A Semantic Framework to Enrich Collaborative Tables with Domain Knowledge *

Anna Goy, Diego Magro, Giovanna Petrone, Marco Rovera and Marino Segnan
Dipartimento di Informatica, Università di Torino, C. Svizzera 185, 10149, Torino, Italy
{annamaria.goy, diego.magro, giovanna.petrone, marino.segnan}@unito.it, europaway@gmail.com

Keywords: Communication, Collaboration and Information Sharing, Metadata Management, Linked Open Data, Semantic Web, Ontology-driven Applications.

Abstract: In this paper we present a project aimed at enhancing a collaborative environment for resource management (SemT++) with domain knowledge, represented by a local ontology and a connection to external data, retrieved from Linked Open Data sets. Our approach is based on the assumption that heterogeneous resources can be viewed as "information objects", and can be organized within collaborative spaces (i.e., "round tables"). Information objects, among other properties, are characterized by their content. Annotations representing resource content (e.g., "Torino") can thus be linked to domain knowledge which provides users with useful information. We tested this approach on the geographic domain, by connecting resources to commonsense geographic knowledge and to information available in GeoNames.

1 INTRODUCTION

In the current ICT scenario, Human-Computer Interaction (HCI) and Personal Information Management (PIM) (Barreau and Nardi, 1995) have to face new challenges. In particular, new web architectures and paradigms, such as Web 2.0, Cloud Computing, Software-as-a-Service, are posing new problems and offering new opportunities. Two aspects are particularly relevant from our viewpoint: first, users have to face the management of a huge amount of heterogeneous resources, possibly related to the same content, but encoded in different formats, handled by different applications, stored in different places, and belonging to different types (documents, emails, videos, bookmarks,...); second, users can actively participate in content creation, can share resources and knowledge, and can collaborate with each other in carrying on many activities.

One of the possible approaches to effectively support both heterogeneous resource management and collaboration relies on semantic technologies, which can be exploited to provide users with smarter and more friendly tools for managing shared resources on the web. This idea is not new. In particular, a significative trend in this direction is the emerging *Social Semantic Web* (Breslin et al., 2009), which relies on the idea that semantic technologies can support the creation of machine readable interlinked representations of social objects (people, contents, resources, tags, etc.) enabling different social "islands" (i.e., isolated communities of users and data) to be connected and integrated. The approach presented in this paper can be seen as part of this project, since it aims at enhancing a collaborative environment for resource management with semantics, in order to provide users with a smarter support to resource management.

Our approach, in particular, is based on the hypothesis that digital resources should be viewed as *information objects*, and should be managed in a uniform way, independently from their possibly heterogeneous types. Awareness about information objects includes different aspects, such as knowledge about the format they are encoded in (e.g., PDF, HTML, JPEG, etc.), about their structure (e.g., if a document contains images or hyperlinks), and about their content (e.g., what a document "is about", or what an image represents). This kind of knowledge has been encoded within *Semantic Table Plus Plus* (SemT++), an environment aimed at

*POSTPRINT VERSION. Cite as: Goy, A., Magro, D., Petrone, G., Rovera, M. and Segnan, M.. A Semantic Framework to Enrich Collaborative Tables with Domain Knowledge. In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 3: KMIS, pages 371-381 - ISBN: 978-989-758-158-8.

supporting users in collaborative resource management on the web.

In this paper, we describe an enhancement of SemT++, leading to DSemT++ (*Domain-aware SemT++*). Besides general knowledge about *information objects*, i.e., information resources as such, DSemT++ relies on knowledge about their *content*. The first type of knowledge is "universal", in the sense that it is (to a certain extent) domain independent, i.e., it models digital resources independently of their specific content (for example, a digital resource always is encoded in a given format, is expressed in one or more languages, and so on); knowledge about resource content is usually domain-specific, since resource content can refer to very different knowledge domains: if a document talks about European Medieval history, the semantic knowledge enabling a tool to deal with the resource content (e.g., for retrieving it) must include a semantic representation (e.g., a Domain Ontology) modeling concepts belonging to the European Medieval history.

A detailed account of the representation of knowledge about *information objects* in SemT++ can be found in (Goy et al., 2014a). In this paper, we will focus on the second type of knowledge, and we will show how a Domain Ontology, coupled with existing resources available as Linked Open Data sets, can be exploited to support users in the organization, retrieval and usage of shared digital resources. The architecture aimed at including domain knowledge in a resource management collaborative environment, together with the support provided by this kind of knowledge, actually represent the major contribution reported in this paper.

The rest of the paper is organized as follows. In Section 2 we set the background, by discussing the main related work, and in Section 3 we briefly summarize the SemT++ project, as it is described in previous papers (mentioned below). In Section 4, which contains the novel contribution with respect to our earlier work and represents the core of this paper, we describe DSemT++, i.e., the enhancement of SemT++ with domain knowledge; moreover, we explain why we chose commonsense geographic knowledge as a testbed domain, we sketch a simple usage scenario, we describe how domain knowledge is linked to knowledge in the Linked Open Data cloud, and how the resulting system supports users in collaboratively handling semantic descriptions of digital resources. Section 5 concludes the paper by discussing open issues and future developments.

