Expert Systems with Applications 42 (2015) 5032–5046

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Ontology-centered environmental information delivery for personalized decision support

Leo Wanner [a,e,*], Marco Rospocher [b], Stefanos Vrochidis [c], Lasse Johansson [d], Nadjet Bouayad-Agha [e], Gerard Casamayor [e], Ari Karppinen [d], Ioannis Kompatsiaris [c], Simon Mille [e], Anastasia Moumtzidou [c], Luciano Serafini [b]

[a] *Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain*
[b] *Fondazione Bruno Kessler, Trento, Italy*
[c] *CERTH/ITI, Thessaloniki, Greece*
[d] *Finnish Meteorological Institute, Helsinki, Finland*
[e] *Pompeu Fabra University, Barcelona, Spain*

A R T I C L E I N F O

Article history:
Available online 6 March 2015

Keywords:
Environmental conditions
Decision support
Ontologies
Reasoning
Search
Multilingual personalized information generation

A B S T R A C T

Data on observed and forecasted environmental conditions, such as weather, air quality and pollen, are offered in a great variety in the web and serve as basis for decisions taken by a wide range of the population. However, the value of these data is limited because their quality varies largely and because the burden of their interpretation in the light of a specific context and in the light of the specific needs of a user is left to the user herself. To remove this burden from the user, we propose an environmental Decision Support System (DSS) model with an ontology-based knowledge base as its integrative core. The availability of an ontological knowledge representation allows us to encode in a uniform format all knowledge that is involved (environmental background knowledge, the characteristic features of the profile of the user, the formal description of the user request, measured or forecasted environmental data, etc.) and apply advanced reasoning techniques on it. The result is an advanced DSS that provides high quality environmental information for personalized decision support.

1. Introduction

Detailed and comprehensive information on observed and forecasted environmental conditions such as weather, air quality and pollen is primary for the assessment of sanitary risks and reasoned daily life decisions for the entire population. Citizens are increasingly aware of it. In the era of the World Wide Web and the increasing digitalization of our society, the bulk of this information comes from the internet. However, as offered in the internet, it is utterly inadequate for assessment by citizens, be it with respect to sanitary risks, societal commitment or social interaction. Firstly, as already a very cursory glance at a few webpages shows, the predictions of the upcoming environmental conditions vary largely. This implies that the quality of the offered data is uncertain. Secondly, the information is, in general, given in terms of raw data or qualitative indices, presented in tables, distribution curves or color scales following the "same information for all" philosophy. This implies that the burden of the interpretation of the data in the personal context of the user and in the context of a specific decision that the user has to take is left to the user herself.

In order to offer appropriate support to citizens in questions related to environmental conditions in their habitat, a novel expert system model must be designed. The design must target the removal of the burden from the user to guess the quality of the environmental webpages in the internet and then to interpret the data of the page chosen perforce arbitrarily in order to come up with a decision concerned with a planned action or precautionary measure.

At the generic level, the main tasks of an environmental decision support model are reflected by the three phases commonly identified in the literature for *Decision Support Systems* (DSSs) (Laskey, 2006):

1. formulation of the decision-making problem;

* Corresponding author at: ICREA and Pompeu Fabra University, Barcelona, Spain.
E-mail addresses: leo.wanner@upf.edu (L. Wanner), rospocher@fbk.eu (M. Rospocher), stefanos@iti.gr (S. Vrochidis), lasse.johansson@fmi.fi (L. Johansson), nadjet.bouayad@upf.edu (N. Bouayad-Agha), gerard.casamayor@upf.edu (G. Casamayor), ari.karppinen@fmi.fi (A. Karppinen), ikom@iti.gr (I. Kompatsiaris), simon.mille@upf.edu (S. Mille), moumtzid@iti.gr (A. Moumtzidou), serafini@fbk.eu (L. Serafini).

http://dx.doi.org/10.1016/j.eswa.2015.02.048
0957-4174/

2. gathering, storing, and fusion of the environmental data relevant to the given problem;
3. reasoning on the data to take a decision.

The implementation of these three phases is usually casted into three modules (Marakas, 1999): (i) a *dialog* or *user* module, which supports the interaction of the user with the system in the formulation of the problem and provision of the result of the DSS computation as output, (ii) a *data* module, which stores the data collected and processed by the DSS, and (iii) a *model* module, which implements the decision support strategy. The realization of an environmental DSS that accounts for the specifics of web-distributed environmental data and the needs of the users in this domain requires a refinement and extension of these modules in that

- to account for the wide range of topics in the environmental domain for which decision support may be solicited, it is essential to formalize the typologies of user requests and relate them to relevant data;
- to be able to provide requested decision support, the data module must contain, apart from environmental data collected from the web, background environmental knowledge, legal warnings, etc.;
- to guarantee that the best and most comprehensive data are used by the model module and communicated to the user, an intelligent data assessment and fusion mechanism must operate on the data module;
- to be able to interpret the collected and assessed data in the light of the background knowledge and the context of user's decision support request, advanced reasoning techniques must be used;
- to ensure the adequate communication of environmental information, recommendations and other decision support hints to the user, the model module must incorporate a full-fledged multilingual natural language generator that selects the content relevant to the inquiry of the user and presents it in the language of her preference.

All of these extensions have in common that they produce or use knowledge as represented in ontologies. Or, in other words, ontologies must be the integrative core of an advanced environmental DSS.

During the last decade, research on DSSs has taken advantage of the achievements in the field of Semantic Web technologies in general, and in the area of ontologies in particular (Ceccaroni, Cortés, & Sànchez-Marrè, 2004; Choras, Kozik, Flizikowski, Renk, & Holubowicz, 2009; Ishizu, Gehrmann, Minegishi, & Nagai, 2008; Saremi, Esmaeili, Habibi, & Ghaffari, 2008; Wyner, 2008; Esposito & De Pietro, 2011; Pangjitt & Sunetnanta, 2011; Zagorulko & Zagorulko, 2010). However, to the best of our knowledge, in none of them ontologies assume the role of an integrative core in the sense that (i) all functions of a DSS operate on ontologies – to feed them, to reason over them or to generate out of them the decision support information; (ii) ontologies are designed to serve all modules of a DSS – which also includes that they are enriched by knowledge elements needed to extend the DSS by novel features (such as dynamic generation of personalized decision support information). Furthermore, despite the fact that ontologies have been extensively used to represent environmental phenomena (see, e.g., the SWEET Ontology[1]) and semantically describe environmental data (see, e.g., the Semantic Sensor Network Ontology[2]), they have had so far small repercussion in state-of-the-art *environmental* DSSs: only few works describe how to exploit ontologies, and in those that do, ontologies have been adopted to support only parts of the environmental DSS. Thus, Ceccaroni et al. (2004) use ontologies in the traditional sense: a static domain ontology models the knowledge on the physical, chemical and microbiological environment of a waste-water treatment plant and provides reasoning support for the decision-making phase. In (Zagorulko & Zagorulko, 2010), the primary use of ontologies is for data representation in the context of power consumption management.

In what follows, we propose to exploit an ontology-based knowledge base (KB) as the main (enhanced) data structure of a DSS, where *all* the content and data for a specific decision support request processed and produced by the system is stored. The ontological representation is used as the main shared data structure accessed by all modules of the DSS and as an intermediate content exchange format. This allows us to design decision support as an open web-based service that facilitates the integration of structured knowledge and data available in the web in heterogeneous formats in terms of shared domain models, and that can be combined with other complex services, such as explanation services or services of other applications. The DSS has been implemented and validated in the PESCaDO[3] service; see (Wanner et al., 2014) for a general presentation of PESCaDO.

The remainder of the paper is structured as follows. In Section 2, we outline the design of the environmental DSS ontology and the architecture of a DSS with such an ontology as its integrative core. The following sections address the realization of the three main phases of decision support: Section 3 deals with the formulation of the problem, Section 4 with data acquisition and fusion, and Section 5 with the interpretation, selection and communication of decision support relevant content to the user. In Section 6, the main modules of the proposed environmental DSS and the whole system are evaluated. In Section 7, finally, some conclusions from the presented work are drawn and lines of future work that shall further advance it are outlined.

2. Designing the architecture of an ontology-centric DSS

The architecture of our ontology-centric DSS is illustrated in Fig. 1. In accordance with the three-partite division of the decision support tasks (formulation of the problem, data processing, and decision support), this architecture consists of three modules (see Sections 3–5) and the ontology-based KB as its core.

To support all subtasks dealt with by the three main modules of the DSS, the KB must capture any data and knowledge these subtasks may require. The required data concern, in particular, the observed, forecasted and derived environmental data of the region relevant to the user inquiry. The range of required knowledge is broader. It involves, for instance, typologies of user requests, plans and actions supported by the DSS, user models, and background environmental, health-related and linguistic information needed to come up with a reasoned well-formulated recommendation or description of meteorological and environmental conditions. Fig. 2 shows the design of an ontology-based KB[4] for environmental decision support that foresees the representation of all of these data and knowledge; see also (Rospocher & Serafini, 2012) for more details. The upper part of the figure, referred to as "*T-Box*" of the KB in Description Logics (DL) terminology, depicts the *Decision Support Ontology* (DSO). This is the ontology that underlies the KB. It comprises data/content type-specific partitions, which are grouped into three components. Each component corresponds to one of the three decision-making phases of a traditional DSS:

[1] http://sweet.jpl.nasa.gov/ontology/
[2] http://www.w3.org/2005/Incubator/ssn/ssnx/ssn

[3] PESCaDO: "Personalized Environmental Service Configuration and Delivery Orchestration".
[4] The KB is specified in terms of the Web Ontology Language OWL (http://www.w3.org/TR/owl2-overview/)

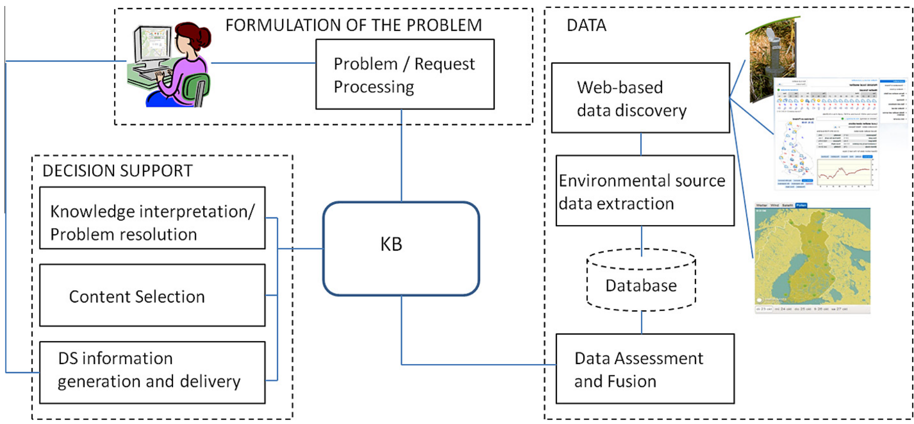

Fig. 1. Architecture of the PESCaDO environmental DSS with a KB as its integrative core.

