Personalized Environmental Service Configuration and Delivery Orchestration: The PESCaDO Demonstrator

Leo Wanner[1,2], Marco Rospocher[8], Stefanos Vrochidis[6], Harald Bosch[3], Nadjet Bouayad-Agha[2], Ulrich Bügel[4], Gerard Casamayor[2], Thomas Ertl[3], Desiree Hilbring[4], Ari Karppinen[5], Ioannis Kompatsiaris[6], Tarja Koskentalo[7], Simon Mille[2], Jürgen Moßgraber[4], Anastasia Moumtzidou[6], Maria Myllynen[7], Emanuele Pianta[8], Horacio Saggion[2], Luciano Serafini[8], Virpi Tarvainen[5], Sara Tonelli[8]

[1]Catalan Institute for Research and Advanced Studies, [2]Dept. of Information and Communication Technologies, Pompeu Fabra University, [3]Visualization Institute, University of Stuttgart, [4]Fraunhofer Institute for Optronics, System Technologies and Image Exploitation, [5]Finnish Meteorological Institute, [6]Informatics and Telematics Institute, Centre for Research and Technology Hellas, [7]Helsinki Region Environmental Services Authority, [8]Fondazione Bruno Kessler
pescado@upf.edu; http://www.pescado-project.eu

Abstract. Citizens are increasingly aware of the influence of environmental and meteorological conditions on the quality of their life. This results in an increasing demand for personalized environmental information, i.e., information that is tailored to citizens' specific context and background. In this demonstration, we present an environmental information system that addresses this demand in its full complexity in the context of the PESCaDO EU project. Specifically, we will show a system that supports submission of user generated queries related to environmental conditions. From the technical point of view, the system is tuned to discover reliable data in the web and to process these data in order to convert them into knowledge, which is stored in a dedicated repository. At run time, this information is transferred into an ontology-based knowledge base, from which then information relevant to the specific user is deduced and communicated in the language of their preference.

1 Research background

Citizens are increasingly aware of the influence of environmental and meteorological conditions on the quality of their life. 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, pollen, and traffic) and take into account a number of specific personal attributes (health, age, allergies, etc.) of the user, as well as the intended use of the information. For instance, a pollen allergic person, planning to do some outdoor activities, may be interested in being notified whether the pollen situation in the area may trigger some symptoms, or if the temperature is too hot for doing physical exercise, while a city administrator has to be informed whether the current air quality situation requires some actions to be urgently taken.

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 [2], [7], [9]. We aim to address the above task in its full complexity.

In this work, carried on in the context of the PESCaDO EU project, 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 users preferred mode.

The demonstration will aim, in particular, at showing how semantic web technologies are exploited to address this challenges in PESCaDO.

2 The PESCaDO Platform: Main modules and key semantic technologies used

The challenges mentioned in Section 1 require the involvement of an elevated number of rather heterogeneous applications addressing various complex tasks: discovery of the environmental service nodes in the web, distillation of the data from webpages, orchestration of the environmental service nodes, fusion of environmental data, assessment of the data with respect to the needs of the addressee, selection of user-relevant content and its delivery to the addressee, and, finally, interaction with the user. Thus, in PESCaDO we developed a service-based infrastructure to integrate all these applications.

For a general overview of the running PESCaDO service platform[1], and the type of information produced, see: http://www.youtube.com/watch?v=c1Ym7ys3HCg. In this section, we focus on presenting three tasks we addressed by applying semantic web technologies.

The back-bone of the PESCaDO service platform, exploited in each of these three tasks, is an ontology-based knowledge base, the PESCaDO Knowledge Base (PKB), where all the information relevant for a user request are dynamically instantiated. The ontology, partially built exploiting automatic key-phrases extraction techniques [8], formalizes a variety of aspects related to the application context: environmental data, environmental nodes[2], user requests, user profiles, warnings and recommendations triggered by environmental conditions, logico-semantic relations (e.g. cause, implication) between facts, and so on. The current version of the ontology consists of 241 classes, 672 individuals, 151 object properties, and 43 datatype properties.

