Supporting Semantic Annotation in Collaborative Workspaces with Knowledge Based on Linked Open Data

Anna Goy[✉], Diego Magro, Giovanna Petrone,
Marco Rovera, and Marino Segnan

Dipartimento di Informatica, Università di Torino, Turin, Italy
{annamaria.goy,diego.magro,giovanna.petrone,
marco.rovera,marino.segnan}@unito.it

Abstract. The management of shared resources on the Web has become one of the most pervasive activities in everyday life, but the heterogeneity of tools and resource types (documents, emails, Web sites, etc.) usually causes users to be lost and to spend a lot of time in organizing resources and tasks. Structured semantic annotation can provide a smart support to collaborative resource organization, but, as demonstrated by our user studies, users have often to deal with ambiguous or unknown expressions, suggested by the system or by other users. As a consequence, it is important to provide them with an "explanation" of unclear annotations, which can be based on formally encoded domain knowledge, retrieved from the LOD Cloud. We chose commonsense geospatial knowledge to implement a proof-of-concept prototype providing such "explanations". After a brief presentation of the background, represented by the SemT++ project, we describe the approach and present a user evaluation of it.

Keywords: Collaborative workspaces · Semantic annotation · Linked Open Data · Semantic Web · Ontology-driven applications · Geospatial knowledge

1 Introduction

The collaborative management of shared resources on the Web has become one of the most pervasive activities, not only for knowledge workers, but for everybody in everyday life. However, due to the pervasive nature of this activity, which is required by almost every task – from buying a ticket for a concert to organizing a holiday, from writing a scientific paper to managing children activities – the number of different tools that users have to use is very large, and useful resources belong to very heterogeneous types (documents, web sites, images, email conversations, posts, etc.). In this scenario, users are often lost and spend a lot of time in trying to organize resources and tasks.

In order to take on this challenge, a mechanism to handle heterogeneous Web resources in a uniform way is required. Moreover, such a mechanism should enable sharing and collaboration among users, and should help them by providing a smart support to resource organization. Semantic technologies can be the key factor to face this challenge. Ontologies and semantic representations of the handled resources, in

fact, can endow collaborative tools with some "expertise" about the objects they are managing, and this can turn such tools into smart companions. In particular, the annotation of resources, based on shared semantic vocabularies, can enable collaborative applications to help users in both resource organization and retrieval.

However, semantic annotation typically imposes a great overload on the users: In order to gain a future advantage in organization and retrieval, users are required an extra-work to annotate resources. To alleviate such an overload, tools for the collaborative management of Web resources should provide automatic annotations or, at least, suggestions. We started the design and implementation of a (semi-)automatic support to semantic annotation of resources (exploiting HTML parsing and Named Entity Recognition); some preliminary results of this work can be found in [1]. As demonstrated by our user studies, in a system that provides users with suggestions for semantic annotation, and where annotations are collaboratively defined, users have often to deal with ambiguous or unknown expressions, suggested by the system or by other users for the annotation of resources. As a consequence, it is of paramount importance providing them with an easy and quick access to the meaning of such unclear annotations. In order to reach this goal, collaborative tools should be equipped with formally encoded domain knowledge and the Linked Open Data (LOD) Cloud (lod-cloud.net) represents a very rich source of knowledge about a wide range of domains, in the form of data and semantic models that can be exploited within collaborative applications. However, a collaborative environment can be used to carry out very different activities, each one referring to a different domain: Which are the most suited datasets? Two considerations can help us answer this question.

First, datasets based on Linked Data best practices typically provide links to related datasets: This means that from a selected set of data referring to a specific domain, other datasets are usually reachable; in particular, many datasets contains links to cross-domain resources such as DBPedia (wiki.dbpedia.org).

Second, there are some types of knowledge that, due to their intrinsic nature, can be considered (almost) universal: *Geospatial knowledge* is one of the most popular type of such a cross-domain knowledge. As demonstrated by the pervasive presence of services based on geolocation, maps, and directions functionalities, geospatial knowledge is involved in a lot of different specific domains and is used by everyone to carry on everyday activities: Geospatial concepts and relations are used when planning a journey, when taking care of environmental issues, when arranging an appointment, when organizing a conference, etc. This knowledge does not represent a scientific perspective, but instead a *commonsense* one (i.e. a perspective enabling people to distinguish different geospatial entities, to identify and to georeference them); in fact, it does not provide a formal precision degree in geographic descriptions, but instead a model enabling people to describe, and ultimately organize, representations of real-world entities, like mountains, cities, or streets.

The cross-domain nature of geospatial information is further confirmed by a recent report by the LOD work team, where geography appears as one of the nine thematic categories the whole LOD cloud is divided into [2]. In particular, GeoNames (www.geonames.org) has assumed, together with DBPedia, a role of hub (see: lod-cloud.net), becoming the de facto reference geospatial dataset in the LOD Cloud. Moreover, geospatial knowledge can act as a "glue" in integrating and linking different datasets [3].

These considerations led us to choose *commonsense geospatial knowledge* as the first testbed for our approach, aimed at endowing shared workspaces with domain knowledge. However, the proposed architecture (see Sect. 4.3) has a more general validity, and can be used to include different types of domain knowledge within collaborative environments.

Ultimately, the main contribution described in this paper shows the role played by (geospatial) ontologies and data retrieved from the LOD Cloud in the implementation of the previously mentioned "explanation functionality" within a collaborative environment for Web resource management.