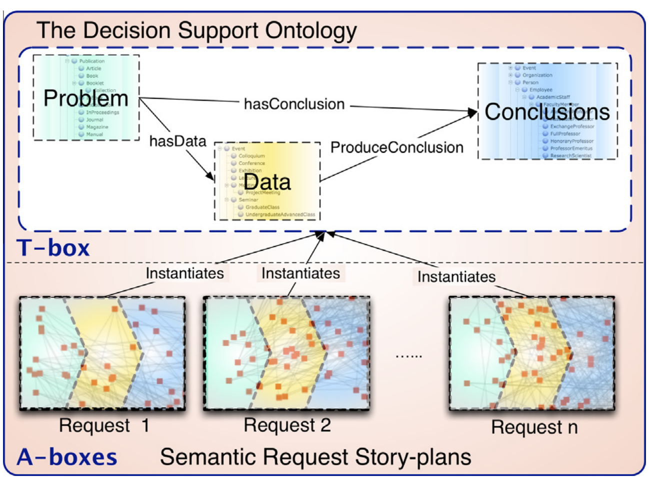

Fig. 2. Design of a knowledge base for environmental DSS.

Problem (formulation), Data (handling), and Conclusions (derivation for Decision Support).

The Problem component formally describes the knowledge needed to interpret a user request. This knowledge concerns:

(i) the type of the request (e.g., "Is there any health issue for me?", "Do environmental conditions require to take some administrative actions?", etc.);
(ii) the reason behind the request, i.e., the type of activity the user plans (e.g., "physical outdoor activity", "car trip", etc.);
(iii) the user characteristics, i.e., the profile of the user who launched the request (e.g., end-user or administrative user, her age, gender, diseases or allergies, etc.).

Accordingly, the Problem component contains three subcomponents; see Fig. 3. The three subcomponents are interrelated by object properties and OWL restrictions (e.g., a request requires a user profile associated with it and may involve an activity the user wants to undertake) that constrain the possible combinations.

The Data component describes the environmental data used by the system to provide decision support – including meteorological data (e.g., temperature, wind direction and strength), pollen data, and air quality data (e.g., NO_2, PM_{10}). The description is detailed enough to comprehensively depict observed, forecasted, and historical (quantitative and qualitative) data, the time period covered by the data, and the type of the data (e.g., instantaneous, average, maximum). In addition to environmental data, a series of other related data are captured: traffic and road conditions, profile of the environmental node from which the data stem (e.g., measurement station, website), its geographical location, confidence rating, qualitative and quantitative data alignment (as, e.g., "*moderate* quantity of birch pollen" ≡ 10–100 pollen grains/m^3), etc.

To build up the Data component, we reuse parts of available environmental domain ontologies such as SWEET, apply techniques for automatic ontology extension (Tonelli, Rospocher, Pianta, & Serafini, 2011), and incorporate contributions of environmental domain experts obtained via interviews.

The Conclusions component has two functions. Its first function is to encode conclusions such as warnings, recommendations and suggestions that may be triggered by environmental conditions or the exceedance of legal thresholds by pollutant concentrations that may be detected in the data. Cf., for illustration, an example of a warning in Fig. 4, together with the associated message to be reported to the users (available in the languages supported in the experimental PESCaDO setup: English, Finnish, and Swedish). A warning issued by the DSO for a given request is represented as a new instance, with an object property asserting the type of warning, and a data type property asserting the relevance (a decimal value in [0 ... 1]) of the warning for the request.

Its second function is to introduce *logico-semantic relations* (LSRs) between entities of the Data component in order to facilitate the deduction of discourse relations for the generation of textual information from these entities, should they be considered relevant to the inquiry and profile of the user. Fig. 5 shows the list of LSRs covered in the DSO (left hand side) and illustrates one of them (namely 'Implication') in more detail (right hand side).

The three components of the DSO are connected by the relations *hasData*, *hasConclusion* and *ProduceConclusion* (see Fig. 2). *hasData* links a problem description with the data relevant to it and *hasConclusion* with the conclusions provided for it by the DSS. *ProduceConclusion* connects the data with the conclusions they trigger. These properties are particularly useful for explanation. For instance, *ProduceConclusion* allows the system to keep track of what data triggered a certain DSS conclusion.

The current DSO[5] consists of 210 classes, 99 object properties, 42 datatype properties, and 641 individuals; cf. Rospocher (2014) for more details.

As shown in the lower part of Fig. 2, each single request submitted to the DSS triggers the instantiation of a new *A-Box* of the KB,[6] i.e., of a set of ontology individuals and assertions on them that fully

[5] Available for download at https://dkm-static.fbk.eu/people/rospocher/resources/pescado/

[6] To handle multiple simultaneous requests in PESCaDO, an *ontology pooling* mechanism has been developed (Mossgraber & Rospocher, 2012).

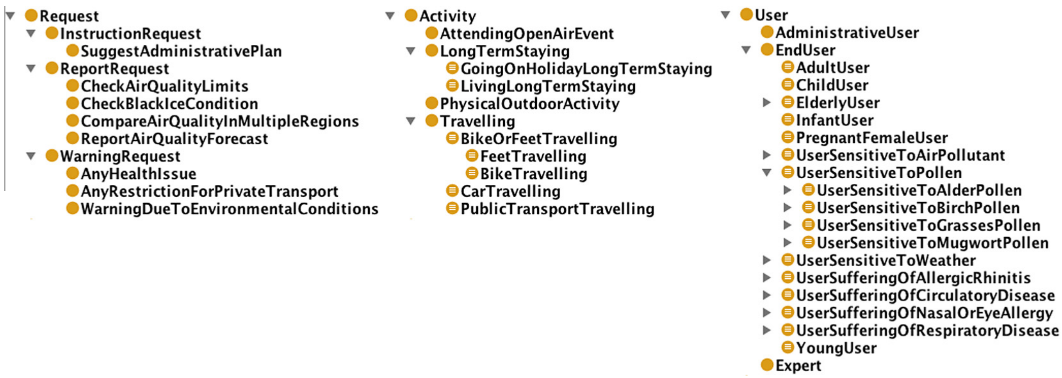

Fig. 3. The problem DSO component.

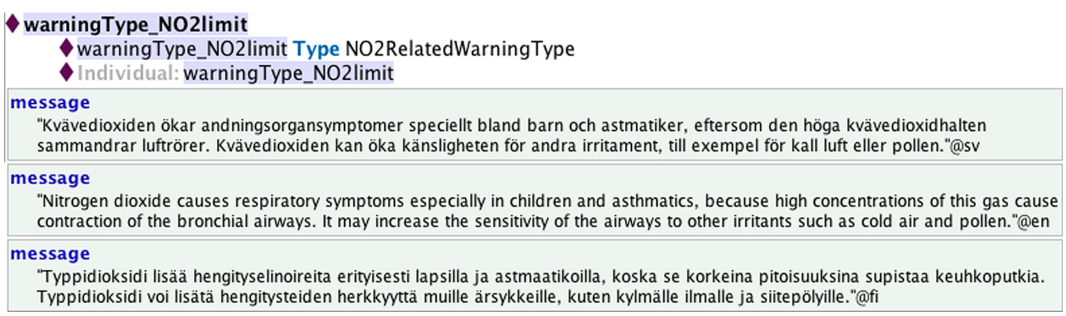

Fig. 4. The conclusions DSO component.

Fig. 5. The LSRs in the conclusions component of the DSO.

describe a specific request: the formulated problem, the specific data that were processed by the system, and the conclusions produced by the system based on the available data. The instantiation of each A-Box is done incrementally in a sequence of steps while the request is processed during the subsequent decision-making process phases.

Fig. 6 shows an excerpt of an A-box describing the instantiation of a problem, some data (namely those related to temperature) and some conclusions for the user request: "I want to go hiking in the Nukksio Park tomorrow/I suffer from birch pollen allergy/Will there be any health issue for me?". This request will be used as a running example to explain how the various modules create, read or process content in the ontology-based KB.

3. Formulation of the problem

The formulation of a request for decision support on a specific problem is controlled by a user interface (UI) wizard, which supervises the introduction of the information needed to fulfill the request. The wizard queries the Problem component of the DSO to retrieve the constraints on each item that are encoded by means of OWL class restrictions, e.g., the types of supported requests, the type of user admitted as requester, the obligatory specification of the time period and geographical area of the action on which decision support is requested, etc. For instance, there is a subclass restriction stating that an end user request for obtaining health-related warnings has to describe also the activity to be performed, which could be either traveling, doing some outdoor activities, or attending some outdoor event. It is important to note that these restrictions are not hardcoded in the UI or in the DSS. Rather, they are dynamically obtained from the ontology. This ensures the maximum flexibility on the operation and extension of the DSS. Fig. 7 shows the selection of the user activity for the request in our running example (note the summary of the request fields compiled so far in the "Your request" box).

