprice = 600,\ r_3 : emptying = monthly, r_4 : containersize = 60\}$ is a set of requirements inconsistent with the defined recommendation knowledge. The recommendation knowledge base induces two minimal conflict sets (CS) [7] in R which are $CS_1 : \{r_1, r_4\}$ and $CS_2 : \{r_1, r_3\}$. For these requirements we can derive two diagnoses: $\Delta_1 : \{r_3, r_4\}$ and $\Delta_2 : \{r_1\}$. For example, to achieve consistency of Δ_1 at least

r_3 and r_4 have to be adapted. Such diagnoses can be determined on the basis of a HSDAG (hitting set directed acyclic graph) (e.g. [13]).

Determining conflict sets [7] at first and afterwards constructing a HSDAG (hitting set directed acyclic graph) to identify diagnoses tends to become inefficient especially in interactive settings. Direct diagnosis algorithms like FastDiag [8] reduce this two-step process to one step by calculating diagnoses directly without conflict determination. This was the major motivation for integrating FastDiag [8] into the WeeVis environment. Like QuickXPlain [7], FastDiag is based on a divide-and-conquer approach that enables the calculation of minimal diagnoses without the calculation of conflict sets. In WeeVis the derived diagnosis are used as a basis for determining repair actions, which lead to the alternative solutions that are be presented to the user. A repair action is a concrete change of one or more user requirements in R on the basis of a diagnosis such that the resulting R' is consistent with C.

Definition 2 (Repair Task). Given a set of requirements $R = \{r_1, r_2, ..., r_k\}$ inconsistent with the constraints in C and a corresponding diagnosis $\Delta \subseteq R$ ($\Delta = \{r_l, ..., r_o\}$), the corresponding repair task is to determine an adaption $A = \{r'_l, ..., r'_o\}$ such that $R - \Delta \cup A$ is consistent with C.

In WeeVis, repair actions are determined conform to Definition 2. For each diagnosis Δ determined by FastDiag, the corresponding solution search for $R - \Delta \cup C$ returns a set of alternative repair actions (represented as adaptation A). In the following, all solutions that satisfy $R - \Delta \cup A$ are shown to the user (see the right hand side of Figure 1).

Diagnosis determination in FastDiag is based on a total lexicographical ordering of the customer requirements [8]. This ordering is derived from the sequence of the entered requirements. For example, if $r_1 : persons = morethanfour$ has been entered before $r_3 : emptying = monthly$ and $r_4 : containersize = 60$ then the underlying assumption is that r_3 and r_4 are of lower importance for the user and thus have a higher probability of being part of a diagnosis. In our working example $\Delta_1 = \{r_3, r_4\}$. The corresponding repair actions (solutions for $R - \Delta_1 \cup C$) is $A = \{r'_3 : emptying = weekly, r'_4 : containersize = 120\}$, i.e., $\{r_1, r_2, r_3, r_4\} - \{r_3, r_4\} \cup \{r'_3, r'_4\}$ is consistent. The item that satisfies $R - \Delta_1 \cup A$ is $\{LargePlan\}$ (see in Figure 2). The identified items (p) are ranked according to their support value (see Formula 1).

$$support(p) = \frac{\#adaptions\ in\ A}{\#requirements\ in\ R} \quad (1)$$

3 Performance Evaluation

3.1 Description of the evaluation

We have conducted a performance evaluation with the goal to highlight the ability of WeeVis to calculate repair actions and if no solutions could be found. Therefore we set up an experiment with three WeeVis recommenders based on the e-government example presented in Section 2. To illustrate the performance of WeeVis, the knowledge base was extended and deployed with different complexity regarding the number of solutions (&PRODUCTS tag in WeeVis), user requirements (&QUESTIONS tag WeeVis), and constraints (&CONSTRAINTS tag WeeVis) (see Table 1). According to these three attributes the knowledge bases were classified as *Small*, *Medium*, and *Large*. To fit the attributes of knowledge base *Small* from Table 1, the running example (see Figure 2) was adapted by adding one question, two products and removing the last two constraints. The *Medium* and *Large* knowledge base are extended versions of the running example.

Knowledge base	**Number of solutions / requirements / constraints**
Small	5/5/5
Medium	20/10/15
Large	50/15/30

Table 1. The different knowledge bases with sizes *Small*, *Medium* and *Large* used for the performance comparison.

