A Concept Recommender Service on The Federated Cloud of the PPP European Project

Boris Moltchanov[1], Oscar Rodriguez Rocha[2]
[1]Telecom Italia, via G. Reiss Romoli, 274, 10148.
[2]Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129.
Turin, Italy.
boris.moltchanov@telecomitalia.it, oscar.rodriguezrocha@polito.it

Abstract: *Conventionally, the Web Service [1], respecting the Client-Server paradigm, consists of a software system or a platform running as a server somewhere in the Internet allowing its customers to benefit from it through the use of a client. Cloud computing [2] (which has gained a lot of popularity in recent years), allows to have this software system or platform deployed and running in a flexible and robust environment where traffic demand is automatically handled and the fault-tolerance and high-availability are provided by the cloud tools and means. However, currently existing private and public cloud models comprise cloud platforms installed on IT facilities located in certain geographical areas. Thus, the final customer location is not considered and the service traffic is not balanced geographically. This is a serious drawback of the existing technology as it creates an overhead of the network (Internet) traffic for the service when the customers use it from different geographical areas. Moreover this increases service delays as the packets in the request-response communication paradigm cross long Internet cross-countries connections and thus delayed. The new concept proposed in this paper is feasible due to the EU attempt to instantiate a number of the public clouds interconnected into a cloud federation over Europe, where the deploying of services in a geographically distributed cloud environment will enable to serve customers in a more efficient and faster way. The location of the customer connected through the network (Internet) will define to which service instance in the federated cloud, its requests will be sent to. This drastically reduces service traffic, and so the costs, over Internet for international traveling customers, makes the service reaction faster as it balances the service traffic over mode nodes and makes significantly more robust overall service solution because of the service geographical redundancy. The recommender service described in this paper is deployed over few nodes of the federated clouds build and interconnected as result of the EU funded PPP XIFI project[3].*

Keywords: Concept Recommender service, Federated Cloud, Web Service, Internet Traffic, European Projects

Recieved: 18 December 2014, Revised 13 January 2015, Accepted 17 January 2015

1. Introduction

The work presented in this paper describes a Web Service (a FIWARE Generic Enabler), deployed in a federated cloud environment and configured in such a way that the service requests received from customers are served by a service instance, which is geographically closest to the customer. The service itself is a Concept Recommender based on Linked Data [4].

Given an input text, the service performs a text analysis process (described in VII) to extract the nouns present and by querying different Linked Data sources, it finds associations between each of these nouns and the corresponding concepts or objects present in the Web of Data (using their URIs). This means that after this process, the original text is annotated semantically. Finally, for each of the semantic annotations found in the previous step, the service computes their semantic relations to generate a list of related concepts that will be the end result or output.

The paper is organized as follows: the PPP XIFI Project itself is introduced and described in the Section II. Section IV describes the federated cloud where the service is deployed. The Recommender Service is detailed in the Section IV. The modules that make up the service are described in Sections V, VI and VII. The Conclusions and the roadmap are presented together in Section VIII, describing the future work within the EU funded initiative and underlying the relevance of this work for a modern future Internet and information enabled society.

2. PPP XIFI Project

In 2010, the EU Commission founded the Future Internet Public Private Partner initiative [5][1]. It resulted into several different projects including the XIFI project [3], which is responsible for the capacity building part of the PPP programme. XIFI paves the way for the establishment of a common European market for large-scale trials for Future Internet and Smart Cities through the creation of a sustainable pan-European federation of Future Internet test infrastructures. The XIFI open federation leverages existing public investments in advanced infrastructures and support advanced large-scale deployment of FI-PPP early trials across a multiplicity of heterogeneous environments and sector use cases that should be sustained beyond the FI-PPP programme. As a result of the XIFI project many national public clouds have been created in different countries (14 clouds providers) and interconnected in order to facilitate the prototyping and trials of services, by using the outcomes of the FIWARE project [5] (more specifically, its Generic Enablers available). Many use-case services are currently deployed and running in the federated cloud infrastructure created by XIFI project. One of this services is the Federated Recommender Service, that has been currently deployed in two federated cloud nodes, one in Spain and the other in Italy.

3. Distributed Service Concept

Depending on the resource usage, the XIFI cloud infrastructure is able to adapt by consuming available resources from other clouds based on the same XIFI federation mechanism. The clouds within the XIFI federation are controlled by their respective owners and operators that can adjust the limitations. Federating different clouds, belonging to different owners and based on different technologies, makes possible to create a more powerful federated cloud that makes available a greater number resources. The conventional cloud concept is extended in the XIFI project, putting all the national public clouds created as XIFI results trials into a federation.