The rest of the paper is organized as follows. In Sect. 2 we discuss the main related work; in Sect. 3 we briefly summarize the main characteristics of the SemT++ project, representing the background of the proposal presented in this paper. In Sect. 4, which describes the contribution with respect to our earlier work, we present the motivations of the presented approach, we sketch a simple usage scenario, and we describe the implementation of the "explanation functionality", together with a user evaluation of it. Section 5 concludes the paper and outlines future developments.

2   Related Work

Several research fields have to be taken into account in order to outline the background reference work for the approach presented in this paper.

As far as the original idea underlying SemT++ is concerned, an interesting reference work is represented by [4], presenting the problems of the *desktop metaphor* and several approaches trying to replace it. In particular, an interesting model presented in the mentioned book is Haystack [5], a flexible and personalized system enabling users to define and manage workspaces referred to specific tasks. Another interesting set of approaches are those grounded into Activity-Based Computing, where the core concept structuring the interaction model is that of *user activity* [6, 7]. The main enhancement of SemT++ with respect to these approaches is the explicit geospatial knowledge model and the exploitation of LOD sets, discussed in this paper.

Another relevant research field that is worth to be considered is represented by research about social tagging systems, where resources can be tagged with meta-data referring to different aspects (*facets*); the user-centered, bottom-up tagging process leads to the creation of multi-facets classifications called *folksonomies* [8]. Interesting semantic enhancements of tagging systems have been developed [9], with particular attention to knowledge workers [10]. With respect to social tagging systems, SemT++ shifts its focus from mass social communities to (small) collaborative groups of people sharing specific activities.

In general, the idea of exploiting semantic technologies to support collaborative resource management is not new. For example, a new research area has recently emerged, the *Social Semantic Web* [8], an approach relying 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.

Another project aimed at coupling desktop-based user interfaces and Semantic Web is the *Semantic Desktop* [11]. In particular, the NEPOMUK project (nepomuk. semanticdesktop.org) defined an open source framework for implementing semantic desktops that rely on a set of ontologies and integrate existing applications to support collaboration among knowledge workers. Drăgan and colleagues [12] propose an approach to connect the Semantic Desktop to the Web of Data: This enables the system to "bring Web data to the user", thus supporting the exploitation of external data within the user personal context. The proposal by Drăgan and colleagues is one of a great number of recent semantic approaches trying to use LOD to enhance services for the users. In the same direction, the LinkZoo tool [13] propose a collaborative annotation platform based on LOD: Semantic annotations are stored as RDF triples and they enable LinkZoo to couple standard keyword search with property-based filtering. [14] contains a survey of the approaches to exploit LOD in metadata for multimedia content, while Linkify [15] is an add-on for major browsers that adds a link to Named Entities recognized in online texts, pointing to a mashup of information items extracted from LOD sources. Passant and Laublet [16] present MOAT (Meaning Of A Tag), a semantic framework for the definition of machine-readable meanings of tags: Tags are represented as quadruples (<*User, Resource, Tag, Meaning*>), their meaning is linked to well-known LOD sets (such as DBpedia and GeoNames) and can be shared with other uses.

Lots of work has also been done, within different research communities, in the field of collaborative semantic annotation. In the NLP field, tools have been implemented aimed at supporting collaborative annotation of textual corpora [17]. NLP-oriented annotation tools enable users to associate "semantic" labels to phrases within a text and usually refer to an annotation schema that can be formally encoded as an ontology (e.g., [18, 19] among many others).

In the Knowledge Management field, a similar notion of "semantic annotation" has been used, where annotations link words or phrases within a document to instances in a semantic knowledge base (and indirectly to classes of a domain ontology); see, for instance, [20], which also contains an interesting survey of annotation frameworks. A good survey of ontology-based annotation environments can also be found in [21]. Usually, ontologies provide the metadata structure, and describe document properties, such as author, date, format, etc. (e.g., Dublin Core: www.dublincore. org). In some cases, annotation systems can rely on more domain-dependent semantic resources (e.g., the Getty Thesaurus of Geographic Names: www.getty.edu/research/ tools/vocabularies/tgn in the geographic domain). Annotations have also been widely used in e-learning [22] and in the so-called "semantic wikis"; for instance, Buffa and colleagues [23] describe SweetWiki, a wiki tool supporting a structured, semantic annotation of resources.

Finally, given that geospatial knowledge plays a major role in the proposal presented in this paper, we dedicate the final part of this related work overview to it. The importance of geospatial knowledge, especially in information retrieval and knowledge organization, is claimed in the literature (see, for instance, [24]), and is demonstrated by the leading role that geography acquired in the Web of Data during the last ten years.