3.2 Results of the evaluation

To provide an optimal user experience a focus of WeeVis is to provide instant feedback after every interaction. Interacting with a WeeVis recommender starts with the the entering of new requirements and the subsequent calculation of solutions for these requirements. If no solution could be found WeeVis calculates one or more diagnoses and the complementing alternative products. With this performance evaluation we show that WeeVis can identify at least one alternative solution even for large knowledge bases within recommended user interface response times [14]:

- below 100ms, the user feels that the system reacts instantaneously
- 1,000ms is the upper limit for keeping the users thought uninterrupted
- 10,000ms is the upper limit for keeping the user's focus on the dialogue

For the first performance evaluation the goal was to measure the time needed for calculating the corresponding solutions to given requirements. After assigning answers to the questions for the three different knowledge bases, the resulting values are depicted in Table 2. The performance values in Table 2 show that for each of the knowledge bases WeeVis identifies solutions fast enough to provide instantaneous feedback from the user interface. If no solution could be found due to inconsistencies between the requirements and the knowledge base, Table 3 shows the time needed to identify at least one alternative solution on the basis of one preferred diagnosis, Table 4 shows the time consumption of calculating all possible solutions. WeeVis is able to calculate either one, two, three or all diagnoses and the corresponding alternative solutions. By taking the response time boundaries for user interfaces into account, the experiment shows that for *small* and *medium* knowledge bases it's possible to calculate all minimal diagnoses within acceptable response times (see Table 3). When it comes to *large* knowledge bases the presented alternative solutions can be reduced to increase the performance of the user interface instead (see Table 4).

	time for identifying solutions
Small	$< 1ms$
Medium	$< 1ms$
Large	$< 2ms$

Table 2. This table shows the time needed to come up with solutions for three knowledge bases. Even for the largest knowledge base the overall time is far below the limit of *an instantaneously reaction*(100ms).

	diagnosis calculation	**repair identification**	**overall time**
Small	$< 1ms$	$< 1ms$	$< 1ms$
Medium	$93ms$	$16ms$	$109ms$
Large	$499ms$	$19ms$	$518ms$

Table 3. This table shows the time needed to come up with at least *one* alternative solution for each of the three knowledge bases. Even for the largest knowledge base the overall time is below the limit of *interrupt a users thought* (1,000ms).

	diagnosis calculation	**repair identification**	**overall time**
Small	$97ms$	$16ms$	$113ms$
Medium	$969ms$	$20ms$	$989ms$
Large	$3,028ms$	$60ms$	$3,088ms$

Table 4. This table shows the time needed to come up with *all* possible alternative solutions for each of the three knowledge bases. For the largest knowledge base the overall time for repair calculation takes about three seconds which is above the recommended time boundaries for *interrupting the user's flow of thought.*

4 Conclusion

In this paper we presented WeeVis which is an open constraint-based recommendation environment. By exploiting the advantages of Mediawiki, WeeVis provides an intuitive basis for the development and maintenance of constraint-based recommender applications. The results of our experiment show that due to the integrated direct diagnosis algorithms the WeeVis user interface provides good the response times for common interactive settings.

References

1. A. Felfernig and R. Burke. Constraint-based Recommender Systems: Technologies and Research Issues. In *10th International Conference on Electronic Commerce*, page 3. ACM, 2008.
2. A. Felfernig, G. Friedrich, D. Jannach, and M. Zanker. An Integrated Environment for the Development of Knowledge-based Recommender Applications. *International Journal of Electronic Commerce*, 11(2):11–34, 2006.
3. G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.
4. J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl. GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 40(3):77–87, 1997.
5. M. Pazzani and D. Billsus. Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Machine learning*, 27(3):313–331, 1997.
6. R. Reiter. A Theory of Diagnosis From First Principles. *Artificial intelligence*, 32(1):57–95, 1987.
7. U. Junker. QUICKXPLAIN: Preferred Explanations and Relaxations for Over-constrained Problems. In *Proceedings of the 19th National Conference on Artifical Intelligence (AAAI'04)*, pages 167–172, 2004.
8. A. Felfernig, M. Schubert, and C. Zehentner. An Efficient Diagnosis Algorithm for Inconsistent Constraint Sets. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 26:53–62, 2012.
9. R. Bakker, F. Dikker, F. Tempelman, and P. Wogmim. Diagnosing and Solving Over-determined Constraint Satisfaction Problems. In *13th International Joint Conference on Artificial Intelligence*, pages 276–281, Chambery, France, 1993.
10. G. Friedrich. Interactive Debugging of Knowledge Bases. In *International Workshop on Principles of Diagnosis (DX'14)*, pages 1–4, Graz, Austria, 2014.
11. J. Baumeister, J. Reutelshoefer, and F. Puppe. KnowWE: a Semantic Wiki for Knowledge Engineering. *Applied Intelligence*, 35(3):323–344, 2011.
12. A. Mackworth. Consistency in Networks of Relations. *Artificial Intelligence*, 8(1):99–118, 1977.
13. A. Felfernig, G. Friedrich, D. Jannach, and M. Stumptner. Consistency-based Diagnosis of Configuration Knowledge Bases. *Artificial Intelligence*, 152(2):213–234, 2004.
14. J. Nielsen. *Usability engineering.* Elsevier, 1994.