As schematically shown in the Figure 1, in the XIFI cloud infrastructure both the developer (or provider) of the service and the end-users can access to the federated service without knowing in which machine and/or where it has been deployed. In order to dispatch the network traffic to the right node, there is a federation manager and a load-balancer. Therefore, the resources available to the developer are seamlessly infinite as they are taken from available public clouds. Additionally the end-user can connect to the service node located closer to her/him without any difference from the service usage perspective and obtaining the best user experience.

The recommender service described in this paper, is currently deployed in two federated cloud nodes, one in Spain and the other in Italy. In general, the targeted stakeholders who can benefit from this federated service scenario, are the following ones:

- Infrastructure owners and operators
- Technology providers
- Intermediaries
- End-users

[1]Thanks to EU PPP Initiative for funding

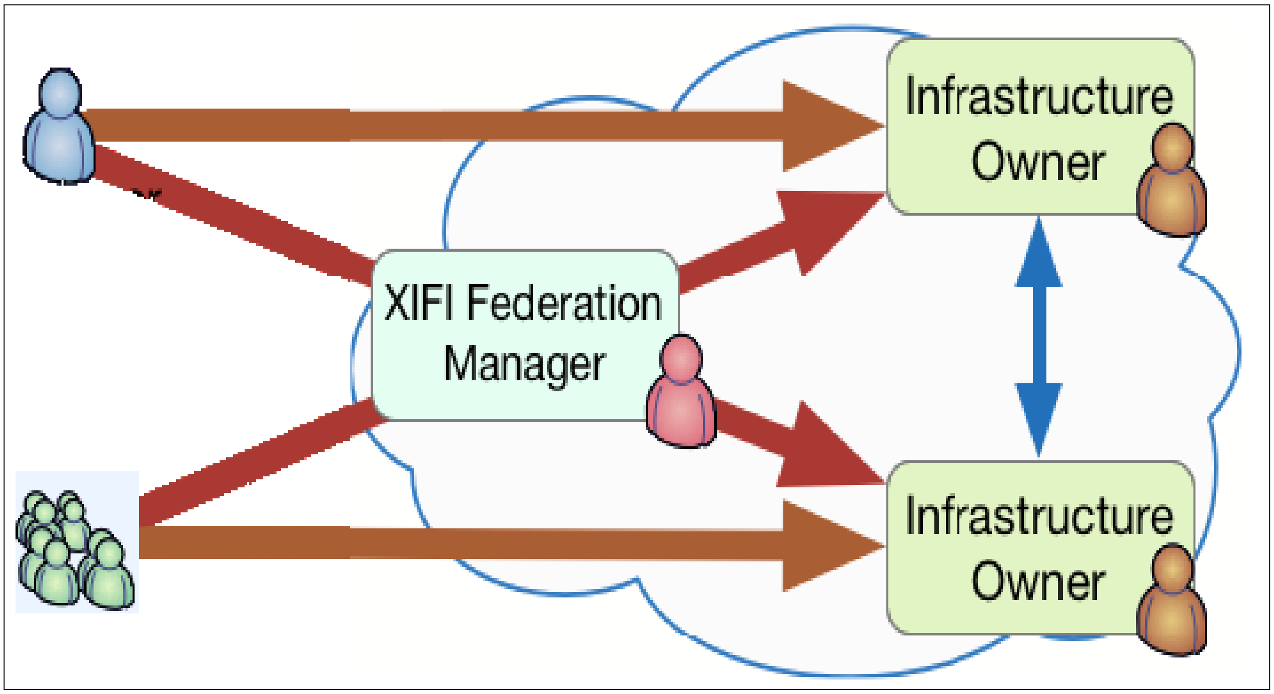

Figure 1. The XIFI project infrastructure, schematically

• Developers

This scenario is important because it shows that:

• A service could be deployed and executed over a big number of clouds located in different geographical areas and running on different cloud infrastructures based on public, private and hybrid clouds in a seamless way for the customers;

• The service end-users do not care where the service is deployed and running, while the service developers and providers have potentially unlimited resources availability provided by federated clouds And the infrastructure owners and operators, developers and endusers will benefit with this service scenario because the infrastructure is exploited to the maximum in a flexible and scalable way, while neither developers nor end-users are required to know about the cloud "location" and cloud composition.