In particular, the growth of Web 2.0 and its related practices, like crowdsourcing, found in Geographic Information a preferential knowledge domain. Goodchild termed the so gathered information *volunteered geographic information*, and considered it an interesting example of user-generated content [25]. This phenomenon emphasizes the major role played by a commonsense perspective over geographic knowledge: Services like OpenStreetMap, WikiMapia, Google Earth, Google Maps were contributing to change the way people interact with the Web, turning them into information prosumers, rather than mere information consumers. Moreover, the recent mobile revolution, the availability of social networks like Foursquare, and the pervasive trend of geolocation and resource geo-tagging, increased the role of geospatial knowledge in our everyday life. Within this scenario, ElGindy and Abdelmoty propose a framework for analyzing folksonomies derived from geo-tagging activity and discovering place-related semantics (e.g., events, activities, personal opinions, and so on). The results of such an analysis reveals "a much richer structure of concepts and relationships than those defined in a formal data source produced by experts" ([26], p. 222). Moreover, the synergy among semantic technologies, Web of Data and Geographic Information resulted in the establishment of the *Semantic Geospatial Web*, a Semantic Web extension based on a set of spatial ontologies that can be exploited in geography-based retrieval systems, leading to better quality results [27]. In conclusion, it is worth mentioning the geospatial ontology *Space*, based on GeoNames, WordNet and MultiWordNet [24]. *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 Shared Workspaces as "Round Tables" and Resources as "Information Objects"

The background of the work described in this paper is represented by the *Semantic Table Plus Plus* (SemT++) project. SemT++ started from the idea that shared workspaces could be seen as "round tables", where people sit together in order to collaboratively carry on an activity (such as planning a journey, organizing children care, participating in the social activities of a NGO, write a scientific paper for a conference) [28]. Table participants typically use different types of resources to perform the tasks required by the specific activity: They get information from Web sites, read papers, write documents, watch videos or photo galleries, write emails and posts, and so on. The resources useful to carry on the activity the table is devoted to are typically encoded in different formats, handled by different applications, and stored in different places, although they typically refer to the same semantic context. We thus designed an interaction model aimed at providing an *abstract view* over table resources by handling them as *information objects*, lying on a table, collaboratively managed (added, deleted, modified, annotated) by table participants.

In SemT++, workspace awareness is guaranteed on each table by standard mechanisms such as a presence panel (showing the list of table participants currently sitting at the table); icon highlighting (to notify users about table events); notification messages (filtered on the basis of the topic context represented by the active table [29]).

One of the most important features of SemT++ is that each table is endowed with semantic representations of the resources lying on it. Such representations are based on the *Table Ontology*, a semantic model grounded in O-CREAM-v2 [30], 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) [31] and one of its extensions, namely the Ontology of Information Objects (OIO) [32]. The Table Ontology enables SemT++ to represent resources (documents, Web pages, email threads, images, etc.) as *information objects*, with properties and relations. Figure 1 shows a simplified example: The semantic representation of a Web page encoded in UTF-8/HTML5, written in Spanish by Carlos, and containing two parts – an image and a link (to a pdf document); moreover, the represented Web page has a main *topic* (*Cuba music tour*) and it refers to a set of entities (called *objects of discourse*: *Havana*, *Santiago de Cuba*, *salsa*, *Camagüey*).

SemT++ also includes a Reasoner that, on the basis of the semantic representations of the resources lying on a table, can infer possible properties of related resources; for example, if a document contains a hyperlink to a resource written in Spanish, the Reasoner infers that probably the document itself is written in Spanish. A detailed description of the Table Ontology, including classes, relations and axioms supporting the reasoning mechanisms can be found in [33].
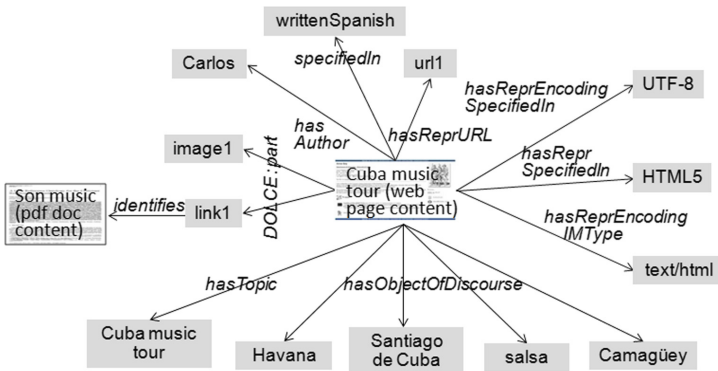

Fig. 1. Semantic representation of a resource (Web page) – simplified example.

The semantic representations of table resources is a kind of *semantic annotation*, in which the structure is provided by the Table Ontology. These representations are collaboratively built by users, with a significant support provided by the system. In fact, when a table participant adds a new resource to a table, a new semantic representation is set up on the basis of contributions from the system and from table participants. In particular, the system (actually, the *Smart Object Analyzer* module) analyzes the resources and, on the basis of the results of the analysis, it performs the following tasks:

- It defines some properties: Typically, the *format* (e.g., UTF-8, HTML) and the *parts* (e.g., images included in the analyzed resource). Parts are proposed to users, who can select the interesting ones and add them to the table.
- It proposes candidate values for other properties: The *author* of the resource, the *language* the resource is expressed in, the main *topic* and the *objects of discourse* (representing the resource *content*). These suggested values are identified as follows:

Authors, from meta information of the document itself;
Language, from meta information of the document itself or from the reasoning mechanisms;
Topic and *objects of discourse*, from meta information about the document (e.g., HTML meta-tags), from the results of a Named Entity Recognition (NER) service; from the Reasoner, which – on the basis of the Table Ontology – infers suggested values.