Once the user introduced/selected all information items needed for her request, the request is instantiated, as mentioned above, in the A-Box in terms of the corresponding individuals and assertions. Like this, any subsequent module of the system is aware of it. Thus, if an end user submitted a request concerning health issues in

Fig. 6. Excerpt of the A-Box associated to a request processed by the system.

doing some activity, an individual X of type *AnyHealthIssue* and an individual Y of type *EndUser* are created and the assertion *hasRequestUser*(X, Y) is added. The complete instantiation of the request for our running example is shown in the top panel (“Problem”) of Fig. 6.

4. The data

The management of the data in a web data-oriented environmental DSS consists of three major tasks: (i) determination of the environmental data that have to be retrieved according to the user request; (ii) search and collection of the data in the web, and (iii) assessment and fusion of the collected data and their integration into the DSO.

4.1. Determining the relevant data

Once the problem description is instantiated in the A-Box associated to a user request, the DSS determines what data are of relevance to this request. This is again done drawing upon the information encoded in the DSO.

The DSO contains mappings between the classes describing requests, activities and user profiles in the Problem component and the types of environmental data defined in the Data component. The purpose of these mappings, which have been defined in collaboration with environmental and health experts, is to identify the data that are to be processed when for a given request + activity + user profile constellation the corresponding conclusions for the user are produced. The mappings are formalized as OWL *hasValue* restrictions on the classes of the Problem component of the ontology. For instance, a restriction of the form *hasRelevantAspect hasValue Rain* on the class that defines the users sensitive to birch pollen states that precipitation data are to be retrieved and processed when providing decision support to this type of user (Fig. 8 shows all the *hasValue* restrictions associated with the class representing users sensitive to birch pollen).

The definition of the mappings as *hasValue* restrictions facilitates the automatic determination of the environmental data types for which data are to be retrieved via Description Logics (DL)-based reasoning. This is done by checking the new assertions inferred by an OWL-reasoner for the request, user, and activity individuals that describe the problem in question. In the case of the request considered in our running example, the OWL reasoner determines

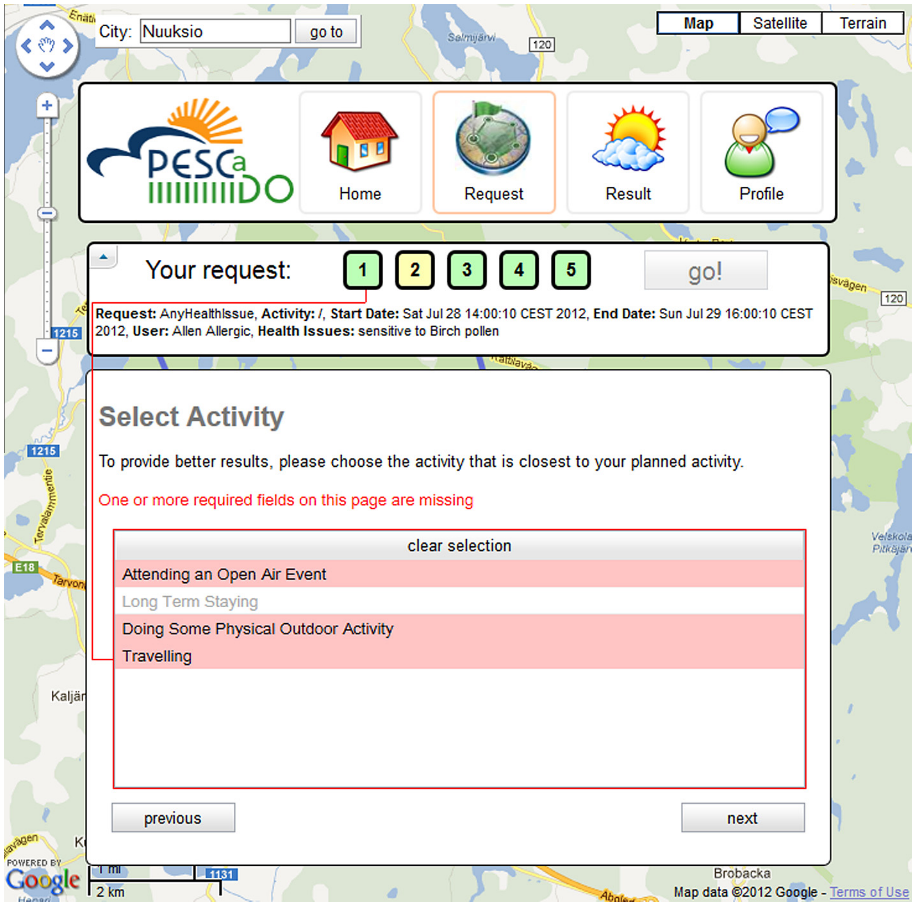

Fig. 7. Request generation wizard.

Fig. 8. The *hasValue* restrictions associated with the class of birch pollen allergic users.

that data about the following phenomena have to be retrieved: birch pollen, rain, wind speed, temperature, humidity, UV index, CO, PM_{10}, $PM_{2.5}$, NO_2, O_3, and SO.

4.2. Searching and collecting the data

The environmental data collection is performed in four phases: (i) environmental data discovery, (ii) content extraction, (iii) environmental data indexing and (iv) retrieval.

The goal of (i)–(iii) is to discover in the web environmental information for the geographical areas of interest and subsequently extract and store/index this information. They are realized in an offline mode. Phase (iv) is performed online (i.e., during the query time). Below, we present details on (i)–(iv), focusing on indexing and retrieval (iii and iv), which are directly relevant to the DS procedure and are therefore driven by the ontology.

4.2.1. Discovery of environmental information and content extraction

The discovery of environmental information webpages is performed using domain specific search techniques. To this end, we exploit an existing general purpose search engine (Yahoo BOSS[7]) by submitting to it domain-specific queries. The results are post-processed using supervised classification based on textual (Moumtzidou, Vrochidis, Tonelli, Kompatsiaris, & Pianta, 2012) and visual information (Moumtzidou, Vrochidis, Chatzilari, & Kompatsiaris, 2013). The discovery is optimized by an administrative user, who interacts with the system to improve and validate the results (Vrochidis et al., 2012).

The retrieved webpages include both textual passages (e.g., pollutant concentrations) and images (e.g., graphs, heatmaps). To distil data from text, advanced natural language parsing techniques were applied, while to transform semi-structured web content into structured data, regular expressions and HTML trees were used (Tonelli et al., 2011). Data extraction from images focused on heatmap analysis by employing a framework built upon Optical Character Recognition (OCR) to identify the text on the image, text processing to improve the quality of the retrieved text, and image processing to map the image information onto numerical values (Moumtzidou et al., 2014).

For our running example, the discovery procedure will have discovered and periodically extracted and indexed at least information relevant to pollen, weather and air pollution from several web resources for the region of Nukksio Park for the time period from: 2012-07-28, T14:00 + 02:00 to 2012-07-29, T16:00 + 02:00). The discovery is initiated by automatic formulation of domain-specific queries such as “birch pollen + Nukksio Park”, “temperature + Nukksio Park”, etc.

[7] http://developer.yahoo.com/boss/search/

4.2.2. Semantic indexing and retrieval of environmental information

To store efficiently the extracted environmental information in the data repository, indexing techniques that support spatial (i.e., geographical location) queries are required. Spatial queries are supported by services that handle sensor data. One of the most effective services for handling sensor data is the Sensor Observation Service (SOS) server (OGC, 2011). SOS server is a standardization of a Sensor Web network that is especially well suited for environmental monitoring. Sensor Web can be considered as a middleware between sensors and applications that involves three architectural layers: (i) the Sensor layer, (ii) the Sensor Web layer, and (iii) the Application layer (Bröring et al., 2011). The Sensor layer tackles, for instance, the definition of a resilient middleware for sensor network management (see, e.g., Benincasa, D'Aniello, Gaeta, Loia, & Orciuoli, 2015a) or, broader, the description of Situation Awareness mechanisms (see, e.g., Benincasa, D'Aniello, De Maio, Loia, & Orciuoli, 2015b for a human perception-oriented Situation Awareness approach). The Sensor Web layer comprises middleware solutions which are particularly designed for making sensors available on the web and enable the access to sensors from the application level by building up Sensor Web infrastructures.[8] The Application layer, finally, comprises, for instance, the emerging centralized web portals for sensors that are considered to be a new type of access to sensor resources for applications.

Each environmental node is described by the type of information/measurement it offers, the data, the date-time stamp, and the area it covers. Therefore, any environmental node can be considered a measurement sensor, and an SOS server can be used in the sense of the Sensor Web layer for the realization of indexing and retrieval of data from it. For this purpose, we implemented 52°North SOS,[9] version 3.1.1. To make the 52°North SOS implementation compatible with the proposed DSO and to facilitate data retrieval and exchange between the repository and the KB, specific enhancements were necessary. Fig. 9 depicts the tables holding the key information of the extended PostgreSQL database (DB) schema defined for the SOS to cover the additional information required for the DSO.[10]

Among the key tables of the DB schema are (i) a table for the data on the area to which the obtained environmental data refer (*feature_of_interest*) and (ii) a table with a short description of the measured phenomena, organized by kind, such as, e.g., 'weather': 'temperature', 'wind speed, wind direction, rain, …; 'air quality': 'CO', 'NO_2', 'O_3', etc.; and 'pollen': birch pollen', 'grass pollen' (*phenomenon*). Further tables include, for instance, a table with the types of the data and the time period for which data are available (*offering*), a table with the specifics of the environmental nodes that deliver the environmental data (*procedure*), and a table with the aggregated data of an observed event such as date and time, procedure, the feature of interest, the phenomenon and the value of the phenomenon (*observation*).

To align the DB schema with the ontology (and thus to allow for a direct data feed from the data repository to the KB), the phenomena in the phenomenon table of the DB have been combined with the value types in the ontology (e.g., *Rating*, or *AirQualityValue*) using the new table fields *phenomenon_owl_class*, and *valueType_owl_class*. Furthermore, OWL-concepts of the ontology have been included as *phenomenon_owl_class* into the phenomenon table of the DB.