4. Recommender Service

The recommendation service we propose, uses Linked Data as a knowledge base to find relationships between concepts and objects belonging to the Web of Data. Initially this recommender, was designed to operate independently of the application domain in which it is used. It has been used and tested in the past to recommend sights and attractions for tourists, as described in [6]. In contrast to that work, in this paper we present the modifications made on its architecture in order to deploy it and execute it in a federated environment. With this architecture, a more efficient operation is obtained as the resources of the cloud nodes in which the service is deployed are dynamically adapted to facilitate its execution. In this way we have been able to benefit from the potential of the federated XIFI cloud.

The federated service architecture in Figure 2, shows the service components present in each node of the cloud where it is deployed.

4.1 Service Operation

The execution flow of the service is as follows:

• The service receives as input an HTTP request (POST or GET), which contains a text from which the enduser wants to get

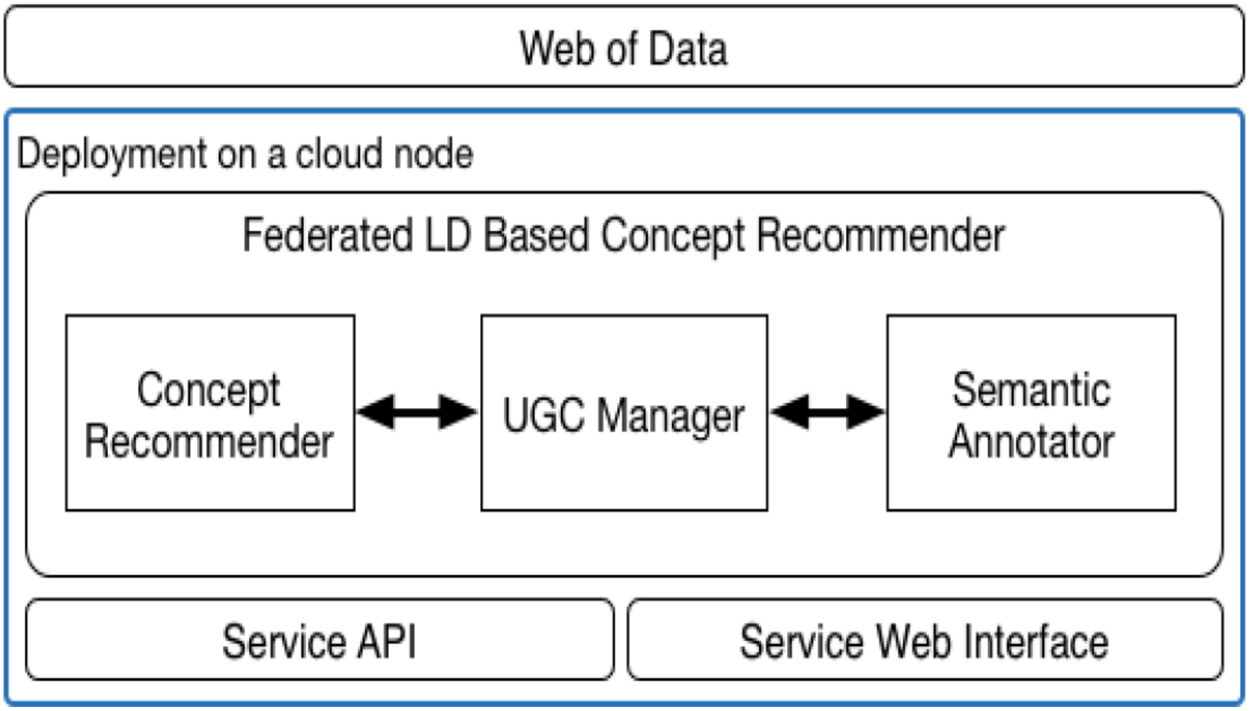

Figure 2. Federated Recommender Service Architecture

recommendations of similar content present on the Internet or on the platform itself (through the UGC manager described in V).

• The federation manager and the load balancer redirect the request to the most appropriate cloud node based on the configuration of the federated cloud.

• The request is received by the service on the selected cloud node, initially the semantic annotator module (described in VII), will perform an analysis to find associations between the nouns found in the text and existing concepts in the Web of Data.

• Each of the associations found (between noun and URI) in the previous step, will be analyzed by the recommender (described in detail in VI), using their knowledge base (Web of Data) will generate a list of related concepts

• Finally, the recommender returns (as output of the service) a list of the candidates obtained either as a list of related URIs or as an enriched view with related content that has been found by performing semantic queries to the UGC module of the platform and/or to different Linked Data sources.