Providing Personalized Cultural Heritage Information for the Smart Region - A Proposed methodology

Antonino Lo Bue[1], Alan J. Wecker[2,3], Tsvi Kuflik[3], Alberto Machì[1], Oliviero Stock[4]

[1]CNR-ICAR Palermo, Italy
{lobue, machi}@pa.icar.cnr.it
[2]University of Haifa, Israel
[3]University of Trento, Italy
ajwecker@gmail.com, tsvikak@is.haifa.ac.il
[4]FBK-irst, Trento, Italy
stock@fbk.eu

Abstract. In this paper we present a methodology to provide visitors, in smart regions, additional cultural heritage attractions based on prior museum visits using user models and Linked Open Data. Visitor preferences and behavior are tracked via a museum mobile guide and used to create a visitor model. Semantic models and Linked Open Data support the representation of regional assets as Cultural Objects. The visitor model preferences are exploited using a graph similarity approach in order to identify personalized opportunities for visitors by filtering relevant Cultural Objects.

Keywords: Personalization, User Models, Linked Open Data, Smart Regions

1 Introduction

In this short paper we show a blueprint how semantic models and Linked Open Data (LOD) support the representation of regional assets in order to identify categories of opportunities for visitors based on different personal characteristics determined by previous visits. Having a broad infobase from which to cull possibilities is an arduous task that can benefit from automation. Due to the overwhelming number of possibilities, it is important to personalize the Cultural Heritage (CH) experience. When considering what is requires from a smart, personalized system, it becomes clear that the reasoning process of the system has to focus on identifying opportunities for intervention. When and how to intervene and what information to deliver/service to offer. Having a user model, a context model, and a model of the cultural objects are essential for successful support. These can lead to the interaction of museums and places of cultural heritage to create mega-tourist experience (similar to Verbke and Rekom [6] concept of the "museumpark") which can have a positive market effect for the region.

We describe our methodology: First we use exhibits in a museum (we use Castle Buonconsiglio in the Trentino Region as examples throughout this paper) and tag them using semantic concepts. Then a mobile museum guide is used to track visitors.

Based on this data a user model is developed consisting of characteristics and preferences. We then use a dataset of Cultural Objects using an ontological representation of the domain to cull opportunities. Visitor Preferences are used to filter which Cultural Objects are relevant, and Characteristics are used to determine whether an event or cultural heritage place is desired. Context is used to filter for proximate locations weather conditions, opening times, etc. Again characteristics are used to determine how best to present this information to the visitor.

2 Background

In this section we review two technology areas: User Modeling and personalization in CH, and Linked Open Data and Semantic relatedness.

According to Ardisonno et al[2], for more than 20 years, cultural heritage has been a favored domain for personalization research and as soon as mobile technology appeared, it was adopted for delivering context-aware cultural heritage information both indoors and outdoors. For personalization, a system needs to have a model of its user. A number of approaches are possible: Overlay, Feature-based, Content based, and Collaborative filtering. In this proposed methodology we use an implicit content based approach, where user interests are represented as sets of words occurring in the textual descriptions of items relevant for the user. Visitors have been observed to behave in certain stereotypical movement patterns [11]; patterns such as Butterfly, Grasshopper Ant, and Fish[10]. The use of personality types to tailor software is not new. We use the SLOAN Big 5 characterization as it is standard and much research has been done using it [5]. We focus on two traits we believe are connected to the museum experience: Inquisitiveness, which is a measure of curiosity and Orderliness, which measures thoroughness and the need for structure. Introversion and Extroversion could also play a part in group visits, but is not examined in this research. In addition we posit a connection between movement types and "identity" types proposed by John Falk [4]. Preliminary ideas for the connection of movement patterns to personality types have been proposed [1].

Public agencies collect organize and manage a vast amount of data. Local and European projects aims to deliver data as freely available, reusable and distributed without any restriction, the so call Open Data. As part of these initiatives, tourism and cultural heritage datasets have been published as Open Data. Semantic Web technologies and in particular the Linked (Open) Data paradigm, introduced by Sir Tim Berners-Lee in 2006 [3], are opening new ways for data integration and reuse, creating a method to make data interoperable at a semantic level. Ontologies formally represent knowledge as a set of concepts and their relationships within a domain. RDF[1] and OWL[2] standards enable the formal representation of ontologies as set of triples (subject, predicate, object). Ontologies are used to express vocabularies of Linked Data

1 http://www.w3.org/RDF/
2 http://www.w3.org/2001/sw/wiki/OWL

triples. On top of RDF and OWL, SPARQL Query Language[3] is used to query and retrieve information stored as triples thus allowing and facilitating access to the so called Web of Data. DBpedia[4], can be seen as the ontological version of Wikipedia, its the core of the Linked Open Data cloud.

In the Natural Language Processing area, semantic relatedness between terms or concepts can be computed using two main approaches: (1) defining a topological graph similarity using ontologies and computing the minimal graph distances between terms, (2) using statistical methods and word co-occurrence in a corpus and calculating the correlation between words. "WikiRelate!" [8], measures correlation among terms using a graph based distance measure on the Wikipedia categories. The system uses the inverse path length measure as a distance metric for terms correlation. Leal et al [9] present an approach for computing the semantic relatedness of terms using the knowledge base of DBpedia, based on an algorithm for finding and weighting a collection of paths connecting concept nodes. The implemented algorithm defines the concept of proximity rather than the inverse path length distance as a measure of relatedness among nodes. Our methodology is based on the inverse path length measure but we apply this to a graph of ontology terms extracted from DBpedia and used as annotation for Open Data resources. Moreover, we also take into account the concept introduced by Moore et al. [7], that evaluates paths calculating the number of outgoing links of each node, in order to improve the precision of the algorithm.