2 RELATED WORK

As far as the aspects related to HCI and PIM are concerned, one of the most relevant research areas is well accounted for by Kaptelinin and Czerwinski (2007), which contains a wide presentation of the problems of the so-called *desktop metaphor*, and of the approaches trying to replace it. In particular, one of the most interesting models discussed in this book is Haystack (Karger, 2007), a flexible and personalized system enabling users to define and manage workspaces referred to specific tasks. Another interesting family of approaches are those grounded into Activity-Based Computing – e.g., (Bardram, 2007; Voida et al., 2008) – where the interaction is designed around the concept of *user activity*. The main "step forward" of DSemT++ with respect to these approaches is the explicit domain knowledge model and the exploitation of Linked Open Data sets, as explained in Section 4.

Strategies used to organize resources have been studied also in social tagging systems, where resources can be tagged with meta-data representing different aspects (*facets*), leading to the creation of *folksonomies*, i.e., multi-facets classifications collaboratively and incrementally built by users in a bottom-up perspective (Breslin et al., 2009). Interesting improvements of such tagging systems have been developed by endowing them with semantic capabilities – e.g., (Abel et al., 2010) – in particular in the perspective of knowledge workers (Kim et al., 2009). With respect to these systems, DSemT++ has a slightly different focus, since it supports collaboration within (small) groups of people working together, instead of mass social communities.

Interesting approaches, based on the definition of a common conceptual framework provided by computational ontologies, have been developed within the Knowledge Management area, with the aim of facilitating communication and shared understanding in collaborative decision-making environments; see, for example, (Evangelou and Karacapilidis, 2005).

Another important research thread, aiming at coupling desktop-based user interfaces and Semantic Web, is represented by the *Semantic Desktop* approach (Sauermann et al., 2005). In particular, the NEPOMUK project (nepomuk.semanticdesktop.org) defined an open-source framework, based on a set of ontologies, for implementing semantic desktops, focusing on the integration of existing applications, in order to support collaboration among knowledge workers. (Drăgan et al., 2009) presents an interesting approach connecting the Semantic Desktop to the Web of Data, underlying how "connecting the two networks of information opens up the possibility of personal services on the desktop which use external data, but in the personal context of the user, highly connected to his personal data and focused on his interests" (Drăgan et al., 2009, p. 34). Moreover, "connecting desktop data with the web enables the system to bring web data to the user, instead of the user having to go find it by himself" (Drăgan et al., 2009, p. 35).

This last proposal is one of a large number of semantic approaches which recently have tried to exploit the potentiality of the Linked Open Data (LOD) paradigm, relying on the fact that most datasets refer to one or more ontologies, or "semantic" vocabularies (e.g., DBpedia: dbpedia.org, GeoNames: www.geonames.org). From this point of view, DSemT++ belongs to the same research thread.

In the same direction, an interesting project, which shares many features with our approach, is LinkZoo (Meimaris et al., 2014) a collaborative platform which exploits LOD to annotate shared heterogeneous resources. Semantic descriptions of resources are stored as RDF triples, and they enable LinkZoo to couple standard keyword search with property-based filtering. (Schandl et al., 2012) contains a survey of the approaches to exploit LOD in metadata for multimedia content, and CAMO (Hu et al., 2014) represents an example of linking LOD to multimedia metadata. Linkify (Yamada et al., 2014) is an add-on for major browsers which adds a link to Named Entities recognized in online texts, pointing to a mashup of information items extracted from LOD sources. MOAT (Meaning Of A Tag) is a framework providing a semantic model for defining machine-readable meanings of tags (Passant and Laublet, 2008). MOAT models tags as quadruples (<*User, Resource, Tag, Meaning*>) and provides a MOAT server, which can be exploited to share tag meanings and retrieve them when tagging resources; in particular, when a user tags content, the MOAT client retrieves tag meanings from the server and let the user choose the most relevant one. Tag meanings are linked to URIs of entities within well-known LOD datasets, such as DBpedia and GeoNames: this solves tagging ambiguity (i.e., in case a tag has more than one URI) and heterogeneity (i.e., in case different tags refer to the same URI), and enables the suggestion of relevant content derived from LOD.

Finally, an example of a different exploitation of LOD can be found in (Giunchiglia et al., 2012), where the authors present a *geospatial* ontology, called *Space*, based on GeoNames, WordNet and MultiWordNet, together with the methodology used for its creation. *Space* is aimed at representing geographic and spatial concepts and relations from the commonsense point of view, an aspect which is shared by our perspective.

3 THE *SEMANTIC TABLE PLUS PLUS* PROJECT

The SemT++ project proposes an interaction model supporting users in collaboratively handling digital resources. Such a model is based on the metaphor of *tables*, populated by *objects*, and is described (Goy et al., 2014b). *Tables* are *thematic contexts*, i.e., shared workspaces devoted to the management of specific activities (e.g., the management of a business project, the organization of children care, a trip planning). SemT++ tables can be seen as "round tables", where users can share information and resources, work together on a document, and so on. Table participants, in fact, can modify objects, delete them, or add new ones; invite people to "sit at the table" (i.e., to become a table participant); define meta-data, such as comments and annotations.

SemT++ provides an *abstract view* over objects lying on tables, by considering them as *information objects* that, despite their heterogeneity (they can be documents, images, to-do items, bookmarks, email conversations, and so on) can be uniformly annotated.

Moreover, SemT++ supports workspace awareness by means of: a table presence panel, showing the list of table participants currently sitting at the table; standard awareness techniques, such as icon highlighting, to notify users about table events (e.g., an object has been modified); notification messages, coming from outside SemT++ or from other tables, filtered on the basis of the topic context represented by the active table; see (Ardissono et al. 2010).