Users can confirm or discard the suggestions, or they can add new values from scratch. The activity performed by table participants in the definition of the semantic representation of resources can be seen as a *collaborative annotation* task: Values for object properties can be added, deleted, or modified, according to the collaboration policy defined on the table: In case of a *consensual* policy, all participants can always edit semantic representations, while in case of an *authored* or *supervised* policy the final decision about the semantic annotation of a resource is taken by its creator or by the table supervisor; a detailed account of collaboration policies in SemT++ can be found in [34]. Semantic representations represent a *shared view* over table resources; however, each table participant can also keep a *personal view* over resources, containing "private" annotations; see [35] for details.

The current SemT++ proof-of-concept prototype is a Java Web application accessible through a Web browser. Backend core functionality is provided by Java components while services relying on heterogeneous technologies are accessible through a RESTful interface: For example, a Python Parser Service provides the HTML analysis, while a Node.js module is in charge of interfacing with the NER Service, based on TextRazor (www.textrazor.com). Files corresponding to table objects are managed through Dropbox, Google Drive, and Google Mail APIs. The User Interface (UI) is based on Bootstrap (getbootstrap.com), which guarantees responsiveness and thus availability on different devices.

The Table Ontology, as well as the knowledge base containing the semantic representations of table resources, are expressed in OWL (www.w3.org/TR/owl2-overview), and the OWL API library (owlapi.sourceforge.net) is used to interact with them. The current Reasoner implementation is based on Fact++ (owl.cs.manchester.ac.uk/tools/fact).

We performed some user evaluations of the system, at different development stages. The first evaluation is presented and discussed in [28], and demonstrated that communication among users, resource sharing, and resources retrieval is significantly faster when using SemT++ with respect to performing the same tasks without it. The second evaluation is presented and discussed in [33], and told us that users

appreciate the functionality of SemT++ User Interface enabling the exploitation of multiple criteria (in particular, resource content) to perform object selection.

Moreover, we carried on a qualitative user study with the goal of defining the model that supports collaborative semantic annotation of table resources, and in particular the suitable collaboration policies [34]. The main results of this user study are the implementation of the mentioned collaboration policies (*consensual*, *authored*, *supervised*) that can be set on each table, and the implementation of the *personal views* functionality, clearly requested by users.

4 Providing Explanations Based on Commonsense Geospatial Knowledge in Collaborative Annotation

4.1 The Need for Knowledge About Resource Content

The "expertise" of SemT++ represented by the Table Ontology mainly refers to resources (documents, Web pages, images, email threads, etc.) as *information objects*, i.e., it includes knowing that they are encoded in specific formats, they usually have one or more authors, and so on. This knowledge also includes knowing that information objects typically have a *content*, structured in a main *topic* and a set of entities the resource "talks about" (the *objects of discourse*). However, the Table Ontology does not provide any knowledge about the content itself: If a document talks about "New York", no semantic representation is associated to such a value, which is simply a string.

In the evaluation aimed at testing the availability of multiple selection criteria [33], as well as in the qualitative user study used to define the model handling collaborative semantic annotation of table objects [34], many users pointed out that the meaning of some topics and objects of discourse were unclear and that a sort of explanation would have been very helpful. The examples of unclear values mentioned by users are system suggestions (see Sect. 3) or annotations provided by other table participants. Since system suggestions and collaboration among users are two core aspects of SemT++, it is clear that the system needs to be enhanced with the required "explanation functionality".

The analysis of the users' answers in the mentioned evaluation and user study also showed that a significant number of examples of unclear values refer to places (villages, monuments, regions, mountains, etc.), and users would like to know their nature (is Saint Barthélemy a village or a valley?) or their geolocation (where is Saint Barthélemy? How far is it from Aosta?). These feedbacks from users suggested us two things:

– Besides the semantic knowledge representing resources as information objects (encoded in the Table Ontology), SemT++ tables need to be endowed with specific domain knowledge, aimed at providing a semantic characterization of the entities representing resource *content*; this knowledge would enable the system to offer "explanations" of the unclear property values.

– A significant aspect of the knowledge about resource content is represented by *commonsense geospatial knowledge*, enabling the system to provide information about places (villages, monuments, regions, mountains, etc.). Obviously, the relevance of such a knowledge depends on the specific activity a single table is devoted to: It is intuitively very important on a table devoted to the organization of a journey, while it seems definitely less useful on a table set up to write a paper about neural networks. However, as already claimed in the Sects. 1 and 2, commonsense geospatial knowledge represents a very important cross-domain knowledge, and it can play a major role in SemT++, in particular as far as the "explanation functionality" is concerned.

On the basis of the just presented motivations, we designed and implemented a new module, the *Geospatial Knowledge Manager*, in charge of managing commonsense geospatial knowledge on SemT++ tables. The *Geospatial Knowledge Manager* will exploit geospatial information retrieved from GeoNames, the most popular geographic dataset in the LOD Cloud. Before describing how the *Geospatial Knowledge Manager* works (Sect. 4.3), we sketch a very simple usage scenario (Sect. 4.2), showing how geospatial knowledge can provide table participants with "explanations" of topics and objects of discourse representing resource content.

4.2 Usage Scenario

Imagine that John, together with a group of friends, participates in a table devoted to the organization of a journey to Cuba. John and his friends are particularly interested in cultural and sustainable tourism. The discussion about the itinerary started a few days earlier, and the table is currently populated by some bookmarks to Web sites describing travels in Cuba. Browsing the Web, John finds another interesting site, proposing a music tour of the island, thus he decides to add it to the table. When the new object is dropped on the table, the system starts its analysis, finding that it is an HTML document (encoded in UTF-8) and it is probably written in Spanish. Moreover, the Smart Object Analyzer module discovers that it contains several images and hyperlinks, which represent its parts; parts are proposed to John: he selects an image (showing the tour steps on a map) and an e-book about Son music (linked in the Web page) and adds them to the table. The system also suggests some candidate topics and a set of candidate objects of discourse (among which: *Cuba*, *Havana*, *Santiago de Cuba*, *Moncada Barracks*, *Camagüey*, *Music of Cuba*, *salsa*).