After the user has specified her aspects of interest for a specific geographical area, a data retrieval service accesses the semantically enriched repository and provides the results in the SOS-compliant XML format. In the case of our running example, the SOS server receives from the data retrieval service the request with the following parameters:

a. the geographical area of Nukksio park, which is defined by the user through the design of a polygon on the map and is resolved to a group of coordinates;
b. the phenomena that may affect the user who is allergic to birch pollen during her outdoor activity (as defined by the reasoner; cf. Section 4.1): pollen, air pollutant substances O_3, SO_2, $PM_{2.5}$, and NO_2, and the weather parameters temperature, rain, wind speed and UV index;
c. the time period of interest; in this case from 2012-07-28, T14:00 + 02:00 to 2012-07-29, T16:00 + 02:00.

The request is sent to the SOS DB. The DB returns the corresponding data for the specified region and time period and passes it to the Fusion Service. It should be noted that inside the DB there is a list of predefined regions described as polygons using coordinates, and the DB is able to match the areas that overlap with the query region in order to return the related observations.

4.3. Fusing and feeding the data into the ontology

To assess the relevance of the retrieved environmental data in the context of a concrete inquiry, to be able to reason its consequences for the user's concerns, and, finally, to deduce targeted decision support information, the data must be integrated into the DSO. Given that the data originate from different, partially competing (e.g., two different temperatures may be given for the same location) and partially complementary (e.g., one site may provide precipitation and another one temperature) sources, a data fusion procedure must precede the integration. Fusion ensures best quality data for a specific location at a specific time and allows for the extrapolation over small geographical and temporal gaps in the input data. The technical and formal details of the fusion procedure are described in (Johansson et al., 2015). Here, we present it briefly in general terms.

The main objective of fusion is to identify the representativeness of each biological, meteorological and chemical weather datum with respect to the user selected location and time. This representativeness needs to be quantified in mathematical terms. For this purpose, the following procedures are carried out:

1. A comprehensive historic measurement DB for each environmental phenomenon (e.g., pollutant substance) is compiled dynamically as data are provided.
2. The data provider's historic performance (e.g., the prediction accuracy of temperature forecasts or air quality sensor accuracy) is constantly monitored.

Using regression techniques on the compiled measurement DB it is possible to associate each datum with an expected variance measure. However, a datum describing recent, close-by conditions may be significantly biased by the characteristics of the location where it has been measured and the time when it has been measured. To detect and straighten out this bias, the fusion service takes two measures:

a. It uses a land-use mask and a population density mask to 'profile' locations. Within a selected range, the distribution of key land uses (urban, suburban, road, vegetation, etc.) and, if appropriate, population density are calculated;
b. It links the expected pollutant concentration, time of day, season and location profiles with multi-variable regression. The multivariable regression model is updated as the historic DB evolves.

[8] SOS is standardized solution for the sensor web layer. A number of other non-standardized approaches for building a Sensor Web exist as well; cf., e.g., the Global Sensor Network (GSN) (Aberer, Hauswirth, & Salehi, 2006).

[9] http://52north.org/

[10] PostgreSQL is used in SOS products as the DB management system due to its efficient handling of spatial data.

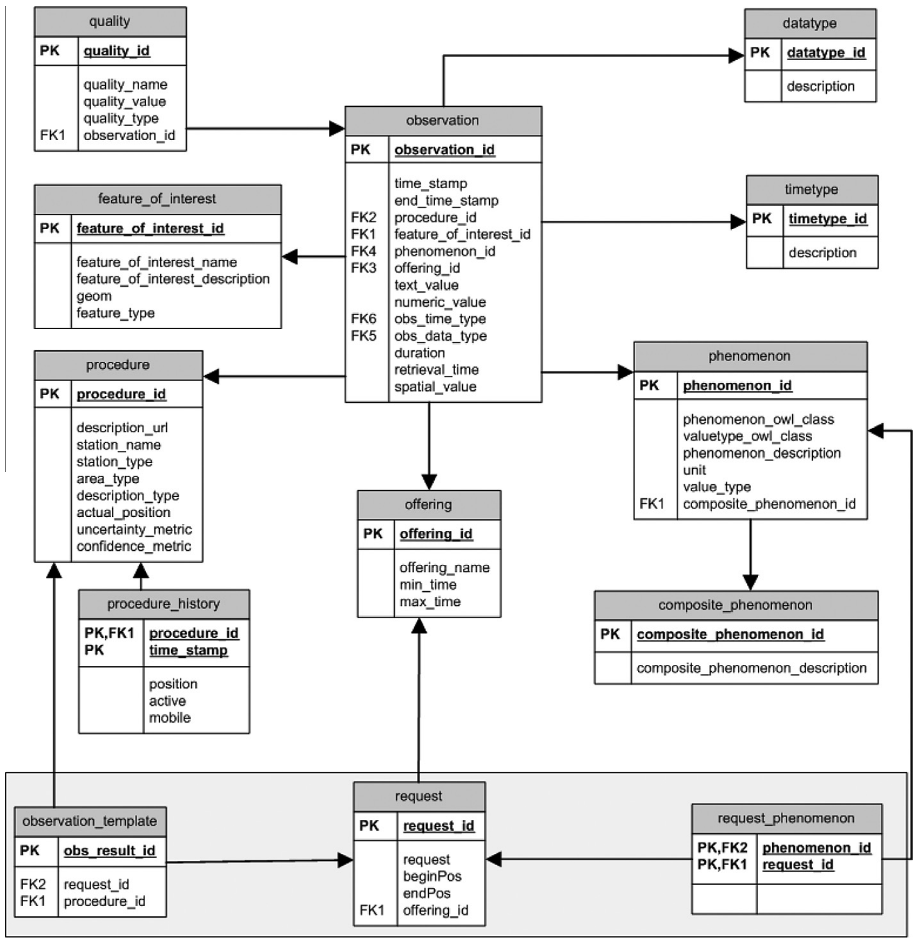

Fig. 9. DB schema of SOS semantically enriched implementation.

Finally, the variance measure is converted into weights by using an algorithm presented in (Potempski & Galmarini, 2009). Each datum d_i, is assigned after bias removal a weight w_i. A large estimated aggregate variance causes the assigned weight to decrease, while the data from more accurate and relevant sources cause the weight to increase – which makes these sources gain more emphasis during fusion. The weighted and bias-corrected d_is are then fused into d_F; consider Equation (1), where '$\mathbf{l}_0$' stands for location and 't_0' for the time of the fused datum, 'd_i'' for the datum considered at location '$\mathbf{l}_i$' and time 't_i':

$$d_F(\mathbf{l}_0, t_0) = \sum_{t=1}^{n} w_i d_i'(\mathbf{l_i}, t_i) \tag{1}$$

The fused data are fed into the A-Box of the KB as shown in Fig. 6 (middle panel – labeled "Data"); see in Fig. 6, for illustration, individuals such as *temperature_2d84e62e-c70e*, values of individuals (such as, e.g., *temperature_2d84e62e-c70e_value, 9.4 degreeC*), or assertions (such as, e.g., *hasValueValues, hasEnvironmentalDataEnvironmentalDataType*, and *hasFromDateTime*) that correspond to some of the data obtained for our running example.

5. Taking decisions/generating conclusions

The task of the decision support proper consists of three subtasks: (i) reasoning over the data, the problem description and the background knowledge to interpret the data and deduce the consequences of the meteorological/environmental conditions for the user; (ii) selection of the content to be included into the decision support bulletin; and (iii) generation of the bulletin.

5.1. Reasoning over the data and the problem description

Once the raw environmental data that are relevant to the given user request have been selected and inserted in the A-Box of the current request, they are processed in order to produce personalized content according to the request and profile of the user. The produced content comprises:

1. aggregations of raw data to favor the communication of environmental conditions to the users;
2. qualitative representation of numerical data (e.g., the fact that 9.4° Celsius are considered a cool temperature in summer and mild in winter);
3. detection of exceptional concentrations of air pollutants or pollen (as in the running example);
4. warnings or recommendations triggered by an environmental condition (e.g., the fact that if the concentration of birch pollen is abundant and the user is allergic to this pollen, a warning must be issued and suggestions on how to behave be proposed);
5. hints on possible causes of exceptional air quality episodes;
6. LSRs that connect facts described in the ontology (e.g., an 'atemporalImplication' relation between a fact of high pollen concentration and a health warning).

All of the above content is derived from the KB. For the computation of data aggregation and the detection of air pollutant concentrations above legally defined thresholds, some highly-modular ontology-based procedures have been implemented that first query the KB to retrieve the data to be considered in the computation, then apply the appropriate function to the retrieved data, detect pollutants whose concentrations are above thresholds, and finally store in the ontology the results of the processing.

For the computation of the qualitative representation of a numerical datum, fuzzy reasoning has been used.[11] Fuzzy reasoning is based on the theory of fuzzy sets, which mimics human reasoning in its use of imprecise information to generate decisions (Zadeh, 1975). An element belongs to a fuzzy set (a qualitative value such as 'abundant' (pollen concentration), 'hot' (temperature), 'low' (pressure), 'satisfactory' (air quality index), etc.) with a degree of membership, a real number from the [0, 1] interval. In formal terms, a fuzzy set is defined by a membership function that maps the universe of discourse to the [0, 1] interval. An example of the membership function for the fuzzy set 'moderate birch pollen concentration' is given in Fig. 10. The interpretation of this function is that when the birch pollen concentration is (more or less) between 10 and 100, it is *moderate*; close to the left border of this interval the "moderateness" is maximal. Below 10 it becomes between *low* and *moderate*, with *low* having a membership value greater than *moderate*. Moving from 10 to 100 the moderateness slowly decreases such that from value above 100 the *abundant* set has a membership value greater than the *moderate* one. The function in the given example is triangular, which is one of the most commonly used membership functions; other widely used functions are Gaussian ("bell") and trapezoid functions. An example of rating instantiated via fuzzy reasoning is the *temperature_2d84e62e-c70e_rating* individual in the "Data" panel of Fig. 6, corresponding to a "cool temperature" rating, computed for the *temperature_2d84e62e-c70e* data starting from the *temperature_2d84e62e-c70e_value* numerical value (9.4 C).