2.1 Discovery of environmental nodes

The first step towards the extraction and indexing of environmental information is the discovery of environmental nodes, which can be considered as a problem of domain spe-

[1] A more comprehensive description of the system workflow can be found in [10]

[2] An environmental node is a provider of environmental data values, like for instance a web-site, a web-service, or a measuring station

cific search. To this end, we implement a node discovery framework, which builds upon state of the art domain specific search techniques, advanced content distillation, ontologies and supervised machine learning. The framework consists of three main parts: a) Web search b) Post processing and c) Indexing and storage. Web search is realized with the aid of a general-purpose search engine, which accesses large web indices. In this implementation we employ Yahoo! Search BOSS API. In order to generate domain specific queries, we apply two complementary techniques. First we use the ontology of the PKB and we extract concepts and instances referring to types of environmental data (e.g. temperature, birch pollen, PM_{10}) and we combine them with geographical city names automatically retrieved by geographical resources. In addition, the queries are expanded by keyword spices [6], which are domain specific keywords extracted with the aid of machine learning techniques from environmental websites.

During the post-processing step we perform supervised classification with Support Vector Machines to separate relevant from irrelevant nodes and we crawl each website to further expand our search in an iterative manner. The determination of the relevance of the nodes and their categorization is done using a classifier that operates on a weight-based vector of key phrases and concepts from the content and the structure of the webpages. Subsequently, we parse the body and the metadata of the relevant webpages in order to extract the structure and the clues that reveal the information presented.

Finally, the information obtained with respect to each relevant node is indexed in a Sensor Observation Service (SOS) [5] compliant repository, which can be accessed and retrieved by the system when a user request is submitted.

The whole discovery procedure is automatic, however an administrative user could intervene through an interactive user interface, in order to select geographic regions of interest to perform the discovery, optimize the selection of keyword spices, and parametrize the training of the classifiers.

2.2 Processing raw environmental data to obtain content

The user interface of the PESCaDO system guides the user in formulating a request, which is instantiated in all its details (e.g. type of request, user profile, time period, geographic location) in the PKB. By exploiting Description Logics (DL) reasoning on the PKB, the system determines from the request description which are the types of environmental data which constitute the raw content necessary to fulfil the user needs. A specific component of the system is then responsible of selecting from the SOS repository the actual values (observed, forecasted, historical) for the selected types of environmental data, and to appropriately instantiate them in the PKB.

At this stage, the raw data retrieved from the environmental nodes are processed to derive additional personalized content from them, such as data aggregations, qualitative scaling of numerical data, and user tailored recommendations and warnings triggered by the environmental data relevant for the specific user query. Logico-semantic relations are also instantiated at this stage, for instance to represent whether a certain pollen concentration value causes the triggering of a recommendation to the user, due to its sensitiveness to that pollen.

The computation of this inferred content is performed by the *decision support* service of the PESCaDO Platform by combining some complementary reasoning strategies, including DL reasoning and rule-based reasoning. A two layer reasoning infrastructure is currently in place. The first layer exploits the HermiT reasoner for the OWL DL reasoning services. The second layer is stacked on top of the previous layer. It uses the Jena RETE rule engine, which performs the rule-based reasoning computation.

2.3 Generating user information from content

As is common in Natural Language Generation, our information generator is divided into two major modules: the text planning module and the linguistic generation module (with the latter taking as input the *text plan* produced by the former).

Text Planning The text planning module is divided into a content selection module and discourse structuring module. As is common in report generation, our content selection is schema- (or template-) based. Therefore, the ontology of the PKB introduced above defines a class `Schema` with an n-ary schema component object property whose range can be any individuals of the PKB itself.

Similar to [1], we assume the output of the discourse structuring module to be a well-formed text plan which consists of (i) elementary discourse units (EDUs) that group together individuals of the PKB, (ii) discourse relations between EDUs and/or individuals of the PKB, and (iii) precedence relations between EDUs. This structure translates into two top classes of the ontology of the PKB: `EDU` with an n-ary EDU component relation and a linear precedence property, and `Discourse Relation` with nucleus and satellite relation. A set of SPARQL query rules are defined to instantiate the various concepts and relations.

Content Selection (CS) operates on the output of the decision support service. It selects the content to be included in the report and groups it by topic, instantiating a number of schemas for each topic. The inclusion of a given individual in a schema can be subject to some restrictions defined in the queries; for example, if the minimum and maximum air quality index (AQI) values are identical, or if the maximum AQI value triggers a user recommendation or warning, then only the maximum AQI value is selected (the minimum AQI rating is omitted).