John confirms the language (*Spanish*), provides *Cuba music tour* as main topic, and looks at the candidate objects of discourse, in order to see if some of them could well represent the Web page content. John is uncertain about one of them, which seems to be interesting, namely *Camagüey*: Is it a city, a small village, or a beach? Is it relevant for the music tour and, in general, for a cultural travel through Cuba? Should it be mentioned as objects of discourse of the selected Web page? John thus clicks on the linked item (*Camagüey*) to get an explanation: The system displays a pop-up window, shown in Fig. 2, providing information about Camagüey, namely the kind of place (a *City*), a short description, and its position on a map. On the basis of this information, John decides to add it as an object of discourse.

Fig. 2. "Explanation" of an object of discourse on a SemT++ table.

Later on, Mary sits at the table to have a look to the new items: She is notified about the new Web page added by John, and she takes a look at the semantic description (annotations), in order to have a view at-a-glance of its content. Intrigued by *Camagüey*, she clicks on it, gets the explanation (Fig. 2), and starts looking for further information about the Cuban city.

4.3   The Geospatial Knowledge Manager

A preliminary version of the Geospatial Knowledge Manager and its role is described in [36]. In the following we describe its current architecture and present a user evaluation of it (Sect. 4.4). The main components of the Geospatial Knowledge Manager module are shown in Fig. 3.
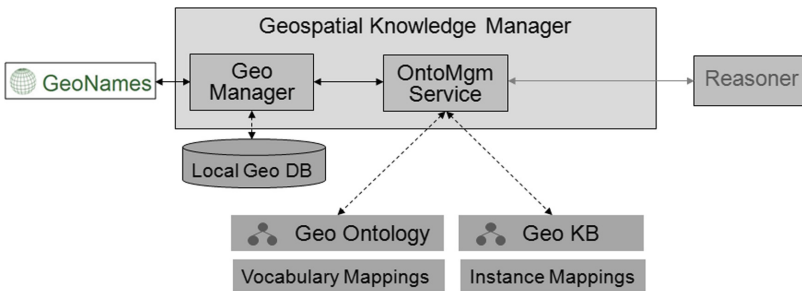

Fig. 3. Architecture of the geospatial knowledge manager.

We describe the role of the different components in the following.

Geo Ontology. The semantic model representing the commonsense geospatial knowledge is provided in the Geo Ontology. This component represents the system "expertise" about geospatial issues, from a commonsense perspective, as described in Sect. 1 and 2. It provides a vocabulary to describe the content of table resources, as far as the geospatial

aspects are concerned, and thus it enables the system to "interpret" geospatial data belonging to potentially heterogeneous sources. This role is one of major importance within SemT++: The Geo Ontology, in fact, provides the conceptual framework needed to integrate data coming from different datasets and possibly originally characterized by means of different ontologies.

The Geo Ontology is a lightweight, application ontology, containing classes (about 240) and properties mainly reflecting the properties used by GeoNames to describe geographic features (latitude, longitude, population, altitude, etc.). The top layer of the taxonomic structure of the Geo Ontology is represented by two classes: *GeoSocialEntity* – that includes all the geospatial entities created by people's activities: For example, infrastructures, human settlements, administrative and political institutions, but also concepts used to partition the geographic space (such as regions or borders) – and *GeoPhysicalEntity* – that includes natural or geophysical entities like valleys, rivers, deserts, mountains, and so on.

It is worth underlining that, although the Geo Ontology partially reflects the GeoNames ontology (see below), it is an independent semantic model, aimed at representing a conceptual vocabulary useful to integrate data from heterogeneous sources. This choice also ensures SemT++ not to be committed to any specific external semantic model, and thus to any specific dataset.

Geo KB. The Geo KB is the knowledge base containing all the semantic assertions, i.e. the "facts", about geospatial instances, expressed according to the vocabulary provided by the Geo Ontology. In particular, each geospatial instance (e.g., the instance representing Camagüey) is classified with respect to at least a class of the Geo Ontology (e.g., as an instance of the *City* class).

GeoNames. GeoNames is an open geospatial gazetteer, released in 2006, gathering different data sources provided by governmental organizations, institutes of geography and statistics, as well as users' contribution. The GeoNames dataset contains over 10 millions of toponyms and 9 millions of *features*, uniquely identified by URIs, and classified according to the GeoNames *ontology*, a taxonomy including 9 high-level classes, called *feature classes*, and 650 subclasses, called *feature codes*. GeoNames offers a number of RESTful Web Services (www.geonames.org/export/ws-overview. html) enabling different types of search: A general purpose string-based search, a search for closest toponyms, for the altitude of a geographic point, for cities and toponyms within a user specified bounding box, for postal codes, for earthquakes, and so on. The most part of GeoNames services return XML or JSON objects, while only in some cases (for example for the general search service) RDF results are available. In SemT++, we used the *search* service, i.e., the general purpose search service returning a list of results in JSON format.