5. UGC Management Module

This layer of the platform is responsible for saving Usergenerated Content and its semantically enriched version. Thus, for each given UGC entry, we have:

• UGC itself (any given multimedia file);

• Original associated plain information (stored in a SQL database);

• Associated semantic information represented in RDF (stored in a triple store).

Additionally, this layer offers the means to retrieve specific content, either through the invocation of a REST API (to get the multimedia contents attached) or performing more complex SPARQL queries in the public endpoint.

6. Concept Recommender Module

Recommender systems (hereafter RSs) are software tools and techniques that suggest items to users [7]. Items can be of different kinds such as songs, news, posts on social networks, persons, services, etc. In this paper, we refer to only one type of RS known as Linked Data-driven Recommenders (LDRs). LDR are recommenders based on knowledge, but unlike traditional knowledge-based recommenders, they are based on datasets modeled, built, and maintained by different organizations and communities around the world. Those datasets contain knowledge from different domains and sources; and can be published on the Web of Data according the LD principles [8].

6.1 LD-driven Recommendations

LDR systems suggest items by measuring semantic distances based on the relationships of concepts in a dataset. Normally, items are associated to concepts of a dataset by using natural language processing techniques, which analyze textual information of items, and extract keywords that can be matched with concepts (or entities) of the dataset. In this way, an item is more likely to be suggested if its related concepts have lower semantic distance with the initial item than other concepts in the dataset.

Semantic distances are used by LDRs to exploit hierarchical or transversal relationships between concepts. Hierarchical relationships are links established over hierarchical properties that organize concepts according the categories they belong to. For example, DBpedia dataset supports three kinds of hierarchical relationships: SKOS (Simple Knowledge Organization System)[2], YAGO (Yet Another Great Ontology)[9], or Word- Net[3] categories. Transversal relationships are links connecting concepts without the aim of establishing a classification or hierarchy [10].

6.2 Recommendation Algorithm

Our LD-driven recommender combines both hierarchical and transversal relationships to generate more accurate recommendations. Additionally, it uses a category-based organization in order to detect a set of contexts in which concepts are arranged according to the application domain they belong to. In this way, the user can access easier to recommended items organized by broader categories, which can be seen also as an explanation for the recommendation. The hierarchical and transversal relationships were obtained from DBpedia because it is being established as the central interlinking hub for the Web of Data, enabling also access to many other datasets in the Linked Open Data Cloud[11] [12].

The algorithm 1 represents our LD-driven recommendation approach. It starts by creating a category graph (categoryGraph) based on hierarchical information extracted from an initial concept (URIin) until reach a maximum level (maxLevel) of categories in the category tree of DBpedia. The maxLevel value is used to limit the levels of super categories that the algorithm extract when navigating the category tree (Lines 1 - 4). Those categories are extracted using the hierarchical relationship skos:broader from the SKOS model of DBpedia.

Next, the algorithm extracts subcategories for all the super categories found in the last step for the categoryGraph, in order to go one level down to increase the possibility for finding more concept candidates (Lines 5 - 8). Then, it obtains the concepts for each category in the categoryGraph (including sub-categories), and arranges them in a map resultMap that relates each concept with the set of categories it belongs to (Lines 9 - 13). The resultMap is the set of candidate concepts found by the recommender, which still needs to be organized before being displayed to the user. For each candidate concept in the resultMap the algorithm calculates a transversal distance (td) counting the input/output properties that the candidate concept shares with the initial concept. Those distances are added to a map in order to make them available for generating the final ranking (distancesMap) (Lines 14 - 17).

Finally, by using the hierarchical information of the categoryGraph, the algorithm organizes the results by context categories and ranks the concepts for each category in descendent order of td (Line 18).

7. Semantic Annotator GE Module

---

[2] http://www.w3.org/2004/02/skos

[3] http://en.wikipedia.org/wiki/WordNet

Algorithm 1 Recommendation Algorithm

Require: An input *URI*: URI_{in}, *maxLevel*

Ensure: A ranked set of recommended concepts RC classified by categories

```
C_in = getCategories(URI_in)
for all c ∈ C_in do
SC = getSuperCategoriesUntilLevel(c; maxLevel)
categoryGraph:add(SC)
for all sc ∈ SC do
subC = getSubCategories(sc)
categoryGraph.add(subC)
end for
for all cg ∈ CategoryGraph do
subC = getConcepts(cg)
resultMap.add(subC, cg)
  end for
end for
for all candidateConcept ∈ result do
td = transversalDistance(candidateConcept, URIin)
distancesMap.add(candidateConcept, td)
end for
classifyResultsByContextCategories(results)
return ranking
```

