

Personalized Environmental Service Orchestration for Quality Life Improvement

Leo Wanner¹, Stefanos Vrochidis², Marco Rospocher³, Jürgen Moßgraber⁴, Harald Bosch⁵, Ari Karppinen⁶, Maria Myllynen⁷, Sara Tonelli³, Nadjat Bouayad-Agha¹, Gerard Casamayor¹, Thomas Ertl⁵, Désirée Hilbring⁴, Lasse Johansson⁶, Kostas Karatzas⁸, Ioannis Kompatsiaris², Tarja Koskentalo⁷, Simon Mille¹, Anastasia Mourtzidou², Emanuele Pianta³, Luciano Serafini³, Virpi Tarvainen⁶

¹ Dept. of Information and Communication Technologies, Pompeu Fabra University

² Centre for Research and Technology Hellas, Informatics and Telematics Institute

³ Fondazione Bruno Kessler

⁴ Fraunhofer Institute of Optronics, System Technologies and Image Exploitation

⁵ Institute for Visualization and Interactive Systems, University of Stuttgart

⁶ Finish Meteorological Institute

⁷ Helsinki Regional Environmental Services Authority

⁸ Informatics Applications and Systems Group, Aristotle University of Thessaloniki

{leo.wanner@upf.edu, stefanos@iti.gr, rospocher@fbk.eu,
juergen.mossgraber@iosb.fraunhofer.de, harald.bosch@vis.uni-stuttgart.de,
ari.karppinen@fmi.fi, maria.myllynen@hsy.fi, satonelli@fbk.eu, nadjat.bouayad@upf.edu,
kkara@eng.auth.gr, gerard.casamayor@upf.edu, thomas.ertl@vis.uni-stuttgart.de,
desiree.hilbring@iosb.fraunhofer.de, lasse.johansson@fmi.fi, ikom@iti.gr,
tarja.koskentalo@hsy.fi, simon.mille@upf.edu, mourtzid@iti.gr, pianta@fbk.eu,
serafini@fbk.eu, virpi.tarvainen@fmi.fi}

Abstract. Environmental and meteorological conditions are of utmost importance for the population, as they are strongly related to the quality of life. Citizens are increasingly aware of this importance. This awareness results in an increasing demand for environmental information tailored to their specific needs and background. We present an environmental information platform that supports submission of user queries related to environmental conditions and orchestrates results from complementary services to generate personalized suggestions. From the technical viewpoint, the system discovers and processes reliable data in the web in order to convert them into knowledge. At run time, this information is transferred into an ontology-structured knowledge base, from which then information relevant to the specific user is deduced and communicated in the language of their preference. The platform is demonstrated with real world use cases in the south area of Finland showing the impact it can have on the quality of everyday life.

Keywords: environmental information service, environmental node discovery, knowledge, personalization, infrastructure, services.

1 Introduction

Environmental and meteorological conditions are considered of utmost importance for the population, as they strongly influence the quality of human life. Citizens are increasingly aware of the important role that environmental data and measurements play in health issues (e.g. allergies), as well as in a variety of important daily activities (e.g. agriculture). One of the consequences of this awareness is the demand for high quality environmental information that is tailored to one's specific context and background (e.g. health conditions, travel preferences, etc.), i.e., which is personalized. Personalized environmental information may need to cover a variety of aspects (such as meteorology, air quality and pollen) and take into account a number of specific personal attributes (health, age, etc.) of the user, as well as the intended use of the information. So far, only a few approaches have been proposed with a view of how this information can be facilitated in technical terms. All of these approaches focus on one environmental aspect and only very few of them address the problem of information personalization [1], [2], [3]. On the contrary, we aim at addressing the aforementioned task in its full complexity.

In our work, we take advantage of the fact that nowadays, the World Wide Web already hosts a great range of services (i.e. websites, which provide environmental information) that offer data on each of the above aspects, such that, in principle, the required basic data are available. The challenge is threefold: first, to discover and orchestrate these services; second, to process the obtained data in accordance with the needs of the user; and, third, to communicate the gained information in the user's preferred mode. To address this problem, we need to involve a considerable number of rather heterogeneous applications and thus create an infrastructure that is flexible and stable enough to support a potentially distributed architecture. This infrastructure is realized by the "PESCaDO platform", which is described in the upcoming sections.

