

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)  
<http://www.disi.unitn.it>

**ADOPTING SEMANTIC  
ANNOTATION SYSTEMS FOR  
CORPORATE PORTAL  
MANAGEMENT: A TELEFONICA  
CASE STUDY**

Ilya Zaihrayeu, German Toro del Valle,  
Juan Pane and Pierre Andrews

December 2010

Technical Report # DISI-10-065



# Adopting Semantic Annotation Systems for Corporate Portal Management: a Telefónica Case Study

Ilya Zaihrayeu<sup>1</sup>, Germán Toro del Valle<sup>2</sup>, Juan Pane<sup>1</sup>, and Pierre Andrews<sup>1</sup>

<sup>1</sup> DISI - University of Trento, Italy  
{ilya, pane, andrews}@disi.unitn.it

<sup>2</sup> Telefónica Investigación y Desarrollo, Spain  
gtv@tid.es

**Abstract.** Corporate portals make an important integral part of the enterprise infrastructure, facilitating the creation, sharing, discovery and consumption of enterprise assets through blogs, news, forums, documents and information in general. However, as the amount of data grows, it becomes much more difficult to access the right asset in the precise moment when it is needed. Annotation systems tried to address this problem to a certain extent by allowing the users to collaboratively annotate assets using tags such that these assets can be found more easily by reusing these tags in queries. However, this model often falls short due to mismatches in the vocabularies of different users, due to the use of synonymous, polysemous, more specific or general terms in tagging and searching, and so on. In this paper we: (a) provide an analysis of the corporate portal of the Telefónica group; (b) identify requirements for a semantics-enabled annotation system that is capable of addressing the above-mentioned shortcomings; (c) define a semantic annotation model that meets the requirements; (d) provide the details of the implementation of the annotation model for the Telefónica portal; and (e) provide an initial evaluation from the use of the semantic annotation system in the enterprise.

## 1 Introduction

Social annotation systems such as Delicious<sup>3</sup>, Flickr<sup>4</sup> and others have laid the fundamentals of the Web 2.0 principles and gained tremendous popularity among Web users. One of the factors of success for these systems is the simplicity of the underlying model, which consists of a resource (e.g., a web page), a tag (e.g., a text string), and a user who annotated the resource with the tag. Despite its simplicity, the annotation model enables a set of useful services for the end user, e.g., searching resources using tags added by a community of users, computing the most popular tags and building the so-called tag clouds<sup>5</sup>, finding users with

---

<sup>3</sup> <http://www.delicious.com/>

<sup>4</sup> <http://www.flickr.com/>

<sup>5</sup> [http://en.wikipedia.org/wiki/Tag\\_cloud](http://en.wikipedia.org/wiki/Tag_cloud)

common interests based on the resources they annotated and on the tags they used and providing a recommendation service on this basis, among others.

Due to the syntactic nature of the underlying model, these systems have been criticized for not being able to take into account the explicit information about the meaning or semantics of each tag. For example, different users can use the same tag with different meanings (i.e., homonyms), different tags with the same meaning (i.e., synonyms), different tags in the same meaning but at different levels of abstraction, morphological variations of the same tag, and so on.

The Semantic Web community proposed to address the aforementioned and related problems by introducing formal models, languages, and methods for representing and manipulating knowledge and data in a machine-processable way. The approach can leverage these formal semantics to provide the user with services of a new qualitative level when computers can “understand” the meaning of data and of user inputs and interact with the user in an intelligent manner. However, for this to happen, a critical mass of content with formal semantics has to be available to enable these services [1].

It has been shown that the automatic extraction of semantics from textual content can hardly be done with an acceptable level of accuracy mainly due to the problem of ambiguity of natural language [2]. In annotation systems this problem can be even harder because user input consists of single words or short phrases that lack contextual information that is present in full-fledged sentences and that is used by natural language processing tools for the disambiguation [3]. On the other hand, it is relatively easy for the user to understand the semantics of an annotation given the contents of the resource and its other annotations. However, in order to get semantic annotations from the user, at least two problems need to be addressed: (1) how to motivate the user to do some extra work before the critical mass is generated and, therefore, the user can see the added value from the invested effort; and (2) how to map the user input to the underlying formal models without requiring the user to learn these models and enabling technologies that would require gaining a non-trivial expertise.