For the generation of warnings and recommendations, the suggestion of possible causes of exceptional air quality episodes, and the instantiation of LSRs between the facts in the A-Box of the current request, DL-reasoning has been combined with rule-based reasoning. For this purpose, a two-layer reasoning infrastructure has been implemented. The first layer exploits the HermiT reasoner[12] for the OWL DL reasoning services. The second layer is stacked on the top of the previous layer. It implements the Jena RETE rule engine,[13] which performs the rule-based reasoning computation. An example of a rule expressed using Jena RETE is that if the maximum value of birch pollen concentration in the period and location considered is abundant and the user is birch pollen-sensitive, then a recommendation has to be issued. An example of instantiated recommendation triggered by such rule is displayed in the "Conclusion" panel of the example in Fig. 6.

Both the DL-reasoning and rule-based reasoning techniques are applied directly to the A-Box of the current problem. The content they produce is put back into the same A-Box, stored as a set of instances and assertions on them. This enables us to relate the personalized content generated by the system with the data and the request description that were processed to produce it. Furthermore, it allows for the codification in a structured format the complete and connected story plan of the processed request: the initial request, the processed data, and the produced conclusions.

5.2. Selecting and structuring relevant content in the ontology

Once the interpretation of the data has been performed and the conclusions for the user in the light of her request have been drawn, the content that is relevant to the support of the user's decision is selected and structured for coherent presentation.

5.2.1. Content selection

The content selection strategy in the PESCaDO DSS is topic-oriented. For each topic (pollen, weather, pollutants, etc.), a "frame" of knowledge elements (more precisely, facts from the DSO) that provide information on this topic is specified. Each fact is assigned a rule (realized in terms of SPARQL queries) for its selection under specific conditions derived mainly from the user request. For instance, for pollen (as in our running example), the following facts are included into the frame:

+ types of pollen measured in region R
+ measured level of pollen X (with X as any type of pollen measured in R)
+ forecasted minimal level of X in R
+ forecasted maximal level of X in R
+ health warnings related to level of X
+ recommendations related to level of X

In the course of content selection, all defined topic frames are processed, such that the result of the content selection is a *content plan* consisting of the selected facts from each frame. Any given frame can be extended as needed to include more data provided by the reasoner (which infers additional facts, such as ratings tendencies, averages, and LSRs between them) and conditions on these data. For instance, although the user did not solicit supportive weather information, if she is allergic to pollen and the meteorological conditions (wind speed, precipitation, etc.) have been determined to be instrumental for the pollen conditions, the corresponding frame will be extended by meteorological condition elements; cf. Fig. 11 for the illustration of the pollen frame in a graphical format. Similarly, when the air quality index is reported on, the frame will be extended by the facts of the air pollutant substances that contribute to the index at the given time and location.

5.2.2. Determining the discourse structure of the selected content

The topic frames are very similar to *schemas* as first introduced into text generation by McKeown (1985) and as commonly used in robust state-of-the-art report generators (Portet et al., 2009; Wanner, Bohnet, Bouayad-Agha, Lareau, & Nicklass, 2010; Yu, Reiter, Hunter, & Mellish, 2007). However, while in a schema the composition of the elements of the schema and the order in which these elements are verbalized is fixed, this is not the case in topic frames. Topic frames need to accommodate for the great variety of content elements that could potentially be relevant to the inquiry of a user and for the dependence of the presentation on the needs and the profile of the user. Therefore, the order of the facts to be communicated to the user should be flexible and depend on the expertise of the user and the nature of her inquiry.

To ensure the required flexibility, PESCaDO's discourse structuring module consists of three stages: (i) elementary discourse unit (EDU) determination, (ii) mapping of the LSRs in the ontology to discourse relations; (iii) EDU ordering.

EDU determination groups thematically related elements into *propositional units*, starting from the frames determined during content selection. Fig. 11 shows two EDUs identified in the pollen-related frame, one for the pollen rating and another for the message with the recommendation concerning the actions of the user in view of this rating. As already content selection, EDU determination is performed by running SPARQL queries on the A-Box that describes the DS request under consideration.

LSRs introduced between facts of the ontology facilitate the derivation of discourse relations of the kind 'Circumstance', 'Elaboration', 'Evidence', etc. in the sense of the Rhetorical Structure Theory (Mann & Thompson, 1988) between the obtained EDUs, and thus also the derivation of a tree-like discourse structure,

[11] Fuzzy reasoning is based on Xfuzzy (http://www2.imse-cnm.csic.es/Xfuzzy/).
[12] Hermit Reasoner. http://hermit-reasoner.com
[13] Jena - A Semantic Web Framework for Java. http://jena.sourceforge.net/index.html.

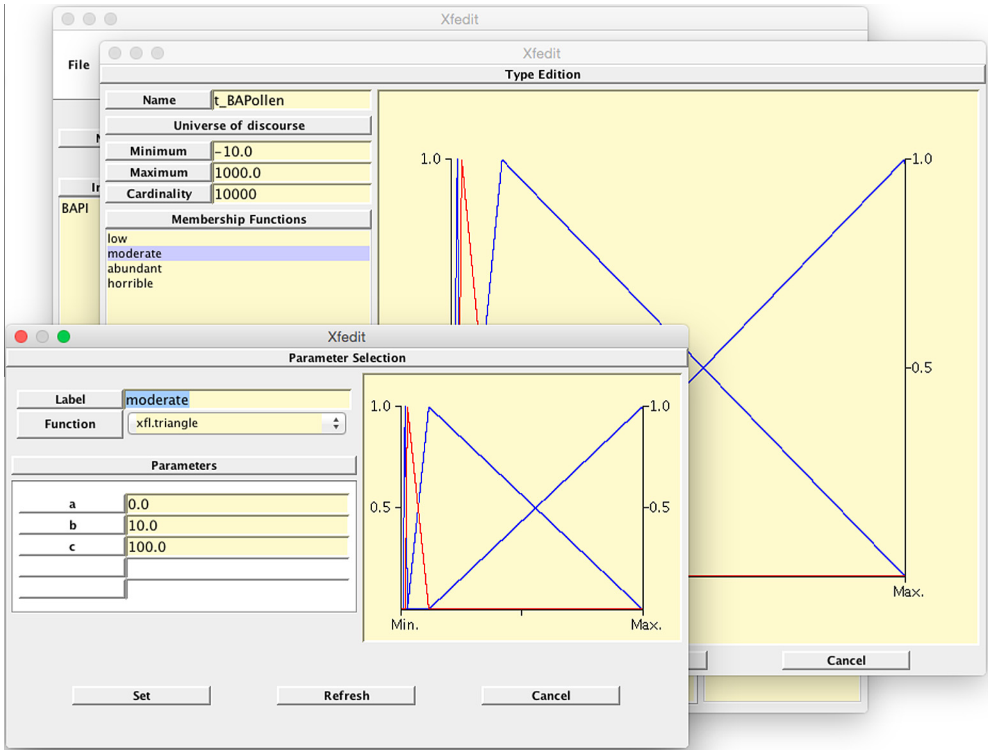

Fig. 10. Membership function for 'moderate birch pollen concentration'.

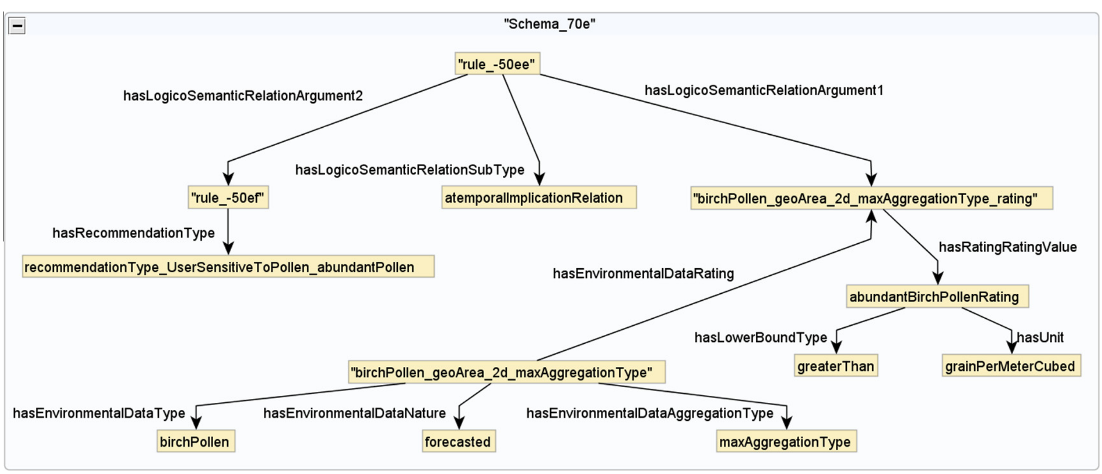

Fig. 11. Pollen frame for selection of content from the ontology for communication.

which is crucial for any coherent narration. Each LSR is mapped onto a specific discourse relation, depending on the context and the profile of the user. For instance, the 'atemporalImplication' LSR between the 'abundant concentration of birch pollen' element and the 'health warning' element will be mapped onto Consequence (which is of justificatory nature) if the user is a citizen and List (which is of notification nature) if the user is an environmental expert.

EDU ordering introduces precedence relations between obtained EDUs. Default order, as, e.g.:

EDU-pollen ← EDU-temperature ← EDU-wind speed ← EDU-wind direction ← EDU-skies ← EDU-humidity ← EDU-precipitation

and condition-based rules, as, e.g. the following one are used:

IF an EDU with the location and time of the user query is available
THEN place this EDU first

The output of the discourse structure module is a "text plan" that is used as input by the linguistic generator; cf. Fig. 12 for the text plan of the frame in Fig. 11, which contains the content for the system's answer to the user request in our running example.

5.3. Producing decision support information from the ontology content

The linguistic generation module produces from the text plan decision support information. The main mode of this information is text in the language of the preference of the user.[14] In the PESCaDO setup, English, Finnish, and Swedish are supported.

[14] The textual information is further enriched by graphical information; cf. (Wanner et al., 2014).