In Section 2, we outline the process of the discovery and extraction of environmental information from services (also referred to as *nodes*) in the Web, which is considered as the prerequisite step for enabling the retrieval capabilities of the system. In Section 3, we describe briefly the processing of the data obtained from the environmental nodes until their delivery to the user. In Section 4, we present the infrastructure designed to accommodate for both the discovery itself and the posterior tasks. Section 5 illustrates the functionality of our system, while Section 6 concludes with a brief summary of our exposé.

2 A Sample User Scenario

In order to set up the context of the use of the PESCaDO platform, let us imagine a non-professional user, Fiona Fit, who is in her late twenties and lives in the municipality of Espoo, located in the south area of Finland. Fiona is rather active in her leisure time and often goes hiking. But since she is allergic to birch pollen, she needs information on the environmental conditions in the hiking resorts before she decides on her route. This afternoon, Fiona wants to go for a hike in the Nuuksio National Park near Helsinki and needs to know whether the forecasted air quality,

weather and pollen conditions are favorable. As a registered user, with her profile uploaded to the system, she seeks decision support from PESCaDO.

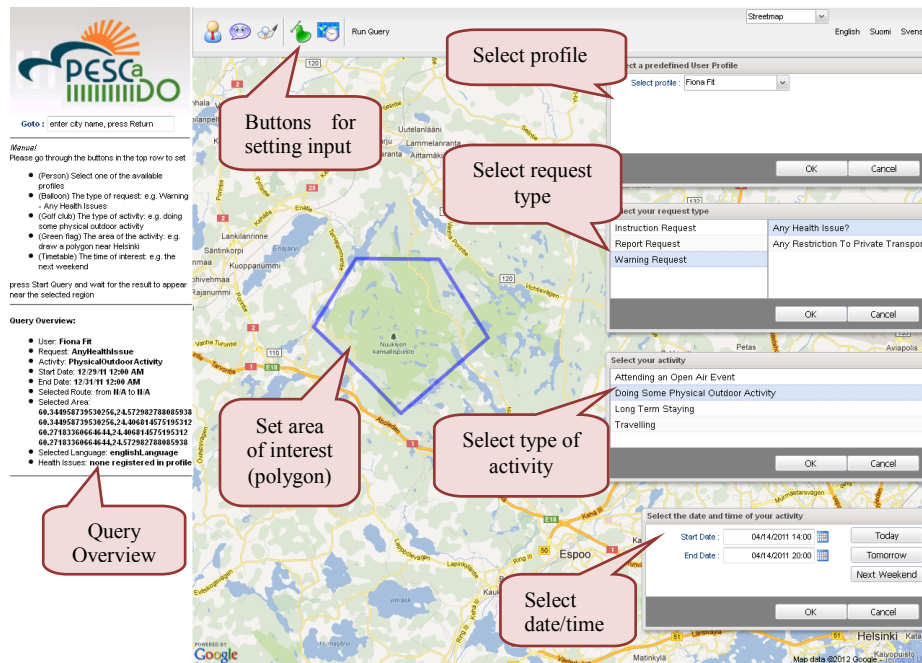


Fig. 1. PESCaDO demonstrator and input for the sample use case

Fiona uses PESCaDO's interface to formulate the aforementioned query.¹ Figure 1 displays the interface and the type of input information a user can submit, i.e., the profile, the request type (whether it is an instruction, report or warning request), the type of activity (e.g. travelling, physical outdoor activity), the start/ end date/ time and the region of interest (depicted as blue polygon on the map).

Based on its knowledge regarding the forecasted air quality, pollen and weather values as well Fiona's health status, the system provides an answer which discourages Fiona from engaging in any sport activities because high concentration of Thoracic Particles is expected during the selected time in the region of Nuuksio Park. The system's answer is displayed in Figure 2.

Although a higher degree of personalization still seems possible, especially in terms of a direct amicable salutation of the user and the like, the offered personalized information without any doubt is already capable of improving the quality of life of their addressees.