In this article we provide an analysis of the requirements for a semantic annotation model that can enable semantics-aware services in the intranet corporate portal of *Telefónica Investigación y Desarrollo*<sup>6</sup> (Section 2). Based on the requirements, we develop a semantic annotation model and describe its properties in detail (Section 3). In Section 4 we show how the model was implemented within the INSEMTIVES platform – a semantic content management platform being developed in the context of the 7th Framework Programme European Union funded project INSEMTIVES<sup>7</sup>. We discuss how the platform and the implemented model was used from inside the Telefónica’s intranet corporate portal to enable enhanced searching, navigation and consumption of the information and services exposed by the portal. Finally, we provide first feedback and evalu-

---

<sup>6</sup> <http://www.tid.es/en>

<sup>7</sup> One of the goals of the INSEMTIVES project is to combine human and computer intelligence in the task of generating semantic contents by providing technological support and incentives for the end users. For more, see <http://www.insemtives.eu>

ation from the adoption and use of the semantic annotation system in *Telefónica Investigación y Desarrollo*. Section 5 summarises the findings and concludes the paper.

## 2 The *Telefónica Investigación y Desarrollo*'s Intranet Corporate Portal: Use Case

Telefónica Investigación y Desarrollo<sup>8</sup> (Telefónica I+D for brevity) is the innovation company of the Telefónica Group<sup>9</sup>. Founded in 1988, it contributes to the Group's competitiveness and modernity through technological innovation. To achieve this aim, the company applies new ideas, concepts and practices in addition to developing advanced products and services.

Over the last few years and within the global market, Telefónica I+D has grown to become a network of centres of technological excellence that stretches far beyond the Spanish borders, extending its R+D activities to offices located in Spain (Barcelona, Granada, Huesca, Madrid and Valladolid), Brazil (São Paulo) and Mexico (Mexico D.F.). Currently, Telefónica I+D employs more than 1200 employees shared out amongst its 7 centres.

### 2.1 The Intranet Corporate Portal

The Telefónica I+D's intranet corporate portal is a live, 24x7, highly active Web portal used by every single employee at the company. To access the portal and all its information and services, the employees have to authenticate themselves using a user name and a password provided by the company. This authentication mechanism allows the company not only to control the access to the information and services but also to audit the use of the portal made by its employees.

Since its first days back in 1998, the intranet corporate portal has become one of the main communication means available at the company. Although it has experimented a noticeable evolution during the last years, including new sources of information, services and technologies (mainly Web 2.0 technologies), it has always been loyal to its early principles as an open and collaborative environment of contact, discussion and information for all the company employees.

The entrance door to the portal is the portal home page (depicted in Figure 1). This home page offers access to all the sources of information and services, including infrastructure management, business management, knowledge management and human resources management information, services and tools.

In its current status, the portal integrates a huge amount of heterogeneous sources of information and services provided by distinct informational systems

---

<sup>8</sup> <http://telefonica.com/en/home/jsp/home.jsp>

<sup>9</sup> The Telefónica Group is a Spanish multinational group of companies operating in the Information Technology and telecommunication domain. The Group stands in the 5th position in the telecommunication sector worldwide in terms of market capitalization, the 1st as an European integrated operator and the 3rd in the Eurostoxx 50 ranking, composed of the major companies in Europe (June 30th 2010).



Fig. 1. The Telefónica I+D's intranet corporate portal home page.

of the company. This heterogeneous environment makes it necessary to use distinct software technologies and programming languages throughout the length and breadth of the portal. In fact, the portal is the result of a considerable technological integration effort where disparate technologies like, for example, Java Server Faces (JSF), Java Server Pages (JSP), HTML, CSS, Javascript, Java applets, Flash, CMS solutions, blogging platforms, etc., coexist.

## 2.2 Problem Statement

Regardless of the enormous usefulness of the information and services exposed by the Telefónica I+D's intranet corporate portal for the company employees, its tremendous growth in the last years has made it very difficult to exploit all the capabilities that the portal provides. Although the portal provides traditional search engine and syndication capabilities, every day corporate news, blog and forum entries, documents and information in general are added to the portal, being almost impossible to be kept up with all of them. In this same sense, it becomes much more difficult to access the right asset in the precise moment when it is needed. This is the reason why new and innovative ways to annotate, retrieve, navigate and, basically, consume all this information have to be provided, that go beyond the capabilities of the conventional annotation and retrieval systems.

### 2.3 Requirements Analysis

In this section we report on the requirements for the annotation model to be used in the Telefónica I+D use case. These requirements were developed from the results of the semi-structured interview reported in [4] that resulted in a generic model that suits the requirements of Telefónica I+D. The requirement analysis phase was structured according to a mutually independent set of features selected based on the intention to demonstrate how much they can contribute to the definition of a semantic annotation model that would enable enhanced search and navigation capabilities in the portal. The analysed features were grouped in the following three categories:

- *structural complexity*: the level structural complexity of annotations (e.g., tags, attributes, and relations), and the granularity of the annotation (i.e., *parts* or *wholes*);
- *vocabulary type*: the level of formality of the annotations defined on the basis of the form of the underlying vocabulary (e.g., free text, thesauri, taxonomies, ontologies) to which the annotations can be linked; and
- *collaboration type*: the level of user collaboration in sharing and reusing semantic annotations (e.g., private annotations, public annotations) and the level of collaboration in the construction and evolution of the underlying vocabulary (static vocabularies vs. collaboratively maintained ones).

