

ERCIM “Alain Bensoussan” Fellowship Scientific Report

Fellow: Giuseppe Pirrò
Scientific Contact: Jérôme Euzenat:
Visited Location : INRIA Grenoble - Rhône-Alpes
Duration of Visit: April 29, 2010 – April 29, 2011

I - Scientific activity

During my visit at INRIA I worked on three main topics. The first topic is semantic similarity in ontologies and its applications to different contexts such as ontology matching and semantic peer-to-peer networks. In particular, this research lead to the definition of a similarity framework, which combines the feature-based and information theoretic based models of similarity. Upon this model, a new measure of similarity has been defined. An extension of this measure to handle different parts-of-speech, especially useful to compute similarity between sentences, has also been proposed. Moreover, a comprehensive similarity library¹ has been implemented in Java, which includes different classes of similarity measures both between words and sentences. Ongoing research is focused on two main directions. The first is how to exploit the huge amount of information available in the Web of Data for the purpose of computing similarity also between individuals. The second one concerns the extension of the proposed similarity framework to expressive knowledge representation languages such as Description Logics.

The second topic concerned thesaurus matching. This research has been particularly interesting since it has been conducted in a real usage context, that is, the Publication Office of the European Commission (OPOCE). Research activities concerned the analysis of multilingual aspects of thesauri and their arrangement in concepts schemes for the purpose of matching. A thesaurus matcher has been implemented, which includes different strategies leveraging both linguistic and structural information. The evaluation of this matcher on a reference alignment provided by OPOCE provided encouraging results. Ongoing work concerns the definition of matching strategies exploiting documents annotated to concepts in different thesauri. This would be particularly helpful since thesauri, per se, are not endowed with instances. Moreover, the implemented matcher can be endowed with a GUI to facilitate user's interactions in composing and executing a thesaurus-matching task.

The third topic has been the study of biological workflow management systems, where it is needed to integrate different sources of information and tools. In particular, this research has been focused on biological networks. Some aspects of data integration from multiple biological databases such as DIP and MINT have been studied. Besides, similarity between proteins has also been studied on the basis of annotations of proteins to concepts of the Gene Ontology. As result of this study, a comprehensive system called BioTRON, has been defined. The system provides a user interface, which helps the user in composing and executing complex biological workflows.

¹ For further information visit: <http://simlibrary.wordpress.com>

II- Publication(s) during your fellowship

- Giuseppe Pirró, Jérôme Euzenat: A Feature and Information Theoretic Framework for Semantic Similarity and Relatedness. 9th International Semantic Web Conference (ISWC). LNCS 6496, pp. 615-630, 2010.

Abstract: *Semantic similarity and relatedness measures between ontology concepts are useful in many research areas. While similarity only considers subsumption relations to assess how two objects are alike, relatedness takes into account a broader range of relations (e.g., part-of). In this paper, we present a framework, which maps the feature-based model of similarity into the information theoretic domain. A new way of computing IC values directly from an ontology structure is also introduced. This new model, called Extended Information Content (eIC) takes into account the whole set of semantic relations defined in an ontology. The proposed framework enables to rewrite existing similarity measures that can be augmented to compute semantic relatedness. Upon this framework, a new measure called FaITH (Feature and Information THEoretic) has been devised. Extensive experimental evaluations confirmed the suitability of the framework.*

- Giuseppe Pirró, Jérôme Euzenat: A Semantic Similarity Framework Exploiting Multiple Parts-of Speech. 9th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE). LNCS 6427, pp. 1118-1125, 2010.

Abstract: *Semantic similarity between words aims at establishing resemblance by interpreting the meaning of the words being compared. The Semantic Web can benefit from semantic similarity in several ways: ontology alignment and merging, automatic ontology construction, semantic- search, to cite a few. Current approaches mostly focus on computing similarity between nouns. The aim of this paper is to define a framework to compute semantic similarity even for other grammar categories such as verbs, adverbs and adjectives. The framework has been implemented on top of WordNet. Extensive experiments confirmed the suitability of this approach in the task of solving English tests.*

- Valeria Fionda, Giuseppe Pirró : BioTRON: A Biological Workflow Managment System. 26th Symposium on Applied Computing (SAC). ACM Press, 2010, to appear.

Abstract: *Bioinformatics tasks may become very complex and usually require to manually integrate both data and results from different knowledge sources and tools. In this scenario, an integrated environment for designing and executing complex biological workflows is a must. Even though several efforts are trying to cope with this aspect, they mostly focus on gene or protein sequence analysis underestimating more complex biological data such as molecular interaction data. The aim of this paper is to present the BioTRON system, which supports biologists in the various steps necessary to perform complex biological tasks such as biological network comparison. BioTRON also features a mechanism to automatically integrate even existing on-line Web services. We present the BioTRON architecture along with a real example, which shows the suitability of the tool.*

- Giuseppe Pirró, Domenico Talia, Paolo Trunfio: A DHT-based Semantic Overlay Network for Service Discovery. Submitted to a journal.

Abstract: *The number of available Internet services increases every day. This trend demands distributed models and architectures to support scalability as well as semantics to enable efficient publication and retrieval of services. Two common approaches toward this goal are Semantic Overlay Networks (SONs) and Distributed Hash Tables (DHTs) with semantic extensions. SONs enable semantic-driven query answering but are less scalable than DHTs, which, on their turn, feature efficient but semantic-free query answering based on exact match. This paper presents a strategy and a system, called ERGOT, that combine DHTs and SONs to enable semantic-based service discovery in distributed infrastructures such as Grids and*

Clouds. ERGOT uses semantic annotations that enrich service specifications in two ways: (i) services are advertised in the DHT on the basis of their annotations, thus allowing to establish a SON among service providers; (ii) annotations enable semantic-based service matchmaking, using a similarity measure between service requests and descriptions. An extensive evaluation of the system is presented and discussed. The experimental evaluation we carried out confirmed the efficiency of the implemented strategy both in terms of accuracy of search and network traffic.

III -Attended Seminars, Workshops, and Conferences

- The 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), May 17-20, 2010, Melbourne, Victoria, Australia.
- The 9th International Semantic Web Conference (ISWC), November 7-11, 2010, Shanghai, China.
- The 9th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), October 26-28, 2010, Heraklion, Greece.
- 3rd KRDB school on Trends in the Web of Data (KRDBs-2010), 17-18 September, 2010, Bressanone, Italy.
- Visit to LOA-CNR, Trento, where I gave a seminar titled: *Computing Semantic Similarity with FaITH*. 22 October 2010.
- KRDB, Free University of Bolzano, Bolzano. Attended a series of seminars, December 2010
- Visit to the Ontology Engineering Group, Faculty of Computer Science, Universidad Politécnica de Madrid. 27-29 April 2011.

IV – Research Exchange Programme (12 month scheme)

- LOA-CNR, Trento, 17-23 October 2010. Contact: dott. Nicola Guarino.
- Ontology Engineering Group, Faculty of Computer Science, Universidad Politécnica de Madrid. 27-29 April 2011. Contact: prof. Asunción Gómez Pérez .