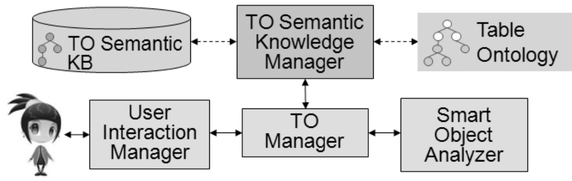

Figure 1: SemT++ architecture.

Figure 1 shows the relevant components of SemT++ architecture. The *User Interaction Manager* handles all tasks related to the interaction with users (User Interface generation, and all communications with the system, namely with the TO Manager). The *TO (Table Object) Manager* plays a "mediation" role between the User Interaction Manager and the components which represent the system "intelligence", i.e. the Smart Object Analyzer and the "semantic" components (see below). In particular, the TO Manager is in charge of all the operations which take place on tables (e.g., adding/deleting objects, comments, etc.). The *Smart Object Analyzer* provides the TO Manager with the analysis of table objects, in order to discover information about them; for example, it detects the encoding format (PDF, HTML) and it looks for parts included in the analyzed object (e.g., images, links, etc.). The *TO Semantic Knowledge Manager* manages the semantic descriptions of table objects, which are stored in the *TO Semantic KB*; such descriptions are based on the *Table Ontology*, which represents the (static) system semantic knowledge concerning *information objects*.

The Table Ontology is grounded in the Knowledge Module of O-CREAM-v2 (Magro and Goy, 2012), a core reference ontology for the Customer Relationship Management domain developed within the framework provided by the foundational ontology DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) (Borgo and Masolo, 2009) and some other ontologies extending it, among which the Ontology of Information Objects (OIO) (Gangemi et al., 2005). The Table Ontology enables us to describe table resources as *information objects*, with properties and relations. For example, a table object can have parts (e.g., images within a document), which are in turn information objects; it can be written in English and it can be stored in a PDF file, or it can be an HTML page; moreover, it has a content, which usually has a main topic and refers to a set of entities (i.e., it has several objects of discourse). Given the object description based on the Table Ontology, reasoning techniques can be applied to infer interesting knowledge, mainly from included parts; for example, if a document contains a hyperlink to a resource written in French, probably the document itself is written in French.

A detailed description of this ontology, including the inferences it enables and how such inferences are exploited to provide users with suggestions about object properties can be found in (Goy et al., 2014a).

Within the SemT++ project, we developed a proof-of-concept prototype, i.e., a Java web application, deployed on the Google App Engine, accessible through a web browser. The backend components, relying on heterogeneous technologies, are implemented as RESTful Web Services communicating by exchanging JSON objects. To store files corresponding to table objects, the current version exploits Dropbox and Google Drive API, while Google Mail is used to handle email conversations. The User Interface (UI) is a dynamic, responsive single page (client side) application, exploiting AJAX to exchange JSON objects with a set of Java Servlets (server-side). UI responsiveness, guaranteeing immediate availability on different devices, is supported by Bootstrap (getbootstrap.com). The Smart Object Analyzer exploits a Python Parser Service, able to analyze HTML documents.

Both the Table Ontology and the TO Semantic KB are expressed in OWL (www.w3.org/TR/owl2-overview); the TO Semantic Knowledge Manager exploits the OWL API library (owlapi.sourceforge.net) to interact with them. The TO Semantic Knowledge Manager also invokes the Reasoner, when required. The current Reasoner implementation is based on Fact++ (owl.cs.manchester.ac.uk/tools/fact).

We also performed some user evaluations of SemT++, which demonstrated that communication, resource sharing, and shared resources retrieval with SemT++ is significantly faster than without it, and user satisfaction is higher. The details of this first evaluation, together with the analysis and discussion of the results, can be found in (Goy et al. 2014b). Moreover, we evaluated the functionality of the User Interface enabling the exploitation of multiple criteria to perform object selection, and we found that users actually appreciate it; see (Goy et al. 2014a).

4 ENHANCING SEMT++ WITH DOMAIN KNOWLEDGE: DSEMT++

Besides the knowledge modeling table resources as information objects, represented in the Table Ontology, DSemT++ tables have been equipped with specific domain knowledge aimed at providing a semantic characterization of the entities table resources refer to, i.e. entities representing the *content* of information objects.

The two properties defined in the Table Ontology whose values refer to resource content are *hasTopic*(x, y, t) and *hasObjectOfDiscourse*(x, y, t), representing, respectively, the relation between an information element (e.g., an email conversation) and its main topic, and what a resource (e.g., a web site) "talks about".

In the evaluation of the User Interface enabling object selection based on multiple criteria, mentioned above, many users claimed that the meaning of some values of *hasTopic* and *hasObjectsOfDiscourse* properties (typically added by other table participants) can result unclear or ambiguous, and they expressed the need of having some explanations about the meaning of such values.

The possibility of classifying an individual representing a property value (e.g., *Torino*, as the value of the *hasTopic* property of an article) in a specific class (e.g., *Municipality*), and providing other information about it (e.g., its location on a map) could represent the "explanation" users were asking for. It is worth mentioning that this support to potentially ambiguous or unknown meanings of property values is particularly important within a collaborative environment such as DSemT++, where a user can be unaware of the meaning of a property value provided by another user.