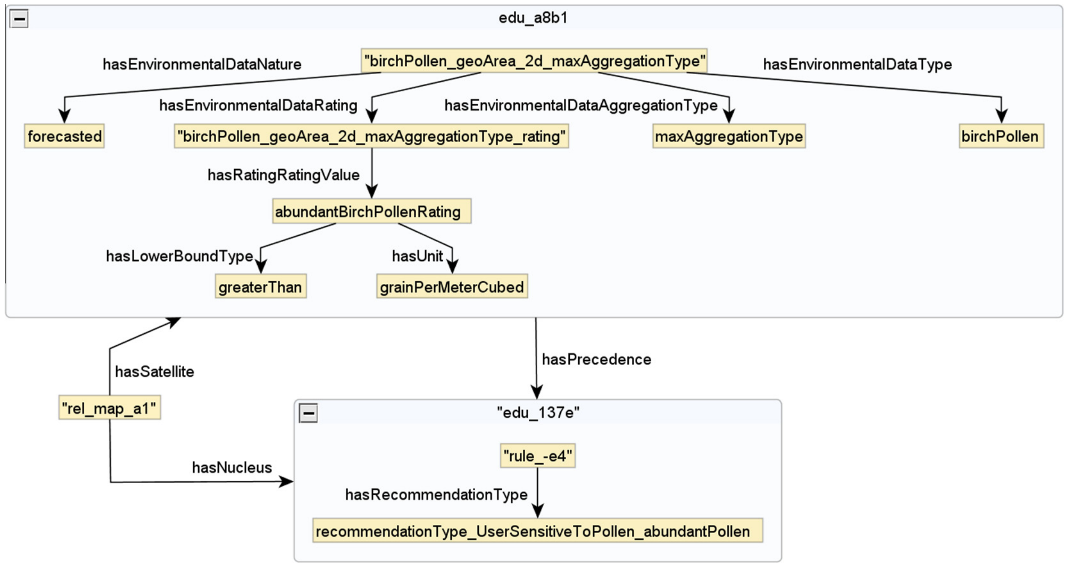

Fig. 12. Text plan for decision support oriented information.

The text generator is based on the extended linguistic model of the Meaning-Text Theory (MTT) (Mel'čuk, 1988). The MTT model foresees different levels of representation for each type of linguistic description (semantics, syntax, morphology), such that the process of the conversion of a given text plan into environmental information is a cascade of transitions between structures of adjacent levels; see, e.g., (Lareau & Wanner, 2007) for theoretical details and (Wanner et al., 2010) for the use of an analogous generator for generation of air quality information. Fig. 13 shows a sample report generated by the PESCaDO DSS in the context of our running example.

6. Evaluation

In what follows, the evaluation of the major decision support modules of the proposed environmental expert system as well as of the system as a whole is presented.

6.1. Evaluation of the data acquisition

In the context of data acquisition, we evaluate data collection and data fusion.

6.1.1. Data collection

The evaluation of data collection considers the following metrics: (i) accuracy of the discovered environmental websites, (ii) accuracy of the extracted data and (iii) time response of the data retrieval module.

Table 1 presents an overview of the outcome of environmental site discovery for South Finland using BOSS Yahoo! API and post-processing classification. Examples of queries submitted to BOSS Yahoo! API are "pollen Nukksio-Park", "temperature Nukksio-Park", "weather + Helsinki", "wind + Tuusula", etc. The performance is measured in terms of well-established information retrieval metrics such as precision (defined as the fraction of retrieved sites that are relevant to the query to the total of the retrieved sites) and recall (defined as the fraction of retrieved relevant sites to the total of relevant sites). The post-processing classification improves the results returned by Yahoo! by 38.5% (from 0.286 to 0.671).

The content extraction from textual websites is performed using dedicated parsers that have been adapted to the formats of the prominent environmental web pages of the regions covered by the system, with an extraction accuracy of 100%. The adaptation turned out to be necessary given the very high level of idiosyncrasy of the codification of information in environmental web pages.[15] On the other hand, the environmental data extraction from heatmap images cannot be flawless since an automatic OCR procedure is employed to detect the position of values on the heatmap axes. The evaluation was realized on a set of 60 images extracted from: Pollen FMI site,[16] SILAM model FMI site,[17] GEMS site,[18] Laboratory of Atmospheric Physics of the Aristotle University of Thessaloniki (LAP) site[19] and the Atmospheric and Oceanic Physics Group site[20]. The error er is calculated as:

$$er = \frac{1}{n} \cdot \sum_{i=0}^{n} \left| \frac{v_i - eV_i}{v_i} \right|, \tag{2}$$

where n is the total number of pixels, v_i is the value of pixel i using manual configuration, and ev_i is the value estimated by the system. The results show that the error introduced in each pixel value is in general low (around 6%).

In order to test and validate the indexing and retrieval module, we have populated the repository using real environmental data from different sources[21] from the geographical area of South Finland.

Table 2 contains the total number of records used to cover either the static data (*offering*, *feature_of_interest*, *phenomenon*) or inserted data (procedure, observation). The time response of the data retrieval module when accessing this data was less than 1 s.

6.1.2. Data fusion

For evaluation of data fusion, temperature was selected as sample variable. Forecast input data were provided by four known

---

15 Experiments with generic data extraction from any arbitrary environmental web page worldwide showed that the formats of the pages are too ad hoc and too irregular, such that no parser can be expected to guarantee output of reasonable quality.

16 http://pollen.fmi.fi/pics/Europe_ECMWF_olive.html

17 http://silam.fmi.fi/AQ_forecasts/Europe_v4_8/index.html

18 http://gems.ecmwf.int/d/products/raq/

19 http://lap.physics.auth.gr/forecasting/fore_images/

20 http://www.fisica.unige.it/atmosfera/bolchem/MAPS/

21 wms.fmi.fi, intermeteo.com, timeanddate.com, ilmanlaatu.fi, gems.ecmwf.int, weatherforecastmap.com, en.ilmatieteenlaitos.fi, http://abclocal.go.com, helsinki-weather.info, ilmanlaatu.fi.

Forecast in the selected area between 28/07/2012 (14h00) and 29/07/2012 (16h00).
The birch pollen count will be abundant. The temperature will be 9.4ºC, the wind weak and the weather will vary from dry to moderate rain. There is no data available for UV.

***Pollen warning**: Most of the sensitive people have symptoms.*

Fig. 13. A sample of decision support information generated by the PESCaDO DSS.

Table 1
Evaluation results of the hybrid approach.

Webpages retrieved by Yahoo!	Precision of web search	Classification precision	Classification recall
2380	0.286	0.671	0.789

Table 2
Records in the repository.

Tables	Records
Offering	3
Feature_of_interest	360
Observation	17941
Procedure	119
Phenomenon	10

service providers (SP) for 50 different locations in Finland in August 2011 for which measured temperatures were readily available. A total of 2500 forecasted-measured temperature pairs from each SP were gathered.

Fig. 14 shows the mean absolute error of temperature forecasts and the fused forecast. The fused temperature forecast has already the lowest mean error with just four different SPs, which is a very good outcome for the fusion procedure.

The performance of the fusion was further tested against a conventional inverse distance weighting extrapolation method (IWD). For this study, the annual measured NO_2 concentration series for 12 local stations around Helsinki metropolitan area in 2011 were used. For each location, the measurements of the corresponding station were removed from the input data and then the NO_2 concentration time series for the whole year was estimated with the fusion and IWD methods. The resulting mean absolute errors of predictions for the 12 test sites are shown in Fig. 15 for both methods. The figure shows that the fusion method has a clearly lower mean absolute error. The largest differences in terms of absolute mean prediction errors occur at the test sites of Mäkelänkatu and Luukki. The first one represents a highly busy urban traffic node, while the Luukki station is a remote background station. This result shows the importance of environmental profiling for the fusion of urban air quality measurements.

In Fig. 16, a sample of estimated NO_2 concentration (fused with profiling) is shown side by side with actual measurements. The location is situated in a suburban background area in Vantaa. The fused values were estimated without the actual measurement data, using the remaining measurement stations as input data nodes. Fig. 16 shows the high accuracy of the PESCaDO fusion method.

6.2. Evaluation of the derivation of decision support content

In the context of ontology interpretation techniques, two questionnaire-based evaluation runs have been carried out—one to assess the quality of the determination of the relevant environmental data for a given request submitted to the system, and one to assess the performance of ontology-based techniques for deduction of relevant personalized content. Both evaluations involved a team of four environmental information experts for three different user problems: (i) planning of an excursion by a citizen with no background in environmental domain, (ii) monitoring of air pollution by an administrative staff member, and (iii) setting up an environmental service by a professional.

In the first evaluation run, the evaluators were asked to judge the appropriateness and the completeness of the environmental data determined as relevant by the system. For this purpose, the experts were presented, for each problem, the textual description of the problem and the list of environmental data determined as relevant to this problem by the system, with the demand to judge the appropriateness of these data in the context of the problem. The outcome of the evaluation revealed an average appropriateness of 94% (with a standard deviation of 11%) and an average completeness of 92% (with a standard deviation of 8%). The appropriateness and completeness figures of all experts and all problems are most often distributed between 92% and 100% for both appropriateness and completeness. The median value is 100% for appropriateness, and 92% for completeness.

In the second evaluation run, the evaluators were asked to judge the appropriateness (e.g., whether it is appropriate to call −26.3 °C "extremely cold") and completeness (e.g., whether the health warnings are included in the case of a high concentration of birch pollen) of the information produced by the system. In this case, the experts were presented for each problem the textual description of the problem, a spreadsheet containing the relevant environmental data retrieved by the system, and the environmental information delivered by the system. The obtained results show an average appropriateness of 90% (with a standard deviation of 25%) and an average completeness of 87% (with a standard deviation of 23%) of the personalized content produced by the system. With the exception of a few outliers, the appropriateness and values obtained for all experts, all problems, and all questions, are as a rule distributed between 95% and 100%, while the completeness figures are mostly distributed between 75% and 100%. The median value is 100% in both cases.

6.3. Evaluation of the decision support information production techniques

To assess the quality of the information production techniques, two different evaluations have been carried out – one on content selection and one on discourse structuring/linguistic generation.