The full specification of each of the categories is provided in [4]. It presents a comparative analysis of the advantages and disadvantages of each feature, with examples of popular systems (both in the research field and among publicly available systems), showing also the trade off between the level of user involvement and provided services. In the following we give examples of typical usages of annotations in the Telefónica portal and specify the identified requirements for each of the three above described categories.

The Telefónica I+D's use case will provide an annotation platform for the users of its intranet (i.e., the company employees). The users will be able to annotate any content exposed on the portal including text (words, sentences and paragraphs), images and areas of images, videos and sections of videos. All these contents are exposed in the shape of corporate news, blog, wiki and forum entries, document repositories, etc., as mentioned in Section 2.1. In what follows we summarise the findings which led to Telefónica specific requirements and group them following the previously mentioned set of features.

**Structural Complexity** The Telefónica I+D use case requires a mixed complexity in the annotation model. The conducted interviews showed that users were interested in an *easy and quick* annotation procedure with a *simple tagging interface*, i.e., the effort for annotating content should be minimal and as seamless as possible.

The annotation model should be able to support *tags* which are not associated to any particular property of the resource, but also *attributes* to describe

specific metadata and *relations* between (and within) resources that can be used for navigation. The annotations should be applicable to whole resources (e.g., blog post, image videos), as well as to specific parts of them (e.g., a particular paragraph, a part of the video, a frame of an image).

When using annotations in search and navigation, the users should be able to *filter-out* search results by using attributes (e.g., the author, a date range).

**Vocabulary Type** The annotation model should not be limited to a domain specific controlled vocabulary and users should also be able to enter free-text annotations. However, the free-text annotation should be mapped, whenever possible, to a general purpose controlled vocabulary, such as Wordnet [5] or DBPedia [6]. This vocabulary should be then extended by the users when providing new terms and concepts in their free-text annotations via consensus and evolution algorithms. The former algorithm finds new entries for the controlled vocabulary by computing a consensus on the meaning of free-text annotations made by different users, whereas the latter algorithm recomputes the mappings from annotations to controlled vocabulary entries as the vocabulary evolves (interested readers are referred to [7–9] for details).

**Collaboration Type** Telefónica I+D has shared resources that users can access, annotate, and search for. With the exception of some resources where there is a need for access control, most of the resources and annotations are available to all users. There is also a need for traceability of the annotations for auditing. The collaborative annotation model thus should be used and embed provenance information (when, who) as well as versioning information. In addition, the controlled vocabulary will be collaboratively extended to capture the dynamic knowledge that evolves within the Telefónica I+D environment.

### 3 Semantic Annotation Models for Telefónica I+D

The purpose of the annotation model developed in the INSEMTIVES project is to represent and seamlessly combine free-text and semantic annotations which describe a (part of a) resource. An in-depth study of the state-of-the-art in the field was conducted and reported in [4]. In this section we present a summary of the model that was built from the analysis presented in [4] and presented in [10]. The elements presented in this section are the most relevant and necessary for the Telefónica I+D use case as defined in the section on requirements (Section 2.3).

#### 3.1 Generic Objects

The generic semantic annotation model defines the following objects: *user*, *resource*, *term* and *controlled vocabulary*. Their relationship is visualized in Figure 2 and their descriptions are provided in the following:



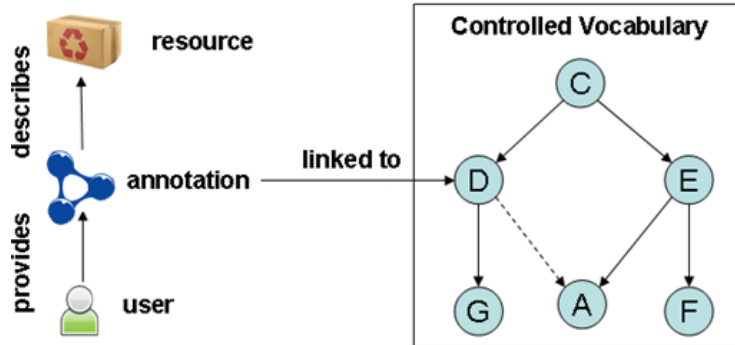


Fig. 2. Generic annotation model with semantics

A **user**  $u$  is a tuple  $u = \langle id, name \rangle$ , where  $id$  is a unique identifier of the user; and  $name$  is the name of the user. We write  $U$  to denote the set of all users.