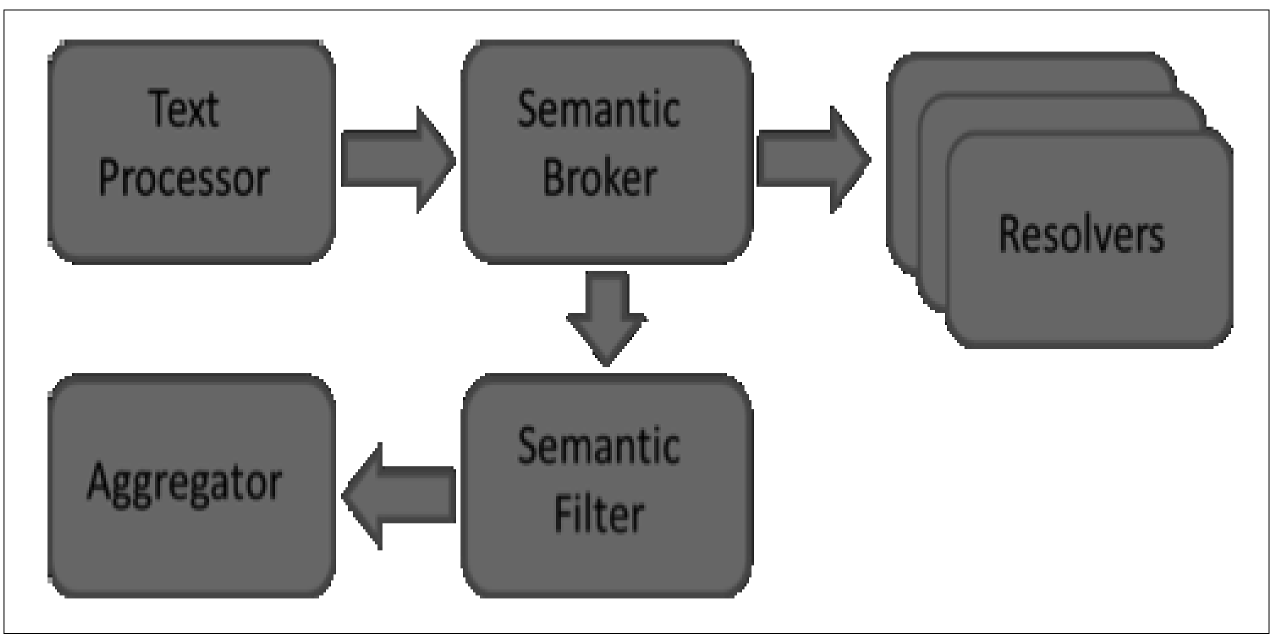

Figure 3. Semantic Annotation GE Architecture

The Semantic Annotation GE aims at performing named entity recognition and semantic annotation from a given text. This enabler is based on the Semantic Annotator described in [13].

The architecture of the Semantic Annotation GE is shown in Figure 3. In order to accomplish the semantic annotation task, it is composed by a Text Processor module, which processes the given text, and identifies the source language. Then, a morphological analysis is performed using FreeLing [14] configured with the identified language. From this analysis, are discarded. At this time, non-numeric proper nouns lemmas with a score of at least 0.2 are preserved and merged with plain tags to compute a well-defined list of unique (multi) words. At this stage, the module uses term frequency to further process the title and to extract other potential relevant words.

The next step involves the Semantic Brokering module assisted by a set of resolvers that perform full-text or termbased analysis based on the previous output. Such resolvers are aimed at providing candidate semantic concepts referring to Linked Data as well as additional related information if available. Resolvers may be domain or language-specific, or general purpose. For term-based analysis, each word of the previously computed list is individually processed to identify a list of candidate Linked Data resources to match with. A set of predefined services, such as DBPedia[4], Sindice[5] and Evri[6] are invoked in parallel.

The Semantic Filtering module processes candidate Linked Data resources received by the broker and performs a disambiguation based on the DBPedia score and the string similarity between each surface form and its corresponding list of candidates, based on the Jaro Winkler distance. This function aims at maximizing both values to identify the "preferred" candidate. In this process, after several empirical tests, candidates with distance lower than 0.8 are discarded at this stage, unless their DBpedia score is the maximum. Automatic annotation is performed using the "preferred" candidate identified during this step.