To implement this functionality on DSemT++ tables, two semantic constituents are required: (a) a Domain Ontology, modeling entities representing topics and objects of discourse; (b) one or more LOD dataset, containing data/information about the chosen domain. These instruments, and knowledge provided by LOD datasets in particular, besides supporting the provision of "explanations" of the meaning of topics and objects of discourse, also offer the possibility of enriching table resources themselves, by providing links to possibly related resources; for example, if a document, lying on a table concerning the organization of a music festival, talks about the French composer Rameau, a link to DBpedia could provide suggestions for adding resources about baroque music on the table.

We thus improved the architecture of our system by adding a *Domain Knowledge Manager*, which manages the semantic knowledge concerning the content of information objects (facts about the individuals involved in the semantic representation of resources content), which is stored in the domain knowledge bases (*Domain KBs*); facts in such knowledge bases are expressed according to the corresponding *Domain Ontologies*; the set of Domain Ontologies included in the system represent the (static) semantic knowledge concerning the domains table resources are about. Domain Ontologies and Domain KBs are currently expressed in OWL and the Domain Knowledge Manager exploits the OWL API library to interact with them. The Domain Knowledge Manager also invokes the Reasoner, if required. Moreover, it handles the connection with Linked Open Data (LOD) sets. To this purpose, it exploits the *Vocabulary Mappings* (mapping LOD datasets classes and properties onto classes and properties belonging to system Domain Ontologies), and the *Instance Mappings* (mapping system and LOD datasets individuals). As we will describe in Section 4.3, in the current prototype, the Domain Knowledge Manager connects to the GeoNames Search Web Service (www.geonames.org/export).

4.1 Commonsense Geographic Knowledge

DSemT++ tables and resources lying on them can refer to a wide range of domains, so, in order to test our approach, we had to choose a specific and well-defined knowledge domain to be modeled in a proof-of-concept prototype (see Section 4.3). We considered *commonsense geographic knowledge* the suitable domain to this purpose. In this perspective, commonsense geographic knowledge is mainly intended to be a testbed, since the whole framework was designed to be reusable and to support data models describing multiple knowledge domains, possibly even on a single table.

However, besides being a testbed, commonsense geographic knowledge has an intrinsic value. In fact, together with time, space is one of the most universal and cross-domain kinds of knowledge, involved in a great number of different domains. Commonsense geospatial knowledge comes in many different ways into people's everyday life: we use geographical concepts and relations when taking a bus or a plane, when planning our holidays or when arranging an appointment with someone. The importance of geospatial knowledge in information retrieval and in knowledge organization is also claimed in the literature; see, for instance, (Giunchiglia et al., 2012).

Further evidence of its centrality can be found in the leading role geography has taken on in the evolution of both the Web 2.0 and the Web of Data (www.w3.org/2013/data) during the last ten years. Services like Google Earth, Google Maps, WikiMapia, and OpenStreetMap are enabling geographically-based user-generated content. Moreover, social networks like Foursquare, the pervasive trend of geolocalization, and resource geo-tagging increased the role of geography in our everyday life. Simultaneously, the combination of semantic technologies, the Web of Data and Geographic Information resulted in the *Semantic Geospatial Web*, a Semantic Web extension based on several spatial ontologies, able to "increase the relevance and quality of results in geographic retrieval systems" (Ballatore et al. 2013, p. 95). In such a process, the cross-domain nature of geographic information acted as a "glue" in integrating and linking different datasets. The connection role assumed by geographic information in the Web of Data is further confirmed by a recent report from the LOD workteam, where geography appears as one of the nine thematic categories the whole LOD cloud is divided in. In particular, this latest crawl of the LOD cloud shows the role of hub, together with DBpedia as general purpose dataset, assumed by GeoNames during the last three years, becoming *de facto* the reference geographic dataset in the LOD scene (Schmachtenberg et al. 2014).

The Domain Knowledge Manager introduced above, has thus been instantiated on commonsense geographic knowledge, becoming the *Geographic Knowledge Manager*. Before describing in detail how it works, we will sketch a very simple usage scenario (Section 4.2) in order to show how domain knowledge (and in particular geographic knowledge) can support table participants in building semantic descriptions and in retrieving table resources on the basis of their content.

4.2 Usage Scenario

The availability of geographic knowledge can help DSemT++ users (at least) in two tasks: the creation (and update) of semantic descriptions of table objects, and the selection of criteria to retrieve them. Consider the new object case (the update case is similar): table participants can create a new object (e.g., when they start writing a new document), or they can add to the table an existing resource (e.g., a bookmark pointing to a web page). In both cases, a new semantic representation is built through the following steps:

1. The Smart Object Analyzer automatically determines the object formats (e.g., UTF-8, HTML), its parts and their type (e.g., images included in it); moreover, it proposes candidate values for authors and languages the information object is expressed in.
2. The Semantic Knowledge Manager, by invoking the Reasoner, provides other candidates for languages, for topics and for objects of discourse (the set of candidates suggested to users is the merge of the sets of candidates proposed by the Smart Object Analyzer and the set proposed by the Semantic Knowledge Manager).
3. Users can confirm suggested values (i.e., candidate authors, languages, topics, and objects of discourse), or they can select values already used on the table for annotating other objects, or they can introduce new ones.