3 System

The mobile guide, at each position of interest (POI), presents a list of relevant media assets. The mobile guide system logs: the POI, which assets are chosen how long they viewed the asset, and in general how long did they stay at the point of interest. We collect two types of information, the first in order to determine general personal characteristics and the second in order to determine specific topic interests. In general we use movement styles, to predict user characteristics (such as personality). We use time viewing presentations in order to determine user topic preferences.

In order to characterize the user we make use of his general movement activities. We use the following statistics: 1) NumberOfPOIsVisted (NPV) – number of positions where a person stayed more than 9 seconds as detected and logged by the mobile guide's positioning system. Nine seconds is a number we have used for previous analysis and has provided good results. 2) POIsWherePresentationsSeen (PPS) – the number of positions where the visitor viewed at least one media asset connected to that position as computed from the logs of the mobile guide. 3)NumberOfPresentationSeen (NPS) – the total number of media assets the visitor viewed as computed from the logs of the mobile guide.

[3] http://www.w3.org/TR/sparql11-query/
[4] http://dbpedia.org

Table 1. Connecting the user behavior to personality and Falk types

Behavior	*Personality*	*Falk*	*Formula*
Fish	Non curious – Unorderly	Recharger	$((PPS/NPV <= T_1)$ &$(NPS/PPS < T_3))$
Ant	Inquisitive – Orderly	Explorer	$(PPS/NPV > T_1)$ & $(NPS/PPS > T_2)$
Grasshopper	Non curious – Orderly	Professional	$(PPS/NPV > T_1)$ & $(NPS/PPS < T_2)$
Butterfly	Inquisitive – Unorderly	Exp. Seeker	$(PPS/NPV < T_1)$ & $(NPS/PPS > T_3)$

3.1 What can we find and match up

The system uses annotated internal and external information about cultural places and events. Internal information is taken from catalogues or websites and is used by the mobile guide app to describe user preferences by storing the relevant topics related to exhibits the user has visited and liked. External information is imported from available Open Data about museums and cultural events and enriched in the domain ontology, using knowledge from the Linked Open Data cloud (DBpedia dataset). Data is stored using a domain ontology for tourism called *eTourism*[5]. The ontology covers methodological and practical aspect of services (hotels, B&B, etc.), cultural objects (museum, cultural places, etc.) and events. It is used as a vocabulary model to map external Open Data into RDF triples validated by the ontology concepts. For the present work we have developed a specific module of the *eTourism* ontology named *Cultural Objects Ontology* (*coo*) that covers (1) properties (such as topic, keywords, geographical information) of museums or events, exploits the semantic identity with LOD/DBpedia concepts (using *owl:sameAs* predicates) and implements (2) user profile types and topics of interests selections.

For each museum source, we extract - as a first step, keywords from exhibits of the Castle Buonconsiglio museum. We exploit the semantic relatedness implementing the graph similarity approach. We annotate keywords - for each description, and we disambiguate them to DBpedia concepts using DBpedia Spotlight APIs[6]. We filter out all the not relevant concepts and we then obtain a bag of concepts (related to cultural heritage) like the following:

{dbpedia[7]:Trentino, dbpedia:Prehistory, dbpedia:Ancient_Rome, dbpedia:Middle_Ages, dbpedia:Hunter-gatherer, dbpedia:Upper_Paleolithic, dbpedia:Bronze_Age}

In DBpedia, each concept is related to a category using the property *dcterms:subject*, then each category is part of a hierarchy structure with nodes connected via *skos:broader* properties. For example the below two DBpedia concepts have as *dcterms:subject* the DBpedia *topic* categories:

1) Last_glacial_period *(dcterms:subject) ->{Climate_history, Glaciology, Holocene, Ice_ages}*
2) Ancient_Rome *(dcterms:subject) ->{Ancient_history, Ancient_Rome, Civilizations}*

---

[5] Currently under development at ICAR-CNR within the framework of the national project Dicet-InMoto-Orchestra, (http://www.progettoinmoto.it).
[6] http://spotlight.dbpedia.org
[7] Prefix for http://dbpedia.org/resource/