A **resource**  $r$  is a triple  $r = \langle id, name[, part] \rangle$ , where  $id$  is a unique identifier of the resource and  $part$  is an optional identifier of a part of the resource. When  $part$  is absent,  $name$  represents the name of the resource and it represents the name of the resource part otherwise. We write  $R$  to denote the set of all resources.

The Telefónica I+D use case requires a granular annotation of resources where also parts of the resource (e.g. a paragraph in a textual document, a part of an image) can be described by an annotation. Therefore, a **part** can take multiple forms. A textual segment is part of a textual document and is defined by a tuple:  $part^t = \langle beginning, end \rangle$  where  $beginning$  and  $end$  represent the number of characters from the beginning of the document to the start of the segment and the end of the segment, respectively. An image segment is part of an image and is defined by a singleton:  $part^i = \langle region \rangle$  where  $region$  denotes a geometrical region of the image. A video segment is part of a video and is defined by a triple:  $part^v = \langle beginning, end, region \rangle$  where  $beginning$  and  $end$  represent the number of seconds from the beginning of the video to the start of the segment and the end of the segment, respectively; and  $region$  denotes a static geometrical region of the video during this segment of time.

A **term**  $t$  is a non-empty finite sequence of characters. We write  $T$  to denote the set of all possible terms. Normally, terms represent natural language words such as “sea”, “bird”, “location” or “semantic web”. A **controlled term**  $ct$  is a tuple  $ct = \langle t, \{c\} \rangle$  where  $t$  is a term and  $c$  is a concept in a controlled vocabulary  $CV$  that represents each token (or word) in the term  $t$ .

The **controlled vocabulary**  $CV$  follows the SKOS [11] model. This means that whenever possible, a SKOS concept  $c$  should be used to convert the term  $t$  into a controlled term  $ct$  in order to enrich the annotation with semantics. The controlled vocabulary can be domain dependent or general purpose, for example Wordnet [5] or other semantic sources such as Yago [12] or DBPedia [6] or any

other SKOS complaint source, for example, those available in Linked Data<sup>10</sup>. The Telefónica I+D use case, however, will initially use Wordnet as initial *CV* and will then be extended by the users via consensus and evolution mechanisms that were described in Section 2.3. Note that in all the cases the term *t* can take the form of free text belonging to an “open” vocabulary and can also be linked to a SKOS concept belonging to a controlled vocabulary, satisfying the Telefónica I+D requirements presented in Section 2.3.

### 3.2 Annotation elements

The annotation element can take the form of tags, attributes and relations. An annotation *ann* is a tuple  $ann = \langle r, ann^x, u, \{he\} \rangle$ , where:

- *r* is the resource being annotated;
- $ann^x$  can take the form of:
  - tag annotation  $ann^{tag} = \langle t, [c, ] \rangle$ , where *t* is the term used for the annotation ( $t \in T$ ); *c* is a concept that represents *t* and belongs to a *CV*;
  - attribute annotation  $ann^{attr} = \langle an, [can, ]av, [cav] \rangle$ , where *an* is a term denoting the attribute name ( $an \in T$ ); *can* is a *CV* concept that represents the attribute value which can belong to any of the primitive data types (e.g., dates, strings) and *cav* is a *CV* concept that represents the attribute value whenever *av* can be linked to a concept;
  - relational annotation  $ann^{rel} = \langle rel, [crel, ]tr \rangle$ , where *tr* is the target resource (i.e., the resource used as an annotation object); *rel* is a term that denotes the name of the relation that exist between *r* and *tr*; *crel* is a SKOS concept that represents *rel* in *CV*;
- $u \in U$  is the user that performed the annotation;
- A history element defined as a 4-tuple:  $he = \langle action[, ov], ts[, \alpha] \rangle$ .
  - *action* is the type of action that was performed  $action \in \{\text{added, removed, updated\_name, updated\_value, updated\_source, updated\_target}\}$ ,
  - *ov*, when applicable, is the value of the changed annotation before *action* was performed;
  - *ts* is the timestamp when this operation was performed; and
  - $\alpha$  provides the accuracy of the algorithm that created the annotation when the annotation is added by a (semi-)automatic algorithm.

The provenance information stored in  $\{he\}$  could be simplified by replacing it with  $u, ts[, \alpha]$  which would remove the versioning capabilities of the model still capturing some minimal auditing capabilities. Because of the centralized organization of the Telefónica I+D use case, there is no issue of synchronization, consistency checking and merging branches in the versioning system.