Now, imagine that Roby participates in a table concerning the activities of a small NGO for environment safeguard, *Save Our Earth*, together with some other volunteers. Roby has to write an article for an online local newspaper, discussing the situation of a local old farm building in Champdepraz (a small municipality in Valle d'Aosta). Roby creates a new table object (an HTML document), writes some text in it, adds a picture of the surrounding mountains and a hyperlink to a resolution by the Municipality of Champdepraz concerning a restoration project for the farm, aimed at transforming it into a hotel. When Roby clicks the "save&update" button, the creation of the object semantic representation is triggered. The Smart Object Analyzer (step 1) discovers that: the object has a HTML representation, encoded in UTF-8; it contains an image and a hyperlink; it may be written in Italian; its author is probably Roby. The Reasoner (step 2) infers the same candidate language, some candidate topics (among them Champdepraz) and a set of candidate objects of discourse (Champdepraz farm building, restoration project, Ayasse river, Mont Avic), mainly derived from topics and objects of discourse of included objects (i.e., the image and the resolution). Roby (step 3) confirms the language (*Italian*), selects *Champdepraz* among the suggested topics, and looks at the candidate objects of discourse, in order to see if some of them could well represent her article content. Roby knows the restoration project by the Municipality for the old farm building, close to the Ayasse river, but she is in doubt about another suggested object of discourse, i.e., *Mont Avic*: she knows there are a park and a mountain with the same name; does the suggested item refer to the mountain or to the park? Is it really close to the Champdepraz farm building? Should she mention it in the article and include it in the set of objects of discourse representing the content of her article?

Roby clicks on the suggested item (*Mont Avic*), to get an explanation of it. The system displays a pop-up window (see Figure 2) telling her that the selected item refers to a 3.006 m. high mountain and showing its position on a map. Moreover, a "more information" link is available: when Roby clicks it, she gets further data about Mont Avic. On the basis of this information, Roby decides to add it as an object of discourse of her article (in fact, although currently it is not explicitly mentioned in the article, the situation of a local old farm building in Champdepraz definitely has a close relation with it).

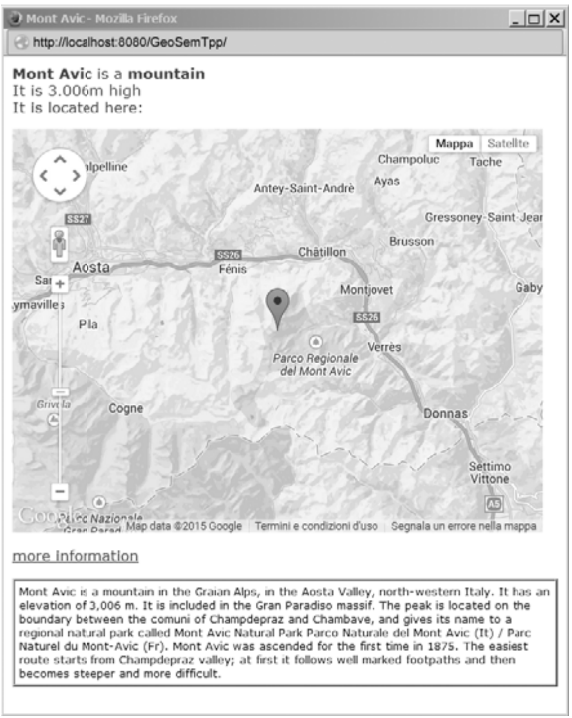

Figure 2: "Explanation" of an object of discourse in DSemT++.

Luca sits at the *Save Our Earth* table, looking for pictures of Valle d'Aosta mountains, for a photographic reportage he is going to create. He selects *topics* as the first criterion for object selection; the table presents him a list of table topics, among which Mont Avic; Luca wonders if it is intended as the mountain or the park: he clicks on the item and gets the "meaning" of the topic Mont Avic, i.e., all the information available for it (see Figure 2); such information enables him to discover that it refers to the mountain, and thus he selects it, so getting a very nice picture of Mont Avic, useful for his reportage.

4.3 The Role of the Geographic Knowledge Manager

As we mentioned, in order to provide tables with geographic knowledge, we instantiated the *Domain Knowledge Manager* module onto the geographic domain, creating the *Geographic Knowledge Manager*. The instruments we need to implement the *Geographic Knowledge Manager* functionality are a Geographic Ontology and a suitable geographic LOD dataset: in the following we will describe them.

Geographic Ontology

The semantic model of the geographic domain is provided in the Geographic Ontology. This component represents the system view of the geographic domain and its role consists in providing a vocabulary to describe the content of table resources (as far as the geographic aspects are concerned). In other words, the Geographic Ontology provides the conceptual view enabling the system to "interpret", and thus integrate, data belonging to potentially heterogeneous sources.

The Geographic KB contains all the "facts", i.e. semantic assertions, about geographic instances: each new geographic instance in DSemT++ (e.g., the instance representing Mont Avic) is classified with respect to the Geographic Ontology (e.g., as an instance of the *Mountain* class).

The Geographic Ontology is a lightweight, task- and application-oriented ontology, containing about 240 classes and a number of properties, mainly reflecting the properties used by GeoNames to describe features (such as latitude, longitude, population, altitude, etc.).

Two classes represent the top layer of the taxonomy:

- *GeoSocialEntity* includes all those geospatial entities whose existence is due to people's activities; it encompasses concrete entities, like infrastructures and human settlements, as well as concepts usually used to partition the geographic space, and administrative or political institutions.
- *GeoPhysicalEntity* includes all natural or geophysical entities like rivers, mountains, deserts, gulfs, valleys, and so on.

Although the Geographic Ontology partially reflects the GeoNames ontology (see below), it is an independent semantic model. DSemT++, in fact, is not committed to any specific external geographic dataset and thus the Geographic Ontology, by providing the system with a conceptual view over the geographic domain, enables the integration within the system of geographic data coming from different datasets and possibly originally characterized by means of different ontologies.