Let us discuss in the following sections how the system processes a query and how it comes up with a personalized answer.

¹ Link to the demonstrator is available at: <http://www.pescado-project.eu/>

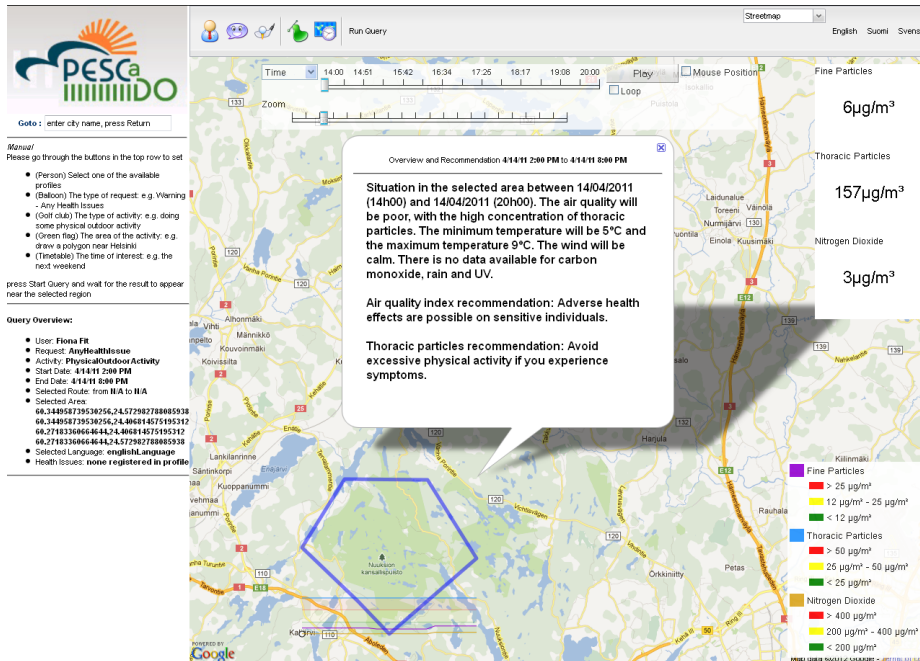


Fig. 2. PESCaDO demonstrator and output for the sample use case

3 Architecture of the PESCaDO Platform

In order to be able to “understand” the query of the user, to gather the required data, to interpret them and then to generate a recommendation or any other type of information useful to the user, PESCaDO needs to tackle a number of tasks:

1. Discovery of the relevant environmental service nodes in the web and extraction of information. As already mentioned above, the web hosts a large amount of environmental (meteorological, air quality, pollen, etc.) distributed services, which include both public webpages that offer environmental information worldwide, as well as dedicated environmental web services with free access; especially the number of meteorological services that cover each major location is impressive. Among all these services those must be discovered that maybe of relevance to PESCaDO and the information offered by them must be extracted. The heterogeneous forms and formats, including text and images, in which this information is encoded make the task of discovery and extraction of information from webpages that provide environmental information a serious challenge. The information extracted from the discovered service nodes is stored in a repository and indexed (together with the references to the nodes). This task is performed off-line, i.e., independent of the queries of PESCaDO’s users, while all the other tasks are performed on-line.

2. Identification of user relevant service nodes. The indexed environmental repository compiled as result of the node discovery task and updated at a predefined time rate serves as the general data source for PESCaDO. That is, when a user poses a query, a process of the identification of environmental service nodes in the compiled repository that are relevant to the query of the user, their profile and their context must be launched. This is not trivial, given that, for instance, a user may be moving, be located in an area which is not directly covered by any node.

3. Orchestration of environmental service nodes: Environmental nodes may provide competing or complementary data on the same aspect for the same or the neighboring location. To ensure the availability of the most reliable and most comprehensive content, the contents proceeding from these nodes must be (i) assessed with respect to its trustworthiness and certainty and selected accordingly (if several nodes offer competing data) or (ii) fused (if several nodes offer complementary information).