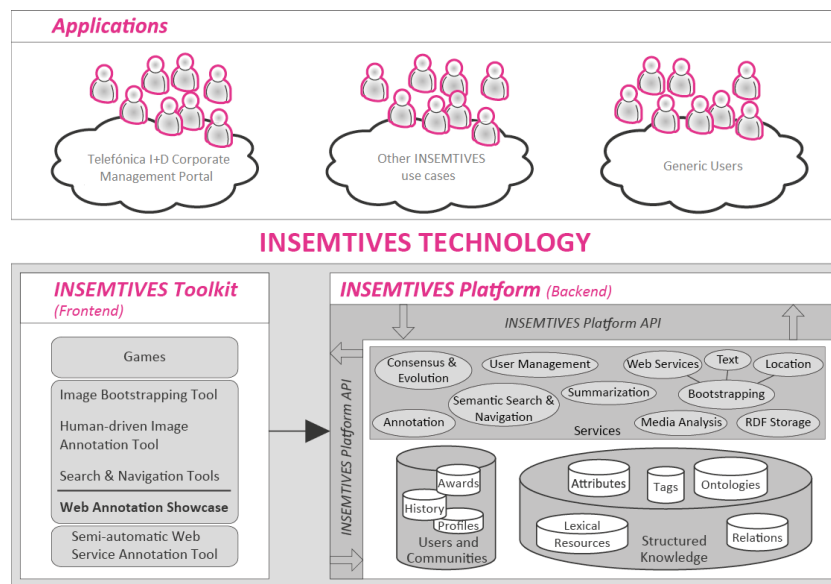
## 4 Implementation

The semantic annotation tool for the Telefónica Corporate Management Portal was developed as a browser extension. In the following sections we will outline the enabling platform as well as the annotation tool.

<sup>10</sup> Linked Data, see <http://linkeddata.org/>

#### 4.1 Enabling Annotation Platform

The Annotation tool of Telefónica I+D Corporate Management Portal is built as an application on top of the INSEMTIVES platform. The architecture of the INSEMTIVES project and its environment is depicted on Figure 3. It includes the application use cases, the toolkit as well as the set of services to be available through the platform API. In this section we will briefly outline the INSEMTIVES platform. Interested readers are referred to [13] for details on the INSEMTIVES *toolkit* and other application use cases. The main components of the INSEMTIVES platform are:



**Fig. 3.** Architecture of the INSEMTIVES project.

*Platform API* The platform services are exposed through the INSEMTIVES platform API's, which are used both for by the toolkits and third party applications. The API's are based on a communication framework which represents a highly modular client-server architecture implemented using communication layer over JMS<sup>11</sup> messaging protocol. Application developers can rely on available client implementations such as ActiveMQ<sup>12</sup> cross-language which provides implementations for 16 languages. Also, it supports 6 protocols including REST<sup>13</sup> and Web Service Notification.

<sup>11</sup> JMS - a Java Messaging Service, see <http://java.sun.com/products/jms/>

<sup>12</sup> ActiveMQ - Apache Integration Provider, see <http://activemq.apache.org/>

<sup>13</sup> REST, see [http://en.wikipedia.org/wiki/Representational\\_State\\_Transfer](http://en.wikipedia.org/wiki/Representational_State_Transfer)

*Structured Knowledge* The structured knowledge stores all artifacts in the form of RDF following the semantic annotation model sketched in Section 3 and defined in [9]. The semantic store relies on OWLIM<sup>14</sup> while the initial Lexical resource is based on Wordnet and DBPedia.

*User and Communities* This module contains user profiles and auditing information and is the responsible for maintaining provenance information of the annotations and resources in the platform. It is also responsible for the ranking of users and supporting the diverse incentive related data.

*Services* The services are implemented on top of the structured knowledge exploiting also the *User and Communities* module. Some relevant services are:<sup>15</sup>

- Annotation: This services manages all the annotations implementing the model sketched in Section 3 and defined in [9]. It stores all the tags ( $ann^{tag}$ ), attributes ( $ann^{attr}$ ) and relations ( $ann^{rel}$ ) annotations in the Structured Knowledge module;
- Bootstrapping: creates a initial set of semantic annotations for new resources using contextual information, e.g., when uploading pictures, it uses the folder structure where the picture resides and the GPS coordinates from the EXIF metadata to suggest semantic annotations; and
- Consensus & Evolution: manages the semi-automatic evolution of the controlled vocabulary and annotations as described in Section 2.3.

## 4.2 Annotation Tool

The INSEMTIVES platform provides all the *backend* services needed to build the annotation tool (see Section 4.1). The annotation tool user interface (UI) was developed to allow users to interact with this back-end. All the interaction between the annotation tools and the platform is managed by means of an HTTP endpoint following the REST paradigm [14] using the *Communication Framework* outlined in Section 4.1.