Discourse structuring is carried out by a pipeline of three rule-based submodules: (i) Elementary Discourse Unit (EDU) Determination, which groups topically related PKB individuals into propositional units starting from the schemas determined during CS; (ii) Mapping logico-semantic relations to discourse relations; and (iii) EDU Ordering, which introduces a precedence relation between EDUs using a number of heuristics derived from interviews with domain communication experts.

Linguistic generation Our linguistic generation module is based on a multilevel linguistic model of the Meaning-Text Theory (MTM) [4], such that the generation consists of a series of mappings between structures of adjacent strata (from the conceptual stratum to the linguistic surface stratum): *Conceptual Structure* (*ConStr*) $\Rightarrow$ *Semantic Structure* (*SemStr*) $\Rightarrow$ *Deep-Syntactic Structure* (*DSyntStr*) $\Rightarrow$ *Surface-Syntactic Structure* (*SSyntStr*) $\Rightarrow$ *Deep-Morphological Structure* (*DMorphStr*) $\Rightarrow$ *Surface-Morphological Structure* (*SMorphStr*) $\Rightarrow$ *Text*. Starting from the conceptual stratum, for each pair of adjacent strata $\mathcal{S}_i$ and $\mathcal{S}_{i+1}$, a transition grammar $\mathcal{G}^i_{i+1}$ is defined; see [3].

The ConStr is derived from each text plan produced by the text planning component. In a sense, ConStr can thus be considered a projection of selected fragments of the ontologies onto a linguistically motivated structure. ConStrs are language-independent and thus ideal as starting point of multilingual generation.

3 System Demonstration

The system demonstration will show how the PKB is instantiated and exploited by the different services composing the PESCaDO Platform in the context of two different application scenarios, one about health safety decision support for end users and one about administrative decision support. In particular, the demo attendees will have the chance to see how the raw environmental data are dynamically processed with ontology-based techniques to obtain reports. Furthermore, we will demonstrate how to use and set-up the tool for environmental node discovery.

Acknowledgments

The work described in this paper has been partially funded by the European Commission under the contract number FP7-248594, PESCaDO (Personalized Environmental Service Configuration and Delivery Orchestration) project.

References

1. N. Bouayad-Agha, G. Casamayor, L. Wanner, F. Díez, and S. López Hernández. Footbowl: Using a generic ontology of football competition for planning match summaries. In *Proceedings of the Eighth Extended Semantic Web Conference (ESWC)*, pages 230–244, 2011.
2. Kostas D. Karatzas. State-of-the-art in the dissemination of aq information to the general public. In *Proceedings of EnviroInfo*, pages 41–47, 2007.
3. F. Lareau and L. Wanner. Towards a generic multilingual dependency grammar for text generation. In T. King and E.M. Bender, editors, *Proceedings of the GEAF07 Workshop*, pages 203–223, Stanford, CA, 2007. CSLI.
4. I.A. Mel'čuk. *Dependency Syntax: Theory and Practice*. SUNY Press, Albany, 1988.
5. 52 North. Sensor observation service (sos), 2004.
6. S Oyama, T Kokubo, and T Ishida. Domain-specific web search with keyword spices. *IEEE Transactions on Knowledge and Data Engineering*, 16(1):17–27, 2004.
7. G. Peinel, Rose T., and R. San José. Customized information services for environmental awareness in urban areas. In *Proceedings of the 7th World Congress on Intel ligent Transport Systems*, Turin, Italy, 2000.
8. S. Tonelli, M. Rospocher, E. Pianta, and L. Serafini. Boosting collaborative ontology building with key-concept extraction. In *Proceedings of 5th IEEE International Conference on Semantic Computing (September 18-21, 2011 - Palo Alto, USA)*, 2011.
9. L. Wanner, B. Bohnet, N. Bouayad-Agha, F. Lareau, and D. Nicklaß. MARQUIS: Generation of user-tailored multilingual air quality bulletins. *Applied Artificial Intelligence*, 24(10):914–952, 2010.
10. L. Wanner, S. Vrochidis, S. Tonelli, J. Moßgraber, H. Bosch, A. Karppinen, M. Myllynen, M. Rospocher, N. Bouayad-Agha, U. Bügel, G. Casamayor, T. Ertl, I. Kompatsiaris, T. Koskentalo, S. Mille, A. Moumtzidou, E. Pianta, H. Saggion, L. Serafini, and V. Tarvainen. Building an environmental information system for personalized content delivery. In *Proceedings of the ISESS 2011, Brno, Czech Republic*, pages 169–176. Springer, 2011.