Thus, DSemT++ Geographic Ontology, along with the suited mappings (see *Vocabulary Mappings* section below), represents a unifying view over heterogeneous geographic semantic models, exploited in the LOD cloud.

The GeoNames dataset

First released in 2006, GeoNames is an open geospatial gazetteer gathering different official data sources (mainly from governmental organizations, institutes of geography and statistics) and combining them with users' contribution. The GeoNames database contains over 10 millions of toponyms and 9 millions of *features*, 2.8 millions of which are populated places. The features are classified according to an OWL taxonomy, the so-called GeoNames *ontology*, made up of 9 high-level classes, called *Feature Class*es, and 650 subclasses, called *Feature Code*s. Each GeoNames feature is uniquely identified by an URI and the whole gazetteer is available both in RDF and as database dump. Moreover, GeoNames makes available RESTful Web Services (www.geonames.org/export/ws-overview.html) enabling different types of queries; for example, besides a general purpose search service, search for closest toponyms, altitude of a geographic point, cities and toponyms within a user specified bounding box, postal codes, earthquakes, timezones. All services can be invoked via HTTP GET requests; the most part of the results are returned by GeoNames as an XML or JSON object, while for the *search* service it is also possible to obtain the results as RDF.

In DSemT++, we employed the *searchJSON* service, i.e., the general purpose search service returning a list of results in JSON format.

Vocabulary Mappings

In order to be exploited in DSemT++, the Geographic Ontology and GeoNames need to be "linked", so that the entities of the latter could be classified into classes of the former. We thus defined a mapping between the entities of the GeoNames ontology (o_{GN}) and the entities of our Geographic Ontology (o_{GO}), relying on the following two relations:

- $o_{GN} = o_{GO}$
- $o_{GN} < o_{GO}$

Two cases are possible: o_{GN} is a *feature code* represented in the GeoNames ontology and o_{GO} is a class of the DSemT++ Geographic Ontology, or o_{GN} and o_{GO} are both properties, the former belonging to the GeoNames ontology and the latter to the DSemT++ Geographic Ontology. Moreover, $=$ expresses conceptual equivalence, and $<$ expresses the fact that the right-hand side concept subsumes the left-hand side one. For example, Figure 3 shows the RDF/XML serialization of the axiom which states the subclass relationship between the class representing all individuals having H.STMH as Feature Code value in GeoNames ontology and the class *WaterSpring* in the Geographic Ontology.

These axioms enable us to achieve the goal of making the two ontologies intelligible to one another, and thus being able to import knowledge from the GeoNames dataset into our system.

DSemT++ Vocabulary Mappings mention 192 classes from the Geographic Ontology and 233 Feature Codes from the GeoNames ontology, establishing 186 equivalence axioms and 31 subsumption axioms.

```
<owl:Restriction>
  <rdfs:subClassOf

rdf:resource="http://www.di.unito.it/ontolog
ies/SemTppOntologies/

SemTppGeographicOntology#WaterSpring"/>
  <owl:onProperty

rdf:resource="http://www.geonames.org/ontolo
gy#featureCode"/>
  <owl:hasValue

rdf:resource="http://www.geonames.org/ontolo
gy#H.STMH"/>
</owl:Restriction>
```

Figure 3: The axiom stating the subclass relationship between a class of the GeoNames ontology and a class of the DSemT++ Geographic Ontology.

Geographic Knowledge Manager

The role of the Geographic Knowledge Manager is twofold:

- It interacts with GeoNames to retrieve information about geographic entities, i.e. about topics and objects of discourse of table resources. The *GeoManager* submodule, shown in Figure 4, is in charge of this activity.
- It interacts with the Geographic Ontology and the Geographic KB and invokes the Reasoner to classify the GeoNames entities obtained at the previous step under the Geographic Ontology schema. In order to achieve this goal it exploits the Vocabulary Mappings (described above). The *OntoMgmService* submodule, shown in Figure 4, is responsible of this activity.

The GeoManager and the OntoMgmService have been implemented, in the proof-of-concept prototype, using different technologies: the asynchronous web framework Node.js for the former, and Java Servlets, exploiting the OWL API library, for the latter. This choice has been mainly suggested by the interactions these modules have with datasets and knowledge bases, i.e. the external dataset GeoNames for the GeoManager and the OWL local ontology and KB for the OntoMgmService. Given such heterogeneity, we designed the OntoMgmService as a RESTful service, accessible through HTTP requests and exchanging data in JSON format. In particular, the OntoMgmService is identified by a URL; the GeoManager invokes it (see step 4(b) below) by sending a POST HTTP request which contains a JSON object (whose main element is the system IRI identifying the geographic instance representing the topic/object of discourse in focus). The OntoMgmService, written in Java, invokes the Reasoner (currently Fact++) in order to classify the instance in the suited class of the Geographic Ontology.

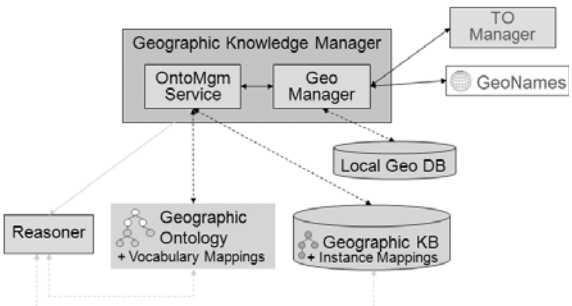

Figure 4: The Geographic Knowledge Manager architecture.