The annotation tool adds a semantic overlay layer on top of the Telefónica I+D's intranet portal. Since it is a live portal used on the 24x7 basis by the company employees, a major requirement for the integration of the tool is its inclusion in a non-intrusive way, this is, not affecting the current users of the portal. To accomplish this requirement, the annotation tool has been implemented as a Google Chrome<sup>16</sup> extension<sup>17</sup>. By doing so, we have been able to implement and test the annotation tool on the latest version of the portal without affecting its current users. On the other hand, the use of Javascript injection techniques<sup>18</sup> allows us to embed the annotation tool directly into the portal at any later extent if desired. Another major consequence of this design decision is

<sup>14</sup> OWLIM, see <http://www.ontotext.com/owlim/>

<sup>15</sup> For the complete reference of services please refer to [13].

<sup>16</sup> <http://www.google.com/chrome/>

<sup>17</sup> <https://chrome.google.com/extensions/>

<sup>18</sup> Javascript injection is a technique which let developers to insert Javascript code into Web pages once they are loaded by an Internet browser. This technique was used

the fact that the annotation tool is not anchored at any means to the Telefónica I+D’s portal, allowing us to apply and to use it on top of any Web site.

Once the annotation tool extension has been downloaded<sup>19</sup> and installed<sup>20</sup> in the Google Chrome browser, users can navigate to any of the pages of the portal. Thanks to the extension, the pages are modified on the fly (using the Javascript injection technique previously mentioned) to include a new INSEMTIVES project logo right on the page the user is visiting (as depicted in Figure 1). When the logo is clicked, the annotation tool side-bar expands as depicted in Figure 4. The side-bar is an overlay layer on top of the portal providing the user with semantic annotation capabilities, although the portal has not been modified.

The annotation tool side-bar provides users with six main accordion menus or areas, these are: “Home”, “Annotate”, “Navigate”, “Search”, “Configure” and “Help”. The first version of the annotation tool includes functionality for the “Home” and “Annotation” areas. We are currently working on developing the rest of the areas to be release in future versions of the annotation tool<sup>21</sup>.

The “Home” area (see Figure 4 a)) includes three modules:

- “*User*” module: this module lets the user manage her personal and contact information and is based on the “Users and communities” module of the INSEMTIVES platform (see Section 4.1);
- “*Ranking & statistics*” module: this module is an example showing how incentives could be included into the annotation tool. In this particular case, we exploit the competition incentive (letting the users of the tool to compete based on the annotations they made, the more annotations the user made the more points the user gets, for example) as well as the reward incentive, rewarding the top annotator with concrete prizes or money; and
- “*My knowledge cloud*” module: this module is another example of how easily incentives could be included into the annotation tool thanks to its modular design. In this particular case, we exploit the reputation incentive including a cloud with the tags ( $ann^{tag}$ ), attributes ( $ann^{attr}$ ) and relations ( $ann^{rel}$ ) that a user has used the most in her annotations and which somehow make her an expert on these topics. Bigger fonts means more frequently used annotations.

The second implemented area in the current version of the annotation tool is the “Annotate” area (see Figure 4 b)). This area allow users to annotate content on the Web page they are visiting. Currently, it includes three main modules:

---

to modify the behavior of the Telefónica I+D’s intranet corporate portal to include the annotation tool without the need to modify or to affect the portal.

<sup>19</sup> The current version of the annotation tool extension can be downloaded from [http://sourceforge.net/projects/insemtives/files/OKenterprise\\_extension\\_for\\_Google\\_Chrome/OKenterprise\\_extension\\_for\\_Google\\_Chrome-v0.3.zip/download](http://sourceforge.net/projects/insemtives/files/OKenterprise_extension_for_Google_Chrome/OKenterprise_extension_for_Google_Chrome-v0.3.zip/download)

<sup>20</sup> Directions to properly install the annotation tool extension can be found at [http://www.insemtives.eu/deliverables/Insemtives\\_d5.2.1.pdf](http://www.insemtives.eu/deliverables/Insemtives_d5.2.1.pdf).

<sup>21</sup> The final version of the annotation tool will be released on September 2011. It will be available at: <http://sourceforge.net/projects/insemtives/>.