For the evaluation of content selection (CS), six system queries were executed and the resulting ontology content dumped into a file.[22] For each dump, all possible instances of the content frames used by the CS module were identified and listed in a readable format. Two environmental experts then marked the relevance of the instances for the problem description they were given: the profile of the user, the request, the geographic area of the query, and the dates. The agreement between the experts was 0.71 in terms of Cohen's Kappa coefficient. The answers of the experts constituted the gold standard content plans against which other plans were

[22] The queries were chosen to yield representative content in the ontology prior to the execution of CS. Three queries correspond to an administrative user scenario, and three to a layperson user scenario.

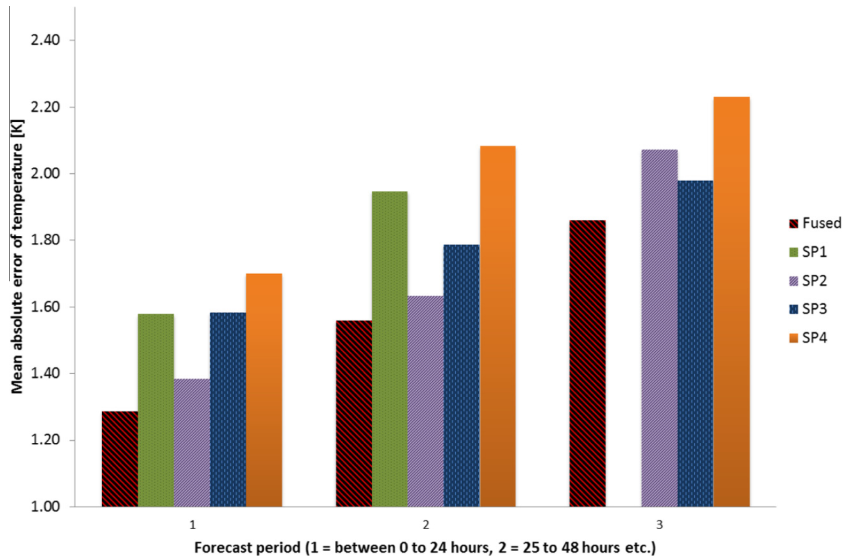

Fig. 14. Mean absolute error of temperature forecasts and the fused forecast for different forecast time spans.

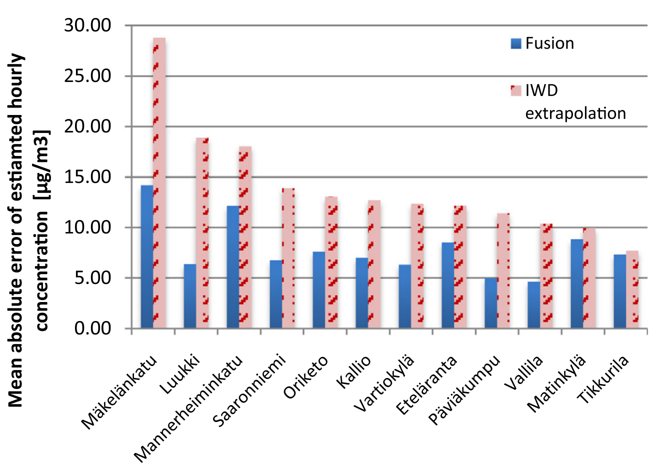

Fig. 15. Comparison of the prediction accuracy of fused concentrations with the accuracy of the IWD extrapolation method.

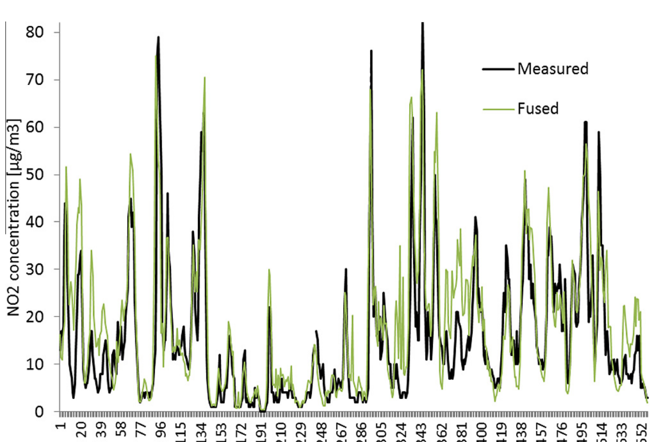

Fig. 16. Estimated fused NO_2 concentration based on local measurement stations (Vantaa and Päiväkumpu) versus the observed concentration in March 2011.

Table 3
Content selection evaluation with two annotators ('*r*' = recall, '*p*' = precision, '*f*' = F-score, '*ir*' = inverse recall, '*ip*' = inverse precision, '*if*' = inverse F-score, '*af*' = average F-score).

	r	*p*	*f*	*ir*	*ip*	*if*	*af*
Baseline	0.94	0.70	0.80	0.56	0.89	0.69	0.74
CS	0.89	0.88	0.88	0.87	0.88	0.87	0.88

evaluated. As baseline content selection, the "majority class" vote was used, which picks a frame if more than half of the instances of this frame are also selected by the experts in the gold standard. The six queries were executed by both the baseline and the CS module. The results were compared with the gold standard to obtain precision, recall and their harmonic mean (F-score) for both the baseline and the CS module. Based on the counts of uninstantiated frames, we also calculated the inverse precision, inverse recall and inverse F-score, which in turn reflect how good the baseline and the CS module are at discarding the content also discarded by the experts in the gold standard. Table 3 summarizes the obtained figures. The figures indicate that the baseline is quite good at selecting the same content as the experts, the average of the F-score and its inverse being 0.74, yet that the CS module still improves this score up to 0.88.

For the evaluation of the information production, the quality of the environmental textual information was evaluated by environmental domain experts according to the following four criteria at a 5-point scale, with 5 being the best score and 1 the worst:

- information packaging: whether information is aggregated/packaged in an optimal way in clauses and sentences,
- intelligibility: whether the text is grammatical and the wording appropriate,
- ordering: whether the order of the clauses is optimal, and
- accuracy: whether all the detailed information available in the content plan is verbalized.

The evaluation was carried out with six content plans with enough variety generated by the CS module, three for planning an excursion by a citizen with no background in environmental domain, and three for monitoring of air pollution by an administrative staff member. Two domain experts produced from these

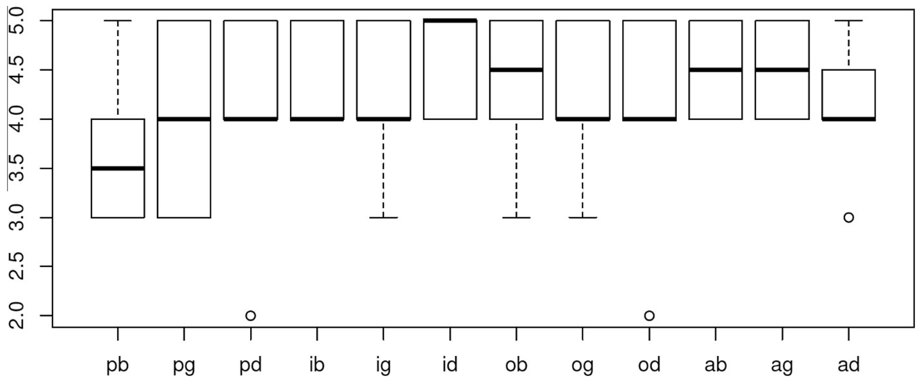

Fig. 17. Box plot of the distribution of packaging, intelligibility, ordering and accuracy values for baseline, generated and gold texts for English ('*p*' = packaging, '*i*' = intelligibility, '*o*' = ordering, '*a*' = accuracy; '*b*' = baseline, '*g*' = generated, '*d*' = gold, such that '*pb*' stands for "packaging baseline", '*pg*' for "packaging gold", etc.).

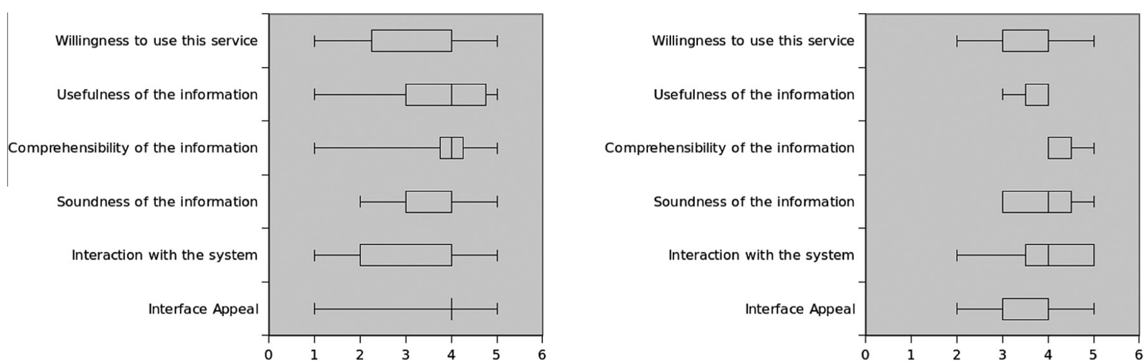

Fig. 18. Evaluation of the proposed DSS by citizens (left) and environmental experts, when using the DSS for administrative decision support (right).

content plans environmental reports, which constituted the gold standards. In parallel, from these content plans, texts were generated by the linguistic generation module. The baseline was obtained by applying a template-based generator which uses sentence templates. The content for the templates was retrieved from the content plans together with ontological information. Fig. 17 displays the results for English (for Finnish and Swedish, the outcome is comparable). It shows that the quality of the PESCaDO text generator across the criteria is rated above 4 out of 5.

6.4. Evaluation of the system as a whole

The evaluation of our complete environmental DSS has been carried out by groups of citizens and experts. In total, 18 citizens and 7 environmental information experts participated in the evaluation. Both groups were asked to fill out questionnaires, rating their consent to a number of statements on a scale from 1 (strongly disagree) to 5 (strongly agree)–including "The interface looks appealing", "The interaction with the system is straightforward", "The provided information is comprehensible", etc. The box plots in Fig. 18 show the dispersion of the ratings. In boxes that do not show the median marker, except in that for "interface appeal" of experts, the median coincides with the right margin of the box. In the box for "interface appeal" of experts, it coincides with the left margin of the box.