4. Converting the data into content. In order to guarantee a motivated orchestration of heterogeneous environmental service nodes and offer user-tailored decision support services and environmental information production, we need to convert the data into structured unified content, which allows for application of intelligent reasoning algorithms. To this end, the extracted and fused environmental data are integrated into an environmental knowledge base (KB). Our KB, which is codified in the standard semantic web ontology language OWL [10], covers environmental content such as meteorological conditions and phenomena, air quality, and pollen, as well as other relevant environment-related content essential for the targeted user-tailored service: travel information, human diseases, geographical data, user profile, etc. In addition, the KB is also capable of formally representing the description of the user's inquiry. The current version of the KB contains around 202 classes, 143 attributes and properties, 463 individuals². Its Description Logic expressivity is *ALCHOIQ(D)*. The KB has been obtained by (i) including customized version of currently available ontologies (e.g., parts of the SWEET ontology), (ii) automatically extracting key concepts from domain relevant text sources, and (iii) manually adding additional properties and attributes.

5. Assessment of the content with respect to the needs of the user: Once the data from the nodes have been incorporated into the KB, they need to be evaluated and reasoned about in order to infer how they affect the addressee, given his/her personal health and life circumstances and the purpose of his/her request. For instance, a citizen may request information, because he/she wants to decide upon a planned action, be aware of extreme episodes or monitor the environmental conditions in a location.

6. Selection of user-relevant content and its delivery. Not all content in the KB is apt to be communicated to the addressee: some of it would sound trivial or irrelevant. Intelligent content selection strategies take into account the background of the user and the intended use of the information to decide which elements of the content are worth and meaningful to be communicated. To deliver the selected

² These data refer to the "empty" KB, i.e. without considering any environmental data coming from the nodes.

content, techniques are required that present the content in a suitable mode (text, graphic and/or table) and in the preferred language.

7. Interaction with the user. The design and development of intuitive interfaces for the interaction between the system and the user forms the final task. The user must be able to formulate the problem in a simple and intuitive format and receive the generated information in a suitable form.

In order to accommodate for all tasks described above, we opted for a service-based architecture. This architecture is based on a methodology which has been developed in ORCHESTRA [11] for risk management, and which has been extended in SANY [12] to cover the domain of sensor networks and standard-based sensor web enablement. The focus of this methodology is on a platform-neutral specification. In other words, it aims to provide the basic concepts and their interrelationships (conceptual models) as abstract specifications. The design is guided by the methodology developed in the ISO/IEC Reference Model for Open Distributed Processing (RM-ODP), which explicitly foresees an engineering step that maps solution types, such as information models, services and interfaces specified in information and service viewpoints, respectively, to distributed system technologies.

We defined application-specific major tasks and actions as abstract service specifications, which can be implemented as service instances on a specific platform. Web service instances for these services are currently being developed. They can be redefined and substituted in the future as needed. Figure 3 displays a simplified sample workflow with the major application services in action. Two services are not cited in Figure 3 since they are consulted by nearly all other services: the Knowledge Base Access Service and the User Profile Management Service. The figure also does not include the services related to data discovery and information distillation from webpages.

A main dispatcher service (called Answer Service, AS) controls the workflow and the execution of the services. First, the user interacts with the system via the User Interaction Service (UIS). In the case that the user is unsure with respect to the types of information they can ask for, they can inquire this information by requesting it from the Problem Description Service (PDS).

To ensure a full comprehension of the problem or user generated question, we decided to operate with controlled graphical and natural language input formats. Once the user has chosen what kind of question they want to submit to the system, the UIS provides them the corresponding formats. Thereupon, the user can formulate their query, which is subsequently translated by the PDS into a formal ontology-based representation understood by the system. After the problem description is generated, it is passed by the UIS to the AS as a "Request Answer" inquiry. Then, the AS assesses what kinds of data beyond environmental data are required to answer the query of the user and solicits these data from the Auxiliary Services (AuxS). Such services can provide for instance travel route information in the case that the user's query concerns the environmental conditions for a bicycle tour from A to B.

After having acquired the complementary data, the AS can request from the Data Retrieval Service (DRS) the environmental data needed to answer the user query. The DRS solicits these data from the environmental nodes identified by the Data Node Retrieval Service (DNRS), which accesses the data node repository.