Vocabulary Mappings. Vocabulary Mappings represent the alignment between SemT++ Geo Ontology and the GeoNames ontology, and enables GeoNames entities to be classified into classes of the Geo Ontology. Vocabulary Mappings rely on two relations, *conceptual equivalence* and *subsumption*; both these relations can involve a GeoNames *feature code* and a class of the SemT++ Geo Ontology, or they can involve

two properties, one belonging to the GeoNames ontology and the other to the SemT++ Geo Ontology.

The *conceptual equivalence* relation is used to state that a GeoNames *feature code* and a Geo Ontology class (or a GeoNames property and a Geo Ontology property) are equivalent; the *subsumption* relation is used to state that a GeoNames *feature code* represents a subclass of a Geo Ontology class (or that a GeoNames property represents a subproperty of a Geo Ontology property). For example, the following is the RDF/XML serialization of the axiom stating the subsumption relationship between the class representing all individuals having H.STMH as *feature code* value in the Geo-Names ontology and the class *WaterSpring* in the Geo Ontology:

```
<owl:Restriction>
  <rdfs:subClassOf
    rdf:resource="http://www.di.unito.it/ontologies/SemTppOntologies/
                  SemTppGeographicOntology#WaterSpring"/>
  <owl:onProperty
    rdf:resource="http://www.geonames.org/ontology#featureCode"/>
  <owl:hasValue
    rdf:resource="http://www.geonames.org/ontology#H.STMH"/>
</owl:Restriction>
```

Current Vocabulary Mappings mention 192 classes belonging to the Geo Ontology and 233 *feature codes* from GeoNames, defining 186 conceptual equivalence relations and 31 subsumption relations. Obviously, new Vocabulary Mappings have to be defined if a new dataset, relying on a different ontology, has to be integrated within the system.

Instance Mappings. Instance Mapping represent the correspondences between SemT++ and GeoNames individuals (e.g. the instance representing the city of Cama-güey in SemT++ and the instance representing it in the GeoNames dataset).

GeoManager. The GeoManager submodule interacts with GeoNames to retrieve information about geospatial entities. The current version of the GeoManager relies on the asynchronous Web framework Node.js and uses a local database (Local Geo DB), implemented in MongoDB (www.mongodb.org), to locally store information retrieved from GeoNames.

OntoMgmService. The OntoMgmService manages all the interactions with the Geo Ontology and the Geo KB and invokes the Reasoner (see Sect. 3) to classify Geo-Names entities in the Geo Ontology, on the basis of the Vocabulary Mappings described above. The current version of the OntoMgmService exploits Java Servlets and the OWL API library, and offers a RESTful service interface, providing results in JSON format.

The Geospatial Knowledge Manager implements the "explanation functionality" depicted in the usage scenario above (Sect. 4.2). In the following we detail the steps performed by the Geospatial Knowledge Manager in order to achieve this goal.

When a new topic or object of discourse is added to the semantic representation of a table resource by a user, or when a candidate topic/object of discourse is proposed by the system (as the case of *Camagüey* mentioned in the usage scenario above), the corresponding string – together with the IRI referring to the instance created by the system for that topic/object of discourse – is passed to the Geospatial Knowledge Manager, more specifically to the GeoManager submodule.

The GeoManager checks in the Local Geo DB if the information about that item had already been retrieved from the LOD, and, if not, it invokes the GeoNames *search* service, getting a JSON object containing a list of entities, along with their descriptions. The GeoManager tries to select the relevant entry using simple heuristics (e.g., the presence of the searched string in the name of the GeoNames entity, the position in the results); in some cases (e.g., the case of *Camagüey* in our usage scenario) this lead to a complete disambiguation, while in other cases the user will be presented with a list of alternative possible "explanations", among which s/he can choose the suited one.

At this point, the GeoManager invokes the OntoMgmService, in order to have the instance (identified by a system IRI) classified with respect to the Geo Ontology (e.g., classifying Camagüey as an instance of *City*). Such a classification enables the system to provide the main information item within the explanaion, i.e., the entity type (allowing users to know, for example, that *Camagüey* is a city and not a beach). Moreover, the 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 Fig. 4), and available to table users: When a table user clicks on that (candidate) topic/object of discourse, the result of the instance classification, together with other relevant GeoNames data (e.g., localization on a map, description usually from linked DBpedia data), are displayed (see Fig. 2, where the information about Camagüey is shown).

As shown in the usage scenario (Sect. 4.2), this knowledge provides table participants with an "explanation" of the meaning of the (candidate) topics/objects of discourse, which can be useful when annotating table resources with semantic properties representing their content. Furthermore, these "explanations" can also support users in resource selection.
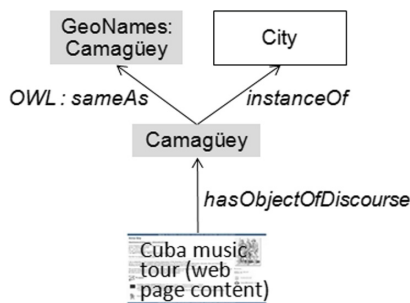

Fig. 4. The semantic representation of a Web page including geospatial knowledge.

In the following section we presents the results of a user evaluation aimed at testing the usefulness of this functionality.