For the second step, we extract from the DBPedia SPARQL endpoint, for each concept, the *topic* categories of the DBpedia taxonomy. As result we obtain a wider bag of DBpedia *topic* categories describing each museum exhibit. Using the hierarchical structure of categories is thus possible to discover similarities among concepts that have ancestor categories in common.
As external sources, we take the Open Data set delivered by the Italian Cultural Heritage Minister[8] (MIBAC) and we map these objects using the *coo* ontology; then, for each object, we exploit the same process applied for the internal resources, in order to annotate and extract the corresponding bag of topics. As a result, we obtain a list of information for each MIBAC *Cultural Object* (cultural place or event), as in the following example:
foaf:name = "Memorie della Grande Guerra",
coo:mainCategory = http://dbpedia.org/resource/Category:History
Bag of Concepts (*dcterms:description*) ->
{*1918_disestablishments, Aftermath_of_World_War_I, Austria-Hungary, Austria_articles needing attention, States_and_territories_established_in_1867, Anoxic_waters, Back-arc_basins, Contemporary Italian_history, History_of_Austria-Hungary, History_of_modern_Serbia, Wars_involving_Italy, World_War_I* }

In order to select suitable *Cultural Objects* candidates for the user, we define a metric to measure the semantic distance between the user profile tags and the available cultural objects tags. As a first step, we measure the *shortest path distance* between each of the *m topic* categories in the bag of topics of the user profile and the *coo:mainCategory* topic of the suitable candidates (see table 2), and we reduce candidates cardinality by applying an upper threshold to the distance.

Table 2. Example path between two DBpedia categories

Distance	*Steps*	*Distance*	*Steps*
0	dboc:[9]Ancient_history	4	dboc:Art_history
1	dboc:Periods_and_stages_in_archaeology	5	dboc:Visual_arts
2	dboc:Archaeology	6	dboc:Arts
3	dboc:Conservation_and_restoration		

After this step, we refine the result by calculating (via SPARQL queries on the DBpedia endpoint) the *shortest path* between the user bag of topics (*m)* and the suitable candidates bag of topics (*n)* on the remaining subset of cultural objects. Its important to underline that when computing the distance measure between topic categories we also take into account, for each hop of the shortest path, the number of outgoing links of the node: the more outgoing links a node has (to other DBpedia taxonomy nodes) the less it is specific. Broad connected nodes receive low weights while nodes with less outgoing connection will get higher values. We use each pairwise distance as a component of a normalized vector of distances, we evaluate, for each museum or

8 http://dbunico20.beniculturali.it/DBUnicoManagerWeb/#home
9 Prefix for http://dbpedia.org/resource/Category:

event an average normalized distance for each *m* user category and we sum all these distances to define the relatedness of each cultural object. Again an empirical threshold on distance is applied to retain a limited number of candidates.

3.2 Use of characteristics

Using behavior types we can tailor the amount and presentation of information. For example for ants and butterflies we can give ten items. For grasshoppers and fish we may only give two items. Ants and grasshoppers may be given places while butterflies and fish may be given events. Additional personalization may be possible.

4 Discussion and Conclusion

The results we get for the four sample users are shown on the table below.

Table 3. Simulated output of the system with Places and Events suggested per each user behavior. Suggested items are marked with a *.

Type	*Preferences*	*Places, Events*
Ant	Bronze_Age (.5), Feudalism (.2), Middle Ages (.5), Ancient Egyptian funerary practices (.1), Civilizations (.2)	Museo archeologico dell'Alto Adige (Archeology) (.6), Area archeologica Palazzo Lodron (Archeology) (.6), Museo delle palafitte del Lago di Ledro (History) (.4), Museo locale di Aldino (Etnography) (.2*)
Grass-hopper	Romantic_art (.4), 20th-century Italian_painters (.3), Postmodern_art (.3), Fresco_painting (.3),Rural_culture (.1)	Museo Rudolf Stolz (Arts) (.6), Museo di arte moderna e contemporanea di Trento Rovereto (Arts) (.5), Museion - Museo d'arte moderna e contemporanea (Arts) (.6), Museo della Val Venosta (Anthropology) (.2*)
Butterfly	World_War_I (.4), Civilizations (.4), 1st-century Roman emperors (.2), History_of_Europe (.6), Rural_culture (.2)	Doni Preziosi, Immagini e Oggetti dalle Collezioni Museali (Exhibition/History) (.5), Storie da Trento all'Europa. Mostra documentaria (Exhibition/History) (.5)
Fish	Romantic_art (.4), 20th-century Italian painters (.3), Bronze_Age (.5), Fresco painting (.3), Rural_culture (.1)	Rinascimenti Eccentrici al Castello del Buonconsiglio (Exhibition/Arts) (.7), Apertura Spazio archeologico Sotterraneo del Sas (Opening/Archeology) (.4)

Our current metric of semantic relatedness doesn't take into account whether the user profile bag of topics is representative of a sufficiently broad range of museums categories to cover their cultural preferences. To balance this, when all/most of the user preferences are of the same topic area (e.g. Prehistory), one or more among suggested items could be chosen from a minor topic category, to elicit variation in user interests. Our current research involves, the implementation of the methodology to the Old City and the Tower of David Museum in Jerusalem, and the evaluation of the user model and the semantic suggestions results.