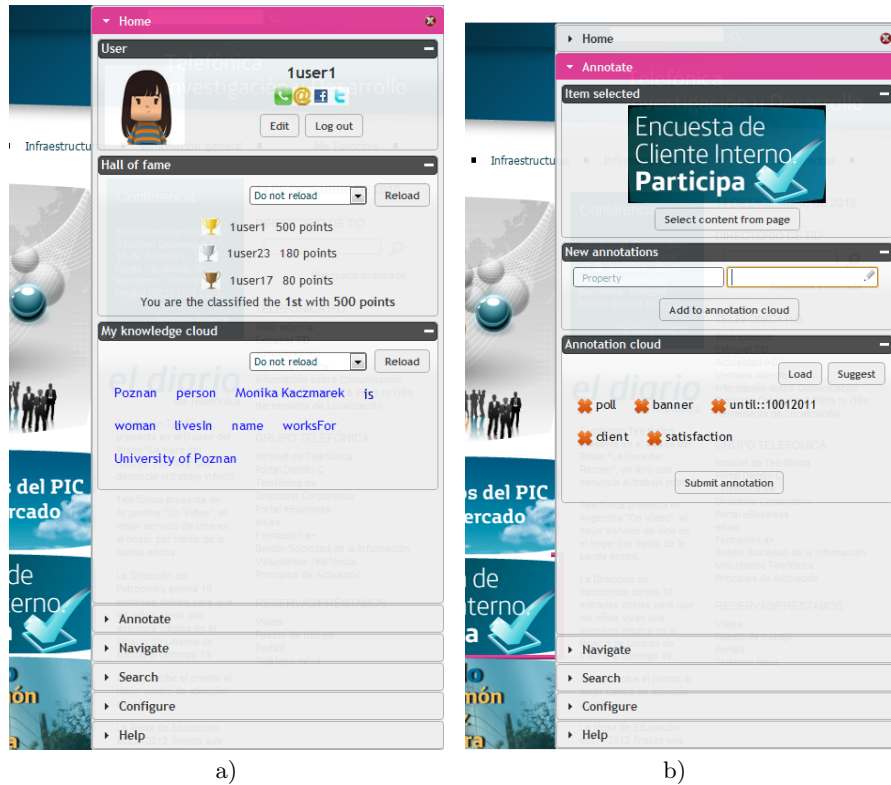


Fig. 4. The INSEMTIVES annotation tool side-bar. a)The “Home” area. b) The “Annotate” area including an image ready to be annotated.

- “*Item selected*” module: this module allows the user select content from the Web page being annotated or see the selected content. Notice that in case if no content is selected, the whole Web page will be annotated. When selecting content from a web page, the annotation tool side-bar collapses letting the user to select any content on the Web page (text, images and videos, mainly). Once the content is selected, the side-bar expands showing the selected content, allowing the user to fine-tune her selection (e.g., select areas of images or sections of the selected video she wants to annotate). Thus the requirement of annotating the whole resource and also parts of it are supported by this functionality (see Section2.3);
- “*New annotations*” module: this module allows users to add concrete annotations to the selected content or to the whole Web page. Users can add tags ( $ann^{tag}$ ), attributes ( $ann^{attr}$ ) and relations ( $ann^{rel}$ ) as defined in Section 3 that will then be stored in the *Structured Knowledge* module of the INSEMTIVES platform (see Section 4.1) using RDF. The input fields offer the user an autocompletion feature suggesting concepts as the user types, based on previous entries and the controlled vocabularies *CV* managed by

the platform. If the input value is not present in the *CV*, then the annotation is stored in the free-text form; and

- “*Annotation knowledge cloud*” module: this module includes the annotations associated to the selected content or to the whole Web page. The user can remove particular annotations directly from the “Annotation knowledge cloud” and can also “Load” annotations made by other users for the same content. The “Suggest” button invokes the bootstrapping service of the IN-SEMTIVES platform to suggest possible annotations (see Section 4.1).

### 4.3 Initial Evaluation

The specification, design and implementation of the annotation tool has followed Agile<sup>22</sup> methodology guidelines, involving the potential final users of the annotation tool (i.e., the Telefónica I+D employees) in the process from the very beginning through in-person interviews. This user-centered approach has let us focus on concrete and real problems the users of the Telefónica I+D’s intranet corporate portal face on their daily work as well as increase the adequacy of the annotation tool to satisfy the potential user real needs.

An in-place evaluation will be carried out at the Telefónica I+D’s facilities on the first quarter of 2011 involving real users to validate the semantic annotation model, the platform and the different incentives models along with the annotation tool being developed. This evaluation is scheduled after the new features listed in Section 4.4 are implemented.