4.4   Evaluation

Following a user-centered design approach, after having implemented a functionality based on the user feedback from the first evaluation round (see Sect. 4.1), we contacted again 8 of the users involved in our previous studies, in order to test the new "explanation functionality". We set up the table of our usage scenario (Sect. 4.2), i.e. a shared workspace were users could organize a cultural and sustainable travel to Cuba. The table was configured with a *consensual* policy for the management of collaborative annotation (see Sect. 3). We invited our participants to imagine that the discussion is now focused on the itinerary and we provided them with a set of bookmarks proposing several tours, some of them being thematic ones (like a music and dance tour, or a horse riding journey).

We left participants some time to take a look at the available resources, in order to become familiar with the context. Then we asked three of them to select a resource each, to be added to the table. This triggers the collaborative annotation process of the added resources, in which the system suggests property values (as described in Sect. 3) and users can select system suggestions or add new values. We asked users to concentrate on *topics* and *objects of discourse* (even though other properties – such as *language*, *author*, etc. – were also available). When the participants reached a stable agreement about the annotation, we stopped the annotation activity.

We recorded the number of times users used the "explanation functionality" available; see Table 1. Moreover, after the test, we asked them to rate, in a 1 to 5 scale, the usefulness of the functionality (1 meaning totally useless, 5 definitely useful); see Table 2. Finally, we collected all free comments they had about the experience.

Table 1. Number of times users used the "explanation functionality".

user1	user2	user3	user4	user5	user6	user7	user8	# users	# times
3	1	2	0	0	3	1	0	5	10

Table 2. Users' evaluation of the "explanation functionality".

user1	user2	user3	user4	user5	user6	user7	user8	mean	st.dev.
5	4	4	3	5	5	5	2	4.125	1.126

From results shown in Table 1 we can see that 5 users out of 8 used the "explanation functionality", while 3 of them did not use it at all. Quite interestingly, one of them (*user5*) rate the usefulness of the functionality at the higher degree (see Table 2), despite that s/he never used it during the test.

In Table 2 we can see that the average rate is 4.125 (on a 1 to 5 scale), indicating that the new feature was appreciated by users; the quite low standard deviation (1.126) tells us that users tend to agree on it (in fact, nobody rated it as totally useless).

Analyzing free comments, we can find interesting suggestions to improve the functionality. Three users said that the functionality would be more interesting if not only geospatial issues were supported; two users pointed out that in the cases in which more than one explanation were available, reading the explanations was quite annoying: Since this derives from the fact that often results from GeoNames search are not unique, it implies that a greater effort should be devoted to the disambiguation phase.

5 Conclusions and Future Work

In this paper we presented the "explanation functionality" implemented within the SemT++ system, providing users with information – retrieved from the LOD cloud – about the entities referred to when describing resource content in a collaborative semantic annotation environment. In particular, we claimed the central, cross-domain role played by commonsense geospatial knowledge in such a context and described the current implementation of the mentioned functionality, together with a user evaluation.

Some open issues clearly emerged from the presented approach. For example, the connection of new datasets, different from GeoNames and in general from geography-oriented datasets, currently requires the manual definition of a local semantic model (taking the role of the Geo Ontology) and of the corresponding Vocabulary Mappings. The investigation of semi-automatic approaches to ontology integration would be interesting; see, for instance, [37].

Moreover, knowledge retrieved from LOD sets could be used to provide users with suggestions about possibly related resources (for example, if a document, lying on a table concerning the organization of travel to Cuba, talks about Son music, a link to DBpedia could provide suggestions for adding resources about Caribbean music on the table).

References

1. Goy, A., Magro, D., Petrone, G., Picardi, C., Rovera, M., Segnan, M.: Semi-automatic support to semantic annotation of web resources in SemT++. Technical report #2015-15, Departent of Computer Science, University of Torino (2015)
2. Schmachtenberg, M., Bizer, C., Paulheim, H.: Adoption of the linked data best practices in different topical domains. In: Mika, P., et al. (eds.) ISWC 2014. LNCS, vol. 8796, pp. 245–260. Springer, Heidelberg (2014). doi:10.1007/978-3-319-11964-9_16
3. Hart, G., Dolbear, C.: Linked Data: A Geographic Perspective. CRC Press, London (2013)
4. Kaptelinin, V., Czerwinski, M. (eds.): Beyond the Desktop Metaphor. MIT Press, Cambridge (2007)

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

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

7. Voida, S., Mynatt, E.D., Edwards, W.K.: Re-framing the desktop interface around the activities of knowledge work. In: UIST 2008, pp. 211–220. ACM Press, New York (2008)

8. Breslin, J.G., Passant, A., Decker, S.: The Social Semantic Web. Springer, Heidelberg (2009)

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

10. Kim, H., Breslin, J.G., Decker, S., Choi, J., Kim, H.: Personal knowledge management for knowledge workers using social semantic technologies. Int. J. Intell. Inf. Database Syst. **3**(1), 28–43 (2009)

11. Sauermann, L., Bernardi, A., Dengel, A.: Overview and outlook on the semantic desktop. In: 1st Workshop on The Semantic Desktop at ISWC 2005, vol. 175. CEUR-WS (2005)

12. Drăgan, L., Delbru, R., Groza, T., Handschuh, S., Decker, S.: Linking semantic desktop data to the web of data. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011. LNCS, vol. 7032, pp. 33–48. Springer, Heidelberg (2011). doi:10.1007/978-3-642-25093-4_3

13. Meimaris, M., Alexiou, G., Papastefanatos, G.: 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, vol. 8465, pp. 407–412. Springer, Heidelberg (2014)