References

1. Antoniou, A., Lepouras, G., Lykourentzou, I. et al.: Connecting Physical Space, Human Personalities and Social Networks.
2. Ardissono, L., Kuflik, T., Petrelli, D.: Personalization in Cultural Heritage: The Road Travelled and the One Ahead. User Modeling and User-Adapted Interaction, 22 (2011) 1-27
3. Berners-Lee, T. (2006). Design issues: Linked data.
4. Falk, J. H.: Identity and the museum visitor experience. Left Coast Press Walnut Creek, CA (2009)
5. Higgins, D. M., Peterson, J. B., Pihl, R. O. et al.: Prefrontal Cognitive Ability, Intelligence, Big Five Personality, and the Prediction of Advanced Academic and Workplace Performance. J. Pers. Soc. Psychol., 93 (2007) 298
6. Jansen-Verbeke, M., & Van Rekom, J.: Scanning Museum Visitors: Urban Tourism Marketing. Ann. Tourism Res., 23 (1996) 364-375
7. Moore, J. L., Steinke, F., & Tresp, V.: A novel metric for information retrieval in semantic networks. ESWC 2011 Workshops 65-79
8. Strube, M., & Ponzetto, S. P.: WikiRelate! Computing semantic relatedness using Wikipedia. AAAI Vol. 6, pp. 1419-1424 (2006).
9. Leal, J. P., Rodrigues, V., & Queirós, R.: Computing semantic relatedness using DBpedia. SLATE 2012, page 3519, 2012.
10. Veron, E., & Levasseur, M.: Ethnographie de l'exposition. Centre Georges Pompidou (1983)
11. Zancanaro, M., Kuflik, T., Boger, Z. et al.: Analyzing Museum Visitors' Behavior Patterns. User Modeling 2007, (2007) 238-246

Personalized Extended Government for Local Public Administrations

Claudio Biancalana[1,2], Alessandro Micarelli[1], and Giuseppe Sansonetti[1]

[1] Roma Tre University, Department of Engineering
Via della Vasca Navale 79, 00146 Rome, Italy
[2] Lazio Innovazione Tecnologica, LAit S.p.A.
Via Adelaide Bono Cairoli 68, 00144 Rome, Italy
claudio.biancalana@laitspa.it,{micarel,gsansone}@dia.uniroma3.it

Abstract. This paper discusses the enterprise organization environment and reports our experience and lessons learned in developing an extension of the traditional virtual enterprise model, we named *personalized extended government (PEG)* model. The aim of such model is to simplify and enhance the effectiveness of e-Government services, by realizing Administration to Administration (A2A) and Administration to Citizen (A2C) processes in a personalized perspective. The features of the proposed model make it suitable for use in local public administrations. As a proof of this, it has been successfully deployed to realize the Italian Open Government Data Portal of Regione Lazio, which allows every citizen to be informed about the employment of public resources on regional territory.

Keywords: personalization, e-Government, virtual enterprise model

1 Introduction

In the late 1990s, the concept of virtual enterprise (VE) model has been introduced [2], where every business organization unit is connected to each other through a data transmission network, in order to explore market opportunities and cooperate, on a temporary basis, to better respond to business opportunities. Hence, a VE can be seen as a heterogeneous network for both enterprises and individuals with integrated cooperation, exploiting ICT technologies and protocols for a specific business process. Over the years, a second model has been developed, substantially similar to the VE model, but based on more stable and long-term agreements: the extended enterprise (EE) model [7]. Recently, an organization model similar to EE has been implemented at a government level and can be recognized in initiatives such as the Italian Open Government Data Portal of Regione Lazio [3]. Such a portal provides a web interface that centralizes access to all open datasets for anyone, in particular for data journalists,

[3] https://dati.lazio.it

public administrations, scientists, and business people. The project has been designed to realize Administration to Administration (A2A) and Administration to Citizen (A2C) processes, and relies on stable and deeply defined agreements. Unlike the EE model, however, the focus is not on business opportunities, but on making e-Government services simpler and more effective. In addition, the need to manage and discover unstructured information has resulted in the gradual awareness of the need to adopt knowledge management systems (KMSs) based on semantic and user-modeling functions. This high degree of similarity allows us to introduce a new definition to refer to the concepts described so far, namely, *personalized extended government (PEG)*, as a personalized and context-oriented extension of EE type. This paper: (1) defines PEG; and (2) provides notes on the design of a KMS that supports PEG.

2 Personalized Extended Government

Hereafter, we refer to personalized extended government (PEG) as an integrated unit of organizations, agreements, protocols and ICT resources able to support public administrations to deploy a context-oriented model to build Administration to Administration (A2A) and Administration to Citizen (A2C) scenarios, in order to simplify and to improve the effectiveness of e-Government services. By analogy with the EE model, we can define the following major PEG model features:

- e-Government service-driven cooperation: A2A and A2C processes are always aimed at providing electronic government services to citizens and businesses, with the goal to simplify and make them more efficient and effective;
- Complementarity: administrations exchange with each other only correct and complete data;
- Process integration and resource sharing: more specifically, data, information and knowledge;
- Interdependence: process integration and resource sharing are carried out according to well defined cooperation agreements.