### 4.4 Future Work

In the future, we plan to extend the current work with new capabilities and functionalities. In particular, we plan to:

- include the “Search” and “Navigate” functionalities to let users semantically search and navigate the annotated content;
- include social capabilities into the annotation tool to allow Telefónica’s users interact not only among them but also with users of other online communities by means of social networks such as Facebook<sup>23</sup>, Twitter<sup>24</sup>, LinkedIn<sup>25</sup> and Yammer<sup>26</sup>;
- extend the annotation tool including capabilities to find experts on specific topics from the annotations made by each user (the annotations made by the users somehow characterize them, their topics of interest and expertise);
- include quick paths (i.e., pop-up menus entry when right-clicking on an image, video or any content susceptible of being annotated) as well as keyboard shortcuts to speed up the annotation process; and

<sup>22</sup> <http://www.agilealliance.org/>

<sup>23</sup> <http://www.facebook.com>

<sup>24</sup> <http://www.twitter.com>

<sup>25</sup> <http://www.linkedin.com>

<sup>26</sup> <http://www.yammer.com>

- convert the knowledge cloud to the semantic knowledge cloud, where each item in the cloud represents a concept, e.g., synonymous annotations such as “car” and “automobile” will be represented by one entry in the cloud and their frequencies merged; and polysemous annotations such as “java (island)” and “java (coffee beverage)” will be represented by different entries in the cloud with their frequencies split (for a more complete reference see [7]).

## 5 Conclusion

In this paper we briefly described how knowledge and data are created, shared, discovered, and consumed in the corporate portal of Telefónica Investigación y Desarrollo – one of the biggest telecommunication provider worldwide. As we pointed out, with the constant growth of contents on the portal, conventional content management tools including annotation and search become less and less effective and often fall short of providing the right asset in the precise moment when it is needed. In order to compensate this shortcoming, we reported on the results of in-depth requirement analysis of an annotation and retrieval system in which semantics play the central role. We then presented a semantic annotation model that meets the identified requirements and reported on the details of its implementation on top of the Telefónica’s corporate portal. Although more work is required to make final conclusions from the introduction of this semantic annotation system for corporate portal management, the initial evaluations are promising as they show that such a system can enable enhanced searching, navigation and consumption of the information and services exposed by the portal.

## Acknowledgements

This work has been partly supported by the INSEMTIVES project (FP7-231181, see <http://www.insemtives.eu>). We would like to thank the members of the INSEMTIVES consortium for their support and collaboration.

## References

1. Hendler, J.: Agents and the semantic web. *IEEE Intelligent Systems* (2) (2001)
2. Navigli, R.: Word sense disambiguation: A survey. *ACM Comput. Surv.* **41** (February 2009) 10:1–10:69
3. Zaihrayeu, I., Sun, L., Giunchiglia, F., Pan, W., Ju, Q., Chi, M., Huang, X.: From web directories to ontologies: Natural language processing challenges. In: *ISWC/ASWC*. (2007) 623–636
4. Bürger, T., Zaihrayeu, I., Andrews, P., Babenko, D., Pane, J., Popov, B.: Report on the state-of-the-art and requirements for annotation representation models. *INSEMTIVES Deliverable 2.1.1*; <http://www.insemtives.eu> (June 2009)
5. Fellbaum, C.: *WordNet: An Electronic Lexical Database*. Bradford Books (1998)
6. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: *DBpedia - A crystallization point for the Web of Data*. *Web Semantics: Science, Services and Agents on the World Wide Web* **7**(3) (Sept. 2009) 154–165



7. Andrews, P., Pane, J., Zaihrayeu, I.: Report on methods and algorithms for bootstrapping semantic web content from user repositories and reaching consensus on the use of semantics. INSEMTIVES Deliverable 2.2.1; <http://www.insemtives.eu> (Nov. 2009)
8. Bürger, T., Morozova, O., Andrews, P., Zaihrayeu, I., Pane, J.: Report on methods and algorithms for linking user-generated semantic annotations to semantic web and supporting their evolution in time. INSEMTIVES Deliverable 2.2.2; <http://www.insemtives.eu> (Nov. 2009)
9. Andrews, P., Zaihrayeu, I., Pane, J., Autayeu, A., Nozhchev, M.: Report on the refinement of the proposed models, methods and semantic search. INSEMTIVES Deliverable 2.4; <http://www.insemtives.eu> (Nov. 2010)
10. Bürger, T., Morozova, O., Zaihrayeu, I., Andrews, P., Pane, J.: Specification of models for representing single-user and community-based annotations of web resources. INSEMTIVES Deliverable 2.1.2; <http://www.insemtives.eu> (Sept. 2009)
11. Miles, A., Pérez-Agüera, J.R.: SKOS: Simple Knowledge Organisation for the Web. *Cataloging & Classification Quarterly* **43**(3) (2007) 69–83
12. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: A Core of Semantic Knowledge. In: 16th international World Wide Web conference (WWW 2007), New York, NY, USA, ACM Press (2007)
13. Siorpaes, K., Konstantinov, M., Popov, B.: Requirement analysis and architectural design of semantic content management platform. INSEMTIVES Deliverable 3.1; <http://www.insemtives.eu> (Sept. 2009)
14. Fielding, R.T.: Architectural styles and the design of network-based software architectures. Doctoral dissertation; <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm> (2000)