Moreover, information retrieved from GeoNames is stored in a local database (Local Geo DB), implemented in MongoDB (www.mongodb.org).

To provide a better understanding of the Geographic Knowledge Manager functionality, we describe its behavior in a typical use case:

1. The GeoManager receives from the TO Manager a string corresponding to a new topic or object of discourse (e.g., "Mont Avic"), together with the IRI referring to the instance created by the system for that topic/object of discourse (e.g., *http://www.di.unito.it/semtpp/resources/mont_avic*). The string is used as a keyword to query the GeoNames dataset, through the *searchJSON* service.
2. GeoNames returns a JSON object containing a list of entities, along with their descriptions.
3. The GeoManager sends these results back to the TO Manager, which passes them to the User Interaction Manager, thus enabling the user to select the proper entity, if any.
4. The system IRI of the instance representing the new topic/object of discourse, together with the GeoNames ID, are sent to the GeoManager, which: (a) uses the GeoNames ID to check if GeoNames data about the entity are already present in the Local Geo DB, and add them if not; (b) uses the system IRI to invoke the OntoMgmService, in order to have the instance classified with respect to the Geographic Ontology (e.g., classifying Mont Avic as an instance of *Mountain*).

In this way, the external semantic knowledge available in LOD sets (GeoNames in our prototype) is brought into the system, linked to the semantic description of table resources (as depicted in Figure 5), and available to table users: when a table user clicks on that topic/object of discourse, the result of the instance classification, together with other relevant GeoNames data (e.g., localization on a map), are displayed (see Figure 2, where the information about Mont Avic is shown).

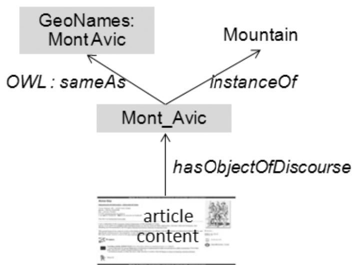

Figure 5: Adding geographic knowledge to semantic descriptions of table objects: an example.

As we shown in the usage scenario (Section 4.2), this knowledge provides table users with an "explanation" of the meaning of the topics/objects of discourse, which can be useful at least in two cases: when annotating table resources with semantic properties representing their content (i.e., topic and objects of discourse), and when selecting table objects on the basis of their content. Moreover, knowledge retrieved from LOD datasets can be exploited to enrich table resources by providing links to possibly related new contents (e.g., a link to the Gran Paradiso massif in case of a resource talking about Mont Avic).

4.4 Preliminary Evaluation

Since the enhancement of SemT++ with domain knowledge started from a need that users pointed out while evaluating our first prototype (Goy et al., 2014a), following a user-centered design approach, we contacted again the same 20 participants of the test which evaluated the use of multiple criteria to select table objects, and we asked them to perform the same task, paying attention to the fact that now an "explanation" is available for some topics and objects of discourse (i.e., for those related to geographic features). Since participants represent potential (D)SemT++ users, we ensured that they were familiar with the system already in the first evaluation.

We asked participants to rate this new functionality, on a 1 to 5 scale. We obtained an average of 4.45, indicating that the new feature was appreciated by users (the low standard deviation tells us that users tend to agree on it). In the free comments section of the brief questionnaire, some users told us that the functionality would be more interesting if not only geographic issues were supported. On the basis of this − quite obvious − observation, we are going to extend the prototype in order to connect other LOD datasets.

5 CONCLUSIONS

In this paper we presented DSemT++, an environment supporting users in the collaborative management of heterogeneous resources, enhanced with domain knowledge partially retrieved from LOD datasets.

We did not explicitly faced here all the issues concerning collaboration, both regarding resource handling and regarding collaborative metadata management. These aspects are discussed in (Goy et al. 2015). Moreover, also some issues concerning the management of semantic knowledge in DSemT++ deserve further study. For example, we are investigating how information and links retrieved from LOD datasets can be used to provide users with suggestions about content items related to the resource in focus, taking into account also the context represented by the activity the table is devoted to. Moreover, the connection of new datasets to DSemT++ currently requires, in many cases, the manual definition of the local Domain Ontology and the Vocabulary Mappings. It would be interesting to investigate the possibility of a semi-automatic support for the integration of ontologies underlying LOD datasets; see, for instance, (Zhao and Ichise, 2014). Furthermore, we are planning a new evaluation of DSemT++ with users, in order to assess the usefulness of domain knowledge within the system.

Finally, we would like to investigate the applicability of the proposed approach to other contexts, in particular to the management of archival resources. Semantic knowledge represented by ontologies and data from the LOD cloud, in fact, could represent precious instruments to enhance the access and management of such resources.

REFERENCES

Abel, F., Henze, N., Krause, D., Kriesell, M., 2010. Semantic enhancement of social tagging systems. In V. Devedžić, D. Gašević (Eds.), *Web 2.0 & Semantic Web*, Heidelberg: Springer, 25–56.

Ardissono, L., Bosio, G., Goy, A., Petrone, G., 2010. Context-aware notification management in an integrated collaborative environment. In A. Dattolo, C. Tasso, R. Farzan, S. Kleanthous, D. Bueno Vallejo, J. Vassileva (Eds.), *Int. Workshop on Adaptation and Personalization for Web 2.0*, vol. 485, CEUR, 21–30.