As already mentioned, the retrieved nodes may deliver complementary or competing data of varying quality (to keep the presentation simple, we dispense with the illustration of the orchestration of service nodes). The Fusion Service (FS) applies uncertainty metrics to obtain the optimal and maximally complete data set, which is passed by the AS to the Decision Service (DS). The DS converts the data set into knowledge, in that it relates it to the knowledge in our KB, reasons about it, and assesses it from the perspective of its relevance to the user. From this content, the Content Selection Service (CSS) compiles a content plan which contains the knowledge to be communicated to the user as the answer. The Information Production Service (IPS) takes the content plan as input and generates information in the language and mode (text, table, or graphic) of the user preference, which then is passed to the user.

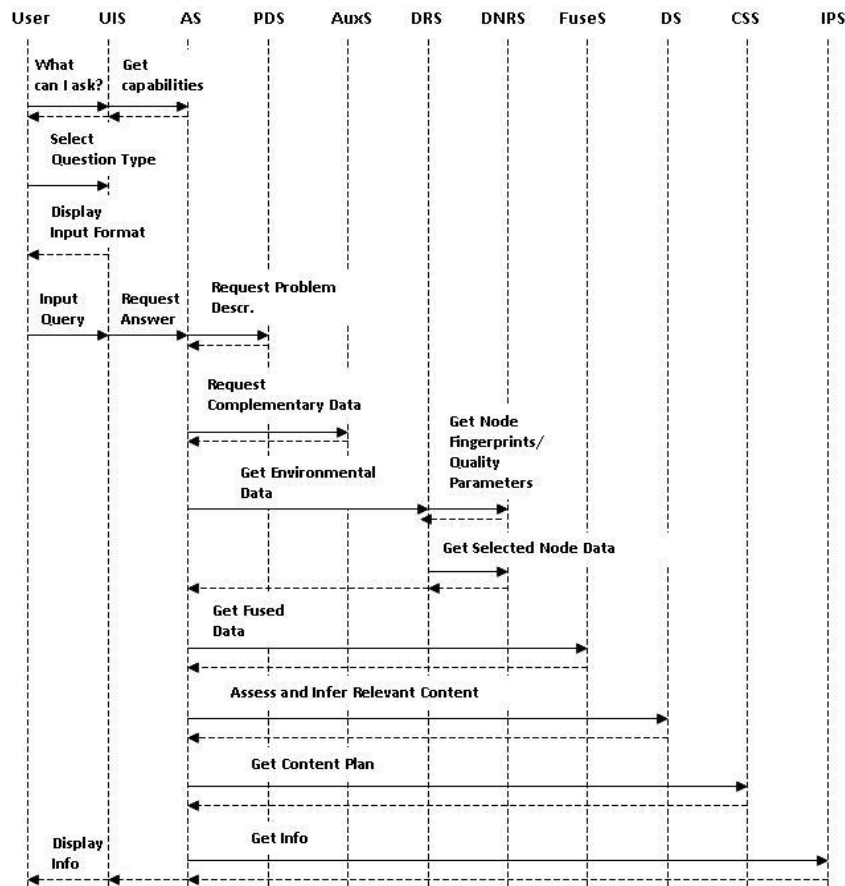


Fig. 3. Sequence diagram service execution for delivery of environmental information

A number of the above services as well as the interaction between selected services have already been discussed in other publications; see, for instance, [14] for the presentation of environmental node orchestration in PESCaDO; [15] for the ontology

management and [16] for the interaction of the DS and CSS and IPS. Therefore, and also due to the lack of space, let us focus in what follows on one aspect of PESCaDO – namely the discovery and extraction of environmental information from the web.