With respect to the appeal of the interface, there is *de facto* no dispersion among citizens, while among the experts, the lower quartile is at 3 (which means "undecided about the appeal of the interface" and the upper quartile at 4 ("I agree that the interface is appealing"). Unsurprisingly, the experts find the interaction with our system easier than the citizens. What might be more surprising is that the soundness of the information is judged in the average higher by experts than by citizens. One might have expected that environmental experts will be more critical with automatically generated information than naïve users. The average 'comprehensibility of information' grade by citizens is 3.88 and by experts 4.3, i.e., in both cases rather high; the dispersion among the ratings is in both cases limited. As the upper quartile of the box plot for the usefulness of the information among citizens shows, citizens tended to find the information provided by the system even somewhat more useful than the experts. The willingness to use the provided service is among citizens slightly lower (3.6 in the average) than among experts (3.9 in the average), but in both cases it is high for an experimental implementation.

7. Conclusions and future work

In the light of the great number of environmental web pages available for the same geographical region and the highly diverging (or even contradicting) data they offer, personalized decision support is needed to remove from the user the burden to guess the quality of the data offered by the individual providers and to interpret them in the context in which a decision is to be taken.

An environmental expert system that is designed to provide personalized decision support to a variety of different users benefits greatly if all types of content it has to deal with are represented in terms of ontologies: (i) measured and observed data from the web as well as assessed and fused data for a given location and time; (ii) background environmental knowledge, (iii) profiles of the users; (iv) the formal language with which a user can solicit decision support; (v) the concerns of the user for which the system can offer decision support, and (vi) discourse knowledge to verbalize the decision support in terms of a coherent and linguistically well-formulated report. The ontological representation facilitates high quality reasoning, content selection and verbalization. Furthermore, allows for a smooth extension/adaptation to new decision support requests by new types of users.

The evaluation of the entire processing chain of the proposed environmental DSS demonstrated both the quality of the individual modules and the capability of the system to compete with the decision support offered by experts.[23] However, a number of challenges must be solved to make our DSS a mature service – including, for instance:

(i) to be able to reliably distil environmental data from any unseen web page, independently of their format (recall that so far we focus on previously inspected prominent pages);
(ii) to be able to dynamically extend the ontologies in order to facilitate the acquisition of additional background environmental, medical and social content that might be relevant to the provision of personalized environmental information, without cost-intensive manual labor;
(iii) to be able to cope with unexpected user decision support requests;
(iv) to ensure user privacy and personal data security: as a prototypical service, the described DSS does not take any measures to ensure that sensitive user data such as domicile, life style preferences, allergies, etc. are handled in compliance with legal data protection regulations;
(v) to facilitate the expansion of the DSS to other related areas (such as water quality, road conditions, etc.).
(vi) These challenges should be on the agenda of any future work in the field of environmental DSSs.

Acknowledgements

The work reported in this paper was supported by the European Commission under the contract FP7-ICT-248594 (PESCaDO).

References

Aberer, K., Hauswirth, M., & Salehi, A. (2006). A middleware for fast and flexible sensor network deployment. In: *Proceedings of 32nd international conference on very large data bases* (pp. 1199–1202), Seoul, Korea.

Benincasa, G., D'Aniello, G., De Maio, C., Loia, V., & Orciuoli, F. (2015b). Towards perception-oriented situation awareness systems. *Intelligent Systems, 322*, 813–824.

Benincasa, G., D'Aniello, G., Gaeta, M., Loia, V., & Orciuoli, F. (2015a). Resilient semantic sensor middleware. *Intelligent Distributed Computing, VIII, 570*, 453–463.

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., & Lemmens, R. (2011). New generation sensor web enablement. *Sensors, 11*(3), 2652–2699.

Ceccaroni, L., Cortés, U., & Sànchez-Marrè, M. (2004). OntoWEDSS: Augmenting environmental decision-support systems with ontologies. *Environmental Modelling and Software, 19*(9), 785–797.

Choras, M., Kozik, R., Flizikowski, A., Renk, R., & Holubowicz, W. (2009). Ontology-based decision support for security management in heterogeneous networks. *Proceedings of the 5th international conference on emerging intelligent computing technology and applications* (pp. 920–927).

Esposito, M., & De Pietro, G. (2011). An ontology-based fuzzy decision support system for multiple sclerosis. *Engineering Applications in Artificial Intelligence, 24*(8), 1340–1354.

Ishizu, S., Gehrmann, A., Minegishi, J., & Nagai, Y. (2008). Ontology-driven decision support systems for management system audit. *Proceedings of the 52nd annual meeting of the ISSS*. .

Johansson, L., Epitropou, V., Karatzas, K., Karppinen, A., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., & Kompatsiaris, I. (2015). Fusion of meteorological and air quality data extracted from the web for personalized environmental information services. *Environmental Modelling and Software, 64*, 143–155.

Lareau, F., & Wanner, L. (2007). Towards a generic multilingual dependency grammar for text generation. In T. King & E. Bender (Eds.). *Proceedings of the GEAF07 workshop*. .

Laskey, K. B. (2006). *Decision Making and Decision Support*. <http://seor.gmu.edu/~klaskey/ SYST542/DSS_Unit1.pdf>.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text, 8*(3), 243–281.

Marakas, G. M. (1999). *Decision support systems in the twenty-first century*. Prentice-Hall.

McKeown, K. R. (1985). *Text generation: Using discourse strategies and focus constraints to generate natural language text*. Cambridge: Cambridge University Press.

Mel'čuk, I. (1988). *Dependency syntax: Theory and practice*. SUNY Press.

Moßgraber, J., & Rospocher, M. (2012). Ontology management in a service-oriented architecture. *Proceedings of the 11th international workshop on web semantics and information processing (WebS2012)*. .

Moumtzidou, A., Epitropou, V., Vrochidis, S., Karatzas, K., Voth, S., Bassoukos, A., Moßgraber, J., Karppinen, A., Kukkonen, J., & Kompatsiaris, I. (2014). A model for environmental data extraction from multimedia and its evaluation against various chemical weather forecasting datasets. *Journal of Ecological Informatics, 23*, 69–82.

Moumtzidou, A., Vrochidis, S., Chatzilari, E., & Kompatsiaris, I. (2013). Discovery of environmental resources based on heatmap recognition. *Proceedings of the IEEE international conference on image processing (ICIP 2013)*. .

Moumtzidou, A., Vrochidis, S., Tonelli, S., Kompatsiaris, I., & Pianta, E. (2012). Discovery of environmental nodes in the web. *Proceedings of the 5th IRF conference*. .

OGC (2011). SOS 2.0 tutorial. Retrieved from <http://www.ogcnetwork.net/SOS_2_0/tutorial>.

Pangjitt, T., & Sunetnanta, T. (2011). A model of ontology driven case-based reasoning for electronic issue management systems. *Proceedings of the international joint conference on computer science and software engineering*. .

Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., & Sykes, C. (2009). Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence, 173*(7–8), 789–916.

Potempski, S., & Galmarini, S. (2009). Est modus in rebus: Analytical properties of multi-model ensembles. *Atmospheric Chemical Physics, 9*, 9471–9489.

Rospocher, M. (2014). An ontology for personalized environmental decision support, FOIS2014. *Formal ontology in information systems – proceedings of the eighth international conference* (267). Rio de Janeiro: IOS Press. September, 22–25, 2014.

Rospocher, M., & Serafini, L. (2012). An ontological framework for decision support. *Proceedings of the 2nd joint international semantic technology conference*. .

Saremi, A., Esmaeili, M., Habibi, J., & Ghaffari, A. (2008). O2dss: A framework for ontology-based decision support systems in pervasive computing environment. In: *Proceedings of the 2008 second asia international conference on modeling & simulation (AMS)*.

Tonelli, S., Rospocher, M., Pianta, E., & Serafini, L. (2011). Boosting collaborative ontology building with key-concept extraction. In: *Proceedings of the IEEE fifth international conference on semantic computing (ICSC-2011)*.

Vrochidis, S., Bosch, H., Moumtzidou, A., Heimerl, F., Ertl, T., & Kompatsiaris, I. (2012). An environmental search engine based on interactive visual classification. *Proceedings of the 1st ACM international workshop on multimedia analysis for ecological data*. .

Wanner, L., Bohnet, B., Bouayad-Agha, N., Lareau, F., & Nicklass, D. (2010). MARQUIS: Generation of user tailored multilingual air quality bulletins. *Applied Artificial Intelligence, 24*(10), 914–952.

Wanner, L., Bosch, H., Bouayad-Agha, N., Casamayor, G., Ertl, T., Hilbring, D., Johansson, L., Karatzas, K., Karppinen, A., Kompatsiaris, I., Koskentalo, T., Mille, S., Mossgraber, J., Moumtzidou, A., Myllynen, M., Pianta, E., Rospocher, M., Serafini, L., Tarvainen, V., Tonelli, S., & Vrochidis, S. (2014). Getting the environmental information across: From the web to the user. *Expert Systems*. http://dx.doi.org/10.1111/exsy.12100.

Wyner, A. (2008). An ontology in OWL for legal case-based reasoning. *Artificial Intelligence and Law, 16*(4), 361–387.

Yu, J., Reiter, E., Hunter, J., & Mellish, C. (2007). Choosing the content of textual summaries of large time-series data sets. *Natural Language Engineering, 13*(1), 25–49.

Zadeh, L. (1975). Fuzzy logic and approximate reasoning. *Synthese, 30*, 407–428.

Zagorulko, Y., & Zagorulko, G. (2010). Ontology-based approach to development of the decision support system for oil-and-gas production enterprise. In: *Proceedings of the 2010 conference on new trends in software methodologies, tools and techniques*.

[23] Detailed information on the PESCaDO system and a demonstration of it are available at http://pescado-project.upf.edu/. Currently, PESCaDO's geographical coverage is restricted to Finland.