In order to deploy the PEG organization model, it is necessary to define the following aspects:

- Guidelines for every single participant IT assets integration. This problem is due to different technologies used by every administration and the need to preserve both investments and administration autonomy. For these reasons, it is necessary to define a technological infrastructure that guarantees interoperability regardless of the organization structures and single participant legacy systems;
- Maturity model, which is a structured collection of elements that describe certain aspects of maturity in an organization, for example, to provide a way to define what improvement means for an organization;
- Common governance model through the administrations of all participants and citizens.

Regione Lazio's experience raises an issue: the realization of A2A and A2C isolated processes leads to fragmented knowledge and to a loss of fundamental information used to integrate management relationship between administrations and citizens or enterprises. For this reason, LAit S.p.A. and Regione Lazio have planned a knowledge management system (KMS) design with basic concepts inspired to both EE model and PEG model, in order to develop research ideas in the EE field. The main principles of KMSs should be the following:

- Affordable setup: no more heavy bulked social networks held by central public administrations. As a normal web user can now start a forum or a blog using third party (often free) software, he should also be able to use a web host or a hosting service;
- Accessibility through (semantic?) search engines: in our vision, this is surely something related to the open nature of KMSs, but it would gain some commitment from search engines, which will be able to improve quality of searches through proper indexing of published semantic annotations;
- Scalable open architecture: a given service may explicitly be built upon a KMS, committing to its ontologies and content organization. Vice versa, in an even more open view, independent services may be linked by a given KMS. This would allow users to tag the content of these services according to the OASIS reference ontologies, thus easily putting traditional (non semantic-driven) services immediately into practice. The same process would be applied to standard web pages. People could write web pages directly connected to a KMS making explicit reference to its vocabulary, as embedded RDFa, or they could semantically bookmark an external web page (or annotate part of its content) against that same vocabulary.

2.1 Knowledge Indexing

Before addressing the problem of knowledge retrieval, it is essential to analyze how the system indexes the available information, that is, which representation has been chosen to guarantee an efficient and effective retrieval phase. The requirements are twofold: it is essential that on the one hand knowledge is quickly retrieved by users, on the other hand this knowledge accurately satisfies users information needs in terms of high precision. An indexing system for business companies must also be able to deal with different kinds of information representations, from unstructured documents based on natural language to ontology-based knowledge and relational databases. Moreover, it should provide a comprehensive and homogenous human-computer interface for knowledge retrieval. In order to provide the aforementioned prerequisites, it is necessary to consider different types of information and the degrees of information "richness". Information based on ontological standards, for instance, expresses relationships between typically non-structured information, such as natural language text and metadata. Such metadata usually describe features or classes related to given pieces of information. A typical example is the association between a document and one particular category in a predefined taxonomy. As for information stored in databases, we have an underlying relational model that clearly states the semantic meaning of each piece of information unit, such as price, address, and location, and therefore enables the interpretation/recovery process. In order to define a unique representation that deals with all the different types of available information (i.e., natural language, ontology-based, and databases) we must define a subset of shared features that is possible to generalize, and automatic or semi-automatic methods and techniques for translating information from one of these representations to the internal one. This sort of intermediate representation consists of traditional non-structured information with associated meta-information related to concepts of a taxonomy of the business domain for the given public administration unit (PAU). Briefly, each information unit is classified in a subset of categories from a simplified ontology. Such meta-information can be exploited both in the retrieval phase, to reduce possible ambiguities in the processed information, and to re-organize the knowledge in more efficient ways for further user search activities, such as online hierarchical clustering. Information based on ontological standards does not pose relevant issues. In this case, the source is based on a rich language while our internal representation simplifies some features, such as the kinds of relations between concepts. In our representation, we have relations $\langle u, C \rangle$ where u is the information unit and C is a set of categories in the given taxonomy related to the concept u. We only have IS-A relationships, so it is not hard to extract them from the initial ontology. The selection of the most important concepts from the initial ontology is the only task that knowledge experts have to perform before populating the internal taxonomy. Considering the current amount of unstructured information available within companies, the problem of making such information accessible to users is likely to be the most important issue to solve in our knowledge system. Specifically, having chosen a particular representation for PAU domains, it is necessary to find a technique that allows us to autonomously process the unstructured information and populate the internal knowledge base. In input we have information objects, typically text documents, reports, hypertext pages, etc. These objects are processed for named entity extraction, text segmentation, and text categorization. Given an information object, we initially locate and classify atomic elements in text into predefined categories, such as names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. To this aim, we used named entity recognition (NER) systems based on linguistic grammar-based techniques, statistical models and dictionaries (or gazetteers). NER is a well-known research field, subtask of information extraction, which does not focus on semantic interpretation of languages but on more practical and easier goals, so obtaining excellent results in terms of precision of results. The proper nouns of companies, persons, etc. in output of the NER module are used to increase the weight of these entities during the indexing/retrieval steps. A further step we follow to process the input information is text segmentation. In short, a given document is divided into sequences of words or other similar meaningful units that are separately stored in the knowledge base. This step is useful whenever we have long documents, such as reports or e-books, which cover several different topics (i.e., categories in the internal taxonomy). In order to increase the retrieval precision, it is better to split them into meaningful coherent regions. Our segmentation algorithm is based on the Choi's work [3]. It performs three steps: (1) tokenization, (2) lexical score determination, and (3) boundary identification. Basically, after breaking the document up into a sequence of tokens, we use a similarity measure to analyze the semantic coherence among contiguous text regions. Finally, we determine the boundaries whenever we have relevant variations of the semantic coherence measure. The last process we perform on unstructured information is text categorization. After recognizing semantically coherent information units and - for each unit - its relevant entities, we assign a subset of taxonomic categories to it through text categorization techniques [8]. As for the training phase, we use a subset of documents already categorized by the knowledge expert, performing an *ad-hoc* feature selection that also exploits the aforementioned NER module to assign more weight to terms that correspond to relevant semantic classes, such as proper names and locations. The categorization output is a tuple of couples $\langle c_i, \alpha_i \rangle$, where for each taxonomic category c_i we have a value α_i between $[0, 1]$ that represents the degree of relatedness of the input document to the class. This information is stored along with the document in the knowledge base and it is used during the retrieval and personalization phases, as described in the following section.