4 Discovery and Extraction of Environmental Information

The discovery of environmental nodes can be considered a problem of domain-specific web search and therefore methodologies from this area can be applied to implement a node discovery framework; see Figure 4 for the architecture. We apply two types of methodologies of domain search: (a) the use of general purpose search engines for the submission of domain-specific queries, and (b) focused crawling of predetermined websites [4]. To generate queries for the general purpose search engine we combine domain information from ontologies and geographical data obtained by geographical web services. In this case, the Yahoo BOSS API³ is used as a general purpose search engine. The queries are expanded by keyword spices [5], which are domain specific keywords extracted with the aid of machine learning techniques from environmental websites. In parallel, a set of predefined environmental websites is enriched using a focused crawler, which is capable of exploring the web in a directed fashion in order to collect other nodes that satisfy specific criteria related to the content of the source pages and the link structure of the web. The focused crawler is built upon Apache Nutch (<http://nutch.apache.org/>)

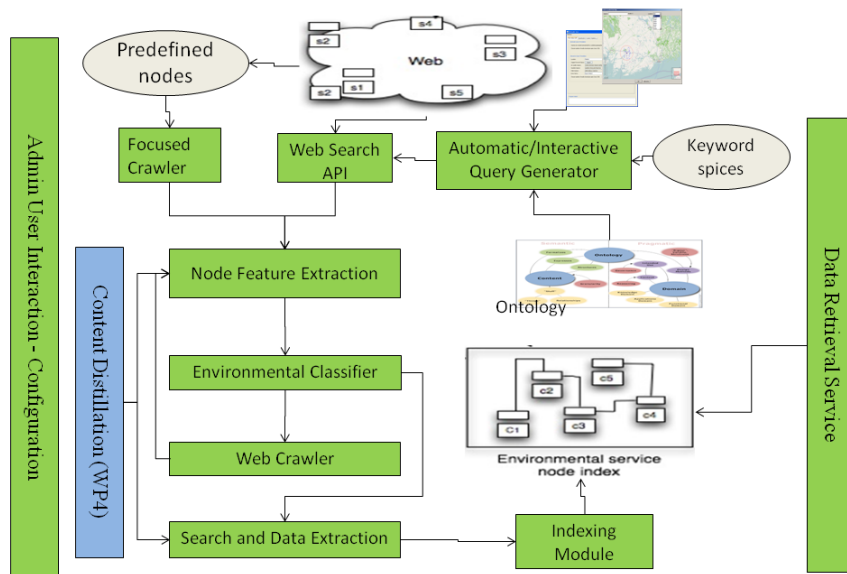


Fig. 4. Architecture for the discovery of environmental service nodes

³ <http://developer.yahoo.com/search/boss/>

Since the output of the search engine and the focused crawler include also many irrelevant results we employ a post-processing classification step in order to improve the precision of the discovery phase, which is realized with the aid of Support Vector Machine (SVM) [6] classifiers trained with manually annotated shots and textual features extracted with KX [7], which is a key phrase extraction tool.

The whole discovery procedure is automatic. However, an administrative user could intervene through an interactive graphical user interface in order to select geographic regions of interest to perform the discovery, optimize the selection of keyword spaces and parameterize the training of the classifiers. Figure 3 above shows the architecture of the discovery of environmental service nodes.

The next step is to extract the measurement data from the environmental nodes in order to index them in a database. It was observed that usually the weather information is mostly encoded in textual formats in the websites, while the air quality and pollen are usually presented in heatmap images.

To distill the textual data from webpages, advanced natural language processing techniques are needed for webpage parsing, information extraction and text mining. Although these techniques can be tuned to deal with the presentation mode of environmental data and information, this task remains very challenging and still only a manually assisted extraction of such information can be supported due to the high variety of websites and presentation formats. In the case of image data, also a semi-automatic procedure is realized. Specifically, we have implemented a configuration tool, in which the administrative user can identify the important images after the discovery process and annotate specific areas in the heatmap (i.e. coordinates, title, map, etc.). Once the configuration is finalized, the template file is used by a data extraction service to automatically retrieve information from this image in the future. The extraction service is built upon OCR technologies for identifying text on the image, while the image is converted to numerical data with the AirMerge tool [8].

Finally the environmental node data are stored and indexed in a Sensor Observation Service (SOS) [9] compliant repository.

For further details on the discovery of environmental nodes in PESCaDO, see [17].

5 Conclusions

In this paper, we have presented a personalized platform based on heterogeneous technologies which can support individuals in planning activities with respect to environmental conditions. As shown in the demonstration, the system can improve the quality of life, since it can offer very important suggestions to people taking into account their health conditions, the intended activities and the environmental conditions.