14. Schandl, B., Haslhofer, B., Bürger, T., Langegger, A., Halb, W.: Linked data and multimedia: the state of affairs. Multimed. Tools Appl. **59**(2), 523–556 (2012)

15. Yamada, I., Ito, T., Usami, S., Takagi, S., Toyoda, T., Takeda, H., Takefuji, Y.: Linkify: enhanced reading experience by augmenting text using linked open data. In: ISWC 2014 Semantic Web Challenge (2014). challenge.semanticweb.org

16. Passant, A., Laublet, P.: Meaning of a tag: a collaborative approach to bridge the gap between tagging and linked data. In: Bizer, C., Heath, T., Idehen, K., Berners-Lee, T. (eds.). Linked Data on the Web (LDOW 2008), vol. 369. CEUR (2008)

17. Bontcheva, K., Cunningham, H., Roberts, I., Tablan, V.: Web-based collaborative corpus annotation: requirements and a framework implementation. In: Witte, R., Cunningham, H., Patrick, J., Beisswanger, E., Buyko, E., Hahn, U., Verspoor, K., Coden, A.R. (eds.) LREC 2010 workshop on New Challenges for NLP Frameworks, pp. 20–27 (2010)

18. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A., Damljanovic, D., Heitz, T., Greenwood, M. A., Saggion, H., Petrak, J., Li, Y., Peters, W.: Text Processing with GATE (Version 6): (2011). gate.ac.uk

19. Fragkou, P., Petasis, G., Theodorakos, A., Karkaletsis, V., Spyropoulos, C.: Boemie ontology-based text annotation tool. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Tapias D. (eds.) Proceedings of the International Conference on Language Resources and Evaluation (LREC 2008), European Language Resources Association (ELRA) (2008)

20. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F.: Semantic annotation for knowledge management: requirements and a survey of the state of the art. J. Web Semant. **4**(1), 14–28 (2006)

21. Corcho, O.: Ontology based document annotation: trends and open research problems. Int. J. Metadata Semant. Ontol. **1**(1), 47–57 (2006)
22. Su, A.Y.S., Yang, S.J.H., Hwang, W.Y., Zhang, J.: A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments. Comput. Educ. **55**, 752–766 (2010)
23. Buffa, M., Gandon, F., Ereteo, G., Sander, P., Faron, C.: SweetWiki: a semantic wiki. Web Semant. **6**(1), 84–97 (2008)
24. Giunchiglia, F., Dutta, B., Maltese, V.: Feroz, F: A facet-based methodology for the construction of a large-scale geospatial ontology. J. Data Semant. **1**(1), 57–73 (2012)
25. Goodchild, M.F.: Citizens as sensors the world of volunteered geography. GeoJournal **69**(4), 211–221 (2007)
26. ElGindy, E., Abdelmoty, A.: Capturing place semantics on the geosocial web. J. Data Semant. **3**(4), 207–223 (2014)
27. Ballatore, A., Wilson, D.C., Bertolotto, M.: A survey of volunteered open geo-knowledge bases in the semantic web. In: Pasi, G., Bordogna, G., Jain, L.C. (eds.) Quality issues in the management of Web information, pp. 93–120. Springer, Heidelberg (2013)
28. Goy, A., Petrone, G., Segnan, M.: A cloud-based environment for collaborative resources management. Int. J. Cloud Appl. Comput. **4**(4), 7–31 (2014)
29. Ardissono, L., Bosio, G., Goy, A., Petrone, G.: Context-aware notification management in an integrated collaborative environment. In: UMAP 2009 Workshop on Adaptation and Personalization for Web2.0, pp. 23–39. CEUR (2010)
30. Magro, D., Goy, A.: A core reference ontology for the customer relationship domain. Appl. Ontol. **7**(1), 1–48 (2012)
31. Borgo, S., Masolo, C.: Foundational choices in DOLCE. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, 2nd edn, pp. 361–381. Springer, Heidelberg (2009)
32. Gangemi, A., Borgo, S., Catenacci, C., Lehmann, J.: Task Taxonomies for Knowledge Content. Metokis Deliverable D07 (2005)
33. Goy, A., Magro, D., Petrone, G., Segnan, M.: Semantic representation of information objects for digital resources management. Intelligenza Artificiale **8**(2), 145–161 (2014)
34. Goy, A., Magro, D., Petrone, G., Picardi, C., Segnan, M.: Ontology-driven collaborative annotation in shared workspaces. Future Gener. Comput. Syst. Spec. Issue Semant. Technol. Collab. Web **54**, 435–449 (2016)
35. Goy, A., Magro, D., Petrone, G., Picardi, C., Segnan, M.: Shared and personal views on collaborative semantic tables. In: Molli, P., Breslin, John, G., Vidal, M.-E. (eds.) SWCS 2013-2014. LNCS, vol. 9507, pp. 13–32. Springer, Heidelberg (2016). doi:10.1007/978-3-319-32667-2_2
36. Goy, A., Magro, D., Petrone, M., Rovera, C., Segnan, M.: A semantic framework to enrich collaborative semantic tables with domain knowledge. In: Proceedings of IC3K 2015, KMIS, vol. 3, pp. 371–381. SciTePress (2015)
37. Zhao, L., Ichise, R.: Ontology integration for linked data. J. Data Semant. **3**(4), 237–254 (2014)