2.2 Semantic Querying and Personalization

The most popular paradigm for querying a textual database is to submit short queries. Users express their information needs through a small set of keywords that must be present in the retrieved documents. The retrieval system returns an ordered list of references, based on matching algorithms that assign a relevance weight to each indexed document. In our knowledge base, along with each document, we have a list of assigned categories referenced in the internal taxonomy. In order to exploit this information, the query should include one or more categories that users are interested in. We named these enhanced user needs *semantic queries* [1]. One of the most important problems that occurs while querying a corpus of textual documents is the choice of the right keywords for retrieval. Synonymy (i.e., two words that express the same meaning) and polysemy (i.e., different meanings expressed by one word) of natural language may decrease the recall and precision of the retrieval process [4]. For that reason, we have included a user modeling component to represent the users needs. This component is involved during the querying in order to help disambiguate the meaning of the query terms. Some user modeling strategies have been proposed in the e-Government services field (see, for instance, [6]). The proposed user modeling is based on a concept network paradigm [5] instantiated on the taxonomy of the PAU domain. Concept networks are usually employed as a form of knowledge representation. They consist of graphs of concepts and edges that represent semantic relations between these concepts. We use concept networks to weight which concepts users are more interested in, that is, concepts related to the user needs. In our first prototype, the relations between concepts are not considered.

3 Conclusion

In this paper we have introduced a natural extension of the virtual enterprise model, we called *personalized extended government (PEG)* model, whose aim is to encourage and facilitate the exchange of public domain and community information in a personalized perspective, respecting the public administrations and citizens information needs. The implementation of the proposed model at a government level has represented a strategic roadmap for Regione Lazio ICT Government.

Obviously, the further development of the PEG model first of all involves planning and performing an in-depth experimentation, in order to assess the actual satisfaction of stakeholders. Furthermore, several future developments of all aspects of this work are possible. More specifically, we would like to enhance the effectiveness of the proposed model by exploiting information extracted from social media. In literature, indeed, some works (e.g., see [9]) show how government services can be improved based on user-generated contents found in publicly available social media.

References

1. Biancalana, C., Gasparetti, F., Micarelli, A., Sansonetti, G.: Social semantic query expansion. ACM Trans. Intell. Syst. Technol. 4(4), 60:1–60:43 (Oct 2013)
2. Camarinha-Matos, L.M., Afsarmanesh, H.: The virtual enterprise concept. In: Proc. of the IFIP TC5 WG5.3 / PRODNET Working Conference on Infrastructures for Virtual Enterprises. pp. 3–14. Kluwer, B.V., Deventer, The Netherlands (1999)
3. Choi, F.Y.Y.: Advances in domain independent linear text segmentation. In: Proc. of the 1st North American Chapter of the Association for Computational Linguistics Conference. pp. 26–33. Stroudsburg, PA, USA (2000)
4. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. CACM 30(11), 964–971 (1987)
5. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, pp. 54–89. Springer-Verlag (2007)
6. Iaquinta, L., Torsello, M.A., Comerio, M., Fanelli, A.M., Semeraro, G.: User segmentation in e-government services. In: Late-Breaking Results, Project Papers and Workshop Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization, Rome, Italy (June 2013)
7. Oliveira, M., Crowcroft, J., Slater, M.: An innovative design approach to build virtual environment systems. In: Proc. of the Workshop on Virtual Environments. pp. 143–151. ACM, New York, NY, USA (2003)
8. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. 34(1), 1–47 (2002)
9. Wan, S., Paris, C.: Improving government services with social media feedback. In: Proceedings of the 19th International Conference on Intelligent User Interfaces. pp. 27–36. ACM, New York, NY, USA (2014)