Ballatore, A., Wilson, D. C., Bertolotto, M., 2013. A survey of volunteered open geo-knowledge bases in the semantic web. In G. Pasi, G. Bordogna, L.C. Jain (Eds.), *Quality issues in the management of web information*, Heidelberg: Springer, 93–120.

Bardram, J. E., 2007. From desktop task management to ubiquitous activity-based computing. In V. Kaptelinin, M. Czerwinski (Eds.), *Beyond the Desktop Metaphor*, Cambridge, MA: MIT Press, 223–260.

Barreau, D.K., Nardi, B., 1995. Finding and reminding: File organization from the desktop. *ACM SIGCHI Bulletin*, 27(3), 39–43.

Borgo, S., Masolo, C., 2009. Foundational choices in DOLCE. In S. Staab, R. Studer, R. (Eds.), *Handbook on ontologies, second edition*, Heidelberg: Springer, 361–381.

Breslin, J. G., Passant, A., Decker, S., 2009. *The social semantic web*. Heidelberg: Springer.

Drăgan, L., Delbru, R., Groza, T., Handschuh, S., Decker, S., 2009. Linking semantic desktop data to the Web of Data. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, E. Blomqvist (Eds.), *The Semantic Web – ISWC 2011*, LNCS 7032, Heidelberg: Springer, 33–48.

Evangelou, C., Karacapilidis, N., 2005. On the interaction between humans and Knowledge Management Systems: A framework of knowledge sharing catalysts. *Knowledge Management Research and Practice*, 3(4), 253–261.

Gangemi, A., Borgo, S., Catenacci, C., Lehmann, J., 2005. Task taxonomies for knowledge content. Metokis Deliverable D07.

Giunchiglia, F., Dutta, B., Maltese, V., Feroz, F., 2012. A facet-based methodology for the construction of a large-scale geospatial ontology, *J. of Data Semantics*, 1(1), 57–73.

Goy, A., Magro, D., Petrone, G., Picardi, C., Segnan, M., 2015. Ontology-driven collaborative annotation in shared workspaces. *Future Generation Computer Systems*, in press.

Goy, A., Magro, D., Petrone, G., Segnan, M., 2014(a). Semantic representation of information objects for digital resources management. *Intelligenza Artificiale*, 8(2), 145–161.

Goy, A., Petrone, G., Segnan, M., 2014(b). A cloud-based environment for collaborative resources management. *Int. J. of Cloud Applications and Computing*, 4(4), 7–31.

Hu, W., Jia, C., Wan, L., He, L., Zhou, L., Qu, Y., 2014. CAMO: Integration of Linked Open Data for multimedia metadata enrichment. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandĕcíc, P. Groth, N. Noy, K. Janowicz, C. Goble (Eds.), *The SemanticWeb – ISWC 2014*, LNCS 8796, Heidelberg: Springer, 1–16.

Kaptelinin, V., Czerwinski, M., 2007. *Beyond the desktop metaphor*. Cambridge, MA: MIT Press.

Karger, D.R., 2007. Haystack: Per-user information environments based on semistructured data. In V. Kaptelinin, M. Czerwinski (Eds.), *Beyond the Desktop Metaphor*, Cambridge, MA: MIT Press, 49–100.

Kim, H., Breslin, J.G., Decker, S., Choi, J., Kim, H., 2009. Personal Knowledge Management for knowledge workers using social semantic technologies. *Int. J. of Intelligent Information and Database Systems*, 3(1), 28–43.

Magro, D., Goy, A., 2012. A core reference ontology for the customer relationship domain. *Applied Ontology*, 7(1), 1–48.

Meimaris, M., Alexiou, G., Papastefanatos, G., 2014. LinkZoo: A linked data platform for collaborative management of heterogeneous resources. In Presutti V., d'Amato C., Gandon F., d'Aquin M., Staab S., Tordai A. (Eds.), *The Semantic Web: Trends and Challenges*, LNCS 8465, Heidelberg: Springer, 407–412.

Passant, A. Laublet, P., 2008. Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In C. Bizer, T. Heath, K. Idehen, T. Berners-Lee (Eds.), *Linked Data on the Web (LDOW 2008)*, vol. 369. CEUR.

Sauermann, L., Bernardi, A., Dengel, A., 2005. Overview and outlook on the semantic desktop. In S. Decker, J. Park, D. Quan, L. Sauermann (Eds.), *Semantic Desktop Workshop*, vol. 175. CEUR.

Schandl, B., Haslhofer, B., Bürger, T., Langegger, A., Halb, W., 2012. Linked Data and multimedia: The state of affairs. *Multimedia Tools and Applications*, 59(2), 523–556.

Schmachtenberg, M., Bizer, C., Paulheim, H., 2014) Adoption of the Linked Data best practices in different topical domains. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandĕcíc, P. Groth, N. Noy, K. Janowicz, C. Goble (Eds.), *The SemanticWeb – ISWC 2014*, LNCS 8796, Heidelberg: Springer, 245–260.

Voida, S., Mynatt, E. D., Edwards, W. K., 2008. Re-framing the desktop interface around the activities of knowledge work. In *Proc. of UIST'08*, New York, NY: ACM Press, 211–220.

Yamada, I, Ito, T., Usami, S., Takagi, S., Toyoda, T., Takeda, H., Takefuji, Y., 2014. Linkify: Enhanced reading experience by augmenting text using Linked Open Data. *ISWC 2014 Semantic Web Challenge*: challenge.semanticweb.org.

Zhao, L., Ichise, R., 2014. Ontology integration for Linked Data, *J. of Data Semantics*, 3(4), 237–254.