8. Conclusions and Future Work

This work presented an application of a Generic Enabler known as Semantic Annotator GE Module in an environment consisting of two federated clouds. The service running in the network is a recommender class service enriching original text of data with Linked Data objects contained in public databases. This service helps to a customer augmenting her/his knowledge about the object and providing with additional semantic information. The service is deployed in a way that the customers is connected to a more closed GE instance based on her/his network-geographical location in order to provide a faster service to the customer and lowest expenses to the service provider. The customer does not care and is not aware about which exact node in the network is serving her/his requests and service is enough robust to survive a failure of one entire geographical node. All this due to the double replication of the service in two different environments and configuration of the load balancers forwarding the traffic of failed node to the survived node.

The future work we still have to accomplish is mainly focused on the following main topics:

- Increase the service coverage by installation of other replications of the service in other geographical areas;
- Standardize and improve the APIs for connection of heterogeneous DataBases in Internet;
- Improve efficiency and precision of the recommendation algorithm;
- Improve the load balancer techniques depending on the available bandwidth and traffic load;
- Improve semantic engine for handling different application semantic domains with different logic of reasoning;
- Create and publish the distribution package of the instance with the configuration script for an easier new instance installation and configuration in the load balancer.

All these features will enable this service to more application domains and so increase its potential benefits and wider its usage

[4]http://dbpedia.org

[5]http://sindice.com

[6]http://www.evri.com

as both a stand-alone service and as a integrated part within other more complex platforms and services.

References

[1] Web services architecture. [Online]. Available: http://www.w3.org/TR/ws-arch/

[2] Jadeja, Y., Modi, K. (2012). Cloud computing - concepts, architecture and challenges, in Computing, *Electronics and Electrical Technologies* (ICCEET), 2012 International Conference on, March, p. 877– 880.

[3] XIFI EU Project. [Online]. Available: https://www.fi-xifi.eu

[4] Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked data - the story so far, *Int. J. Semantic Web Inf. Syst*., 5 (3) 1–22.

[5] FIWARE Project. [Online]. Available: http://www.fiware.org

[6] Rocha, O. R., Figueroa, C., Vagliano, I., Moltchanov, B. (2014). Linked data-driven smart spaces, in Internet of Things, *Smart Spaces, and Next Generation Networks and Systems* - 14th International Conference, NEW2AN 2014 and 7th Conference, ruSMART 2014, St. Petersburg, Russia, August 27-29, 2014. *In:* Proceedings, 2014, p. 3–15.

[7] Manning, C. D., Raghavan, P., Schutze, H. (2009). Introduction to Information Retrieval, *Computational Linguistics*, 35, p. 307–309, 2009. [Online]. Available: http://dspace.cusat.ac.in/dspace/handle/123456789/2538

[8] Berners-Lee, T. (2006). Linked Data - Design Issues, http://www.w3.org/DesignIssues/LinkedData.html, p. 7, 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html

[9] Suchanek, F. M., Kasneci, G., Weikum, G. (2007). Yago: A core of semantic knowledge, *In*: Proceedings of the 16th International Conference on World Wide Web, ser. WWW '07. New York, NY, USA: ACM, p. 697–706.

[10] Damljanovic, D., Stankovic, M., Laublet, P. (2012). Linked Data-Based Concept Recommendation: Comparison of Different Methods in Open Innovation Scenario, in The Semantic Web: Research and Applications, ser. Lecture Notes in Computer Science, E. Simperl, P. Cimiano, A. Polleres, O. Corcho, and V. Presutti, Eds. *Springer Berlin Heidelberg*, March, p. 24–38. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-30284- 8 9 files/119/Damljanovic et al. - 2012 - Linked Data-Based Concept Recommendation Comparis.pdf files/120/10.html

[11] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S. (2009). DBpedia - A crystallization point for the Web of Data, *Journal of Web Semantics*, 7, p. 154–165.

[12] Chiarcos, C., Hellmann, S., Nordhoff, S. (2011). Towards a linguistic linked open data cloud: *The open linguistics working group. TAL*, 52 (3) 245–275.

[13] Rocha, O. R., Goix, L. W., Mondin, F. L., Criminisi, C. (2012). *Lodifying personal content sharing*, New York, p. 124–130, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2320765.2320805

[14] Freeling language analyzer. [Online]. Available: http://nlp.lsi.upc.edu/freeling/