Acknowledgments. This work is partially funded by the European Commission under the contract number FP7-248594 “Personalized Environmental Service Configuration and Delivery Orchestration” (PESCaDO).

6 References

1. Karatzas, K.: State-of-the-art in the dissemination of AQ information to the general public. In: Proceedings of EnviroInfo, vol. 2, pp. 41--47. Warsaw (2007)
2. Peinel, G., Rose, T., San José R.: Customized Information Services for Environmental Awareness in Urban Areas. In: Proceedings of the 7th World Congress on Intelligent Transport Systems. Turin (2000)
3. Wanner, L., Bohnet B., Bouayad-Agha, N., Lareau F., Nicklass, D.: MARQUIS: Generation of User-Tailored Multilingual Air Quality Bulletins. *J. Applied Artificial Intelligence*. 24 (10), 914--952 (2010)
4. Wöber, K.: Domain Specific Search Engines, In: Fesenmaier, D. R., Werthner, H., Wöber, K. (eds.) *Travel Destination Recommendation Systems: Behavioral Foundations and Applications*, 205—226. Cambridge, MA: CAB International, (2006)
5. Oyama, S., Kokubo, T., Ishida, T.: Domain-Specific Web Search with Keyword Spices Awareness in Urban Areas. *J. IEEE Transactions on Knowledge and Data Engineering*. 16 (1), 17--24 (2004)
6. Boser, B. E., Guyon, I. M. and Va, V. N.: A training algorithm for optimal margin classifiers. In: COLT '92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. New York, NY, USA: ACM Press, pp. 144–152 (1992)
7. Pianta, E. and Tonelli, S.: KX: A Flexible System for Keyphrase Extraction. In: Proceedings of SemEval 2010, Uppsala, Sweden (2010)
8. Epitropou V., Karatzas K. and Bassoukos A. (2010), A method for the inverse reconstruction of environmental data applicable at the Chemical Weather portal, In: Proceedings of the GI-Forum Symposium and exhibit on applied Geoinformatics, pp. 58-68, Wichmann Verlag, ISBN 978-87907-496-9
9. Sensor Observation Service (SOS), https://wiki.52north.org/bin/view/Sensornet/SensorObservationService#SOS_tutorial
10. World Wide Web Consortium: OWL Web Ontology Language Reference, <http://www.w3.org/TR/owl-overview/>
11. Usländer, T. (ed.): Reference Model for the ORCHESTRA Architecture Version 2.1. OGC Best Practices Document 07-097, http://portal.opengeospatial.org/files/?artifact_id=23286 (2007)
12. Usländer, T.: Specification of the Sensor Service Architecture, Version 3.0 (Rev. 3.1). OGC Discussion Paper 09-132r1. Deliverable D2.3.4 of the European Project SANY, FP6-IST-033564, http://portal.opengeospatial.org/files/?artifact_id=35888&version=1 (2009)
13. Cooper, A.: *The Inmates Are Running the Asylum*. Sams Publishing, Indianapolis, (1999).
14. Epitropou, V., Johansson, L., Karatzas, K.D., Bassoukos, A., Karppinen, A., Kukkonen, J. and Haakana, M.: Fusion of Environmental Information for the Delivery of Orchestrated Services for the Atmospheric Environment in the PESCaDO project. In Proceedings of the International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Leipzig, Germany, (2012)
15. Rospocher, M. and Moßgraber, J.: Ontology Management in a Service-oriented Architecture. In Proceedings of the International Workshop on Web Semantics and Information Processing, (2012)
16. Bouayad-Agha, N., Casamayor G., Mille S., Rospocher M., Saggion H., Serafini L., Wanner L.: From Ontology to NL: Generation of Multilingual User-Oriented Environmental Reports. In NLDB'12, (2012)
17. Moutzidou, A., Vrochidis, S., Tonelli, S., Kompatsiaris, I. and Pianta, E.: Discovery of Environmental Nodes in the Web. In Proceedings of the 5th Information Retrieval Facility Conference, Austria, Vienna, (2012)