

Co-extracting content ontologies and expressivity from textual resources (competitive selection).

Textual resources are nowadays a primary source of information for several applications. However, textual resources cannot often be used as they are because they cannot be automatically processed. It is therefore required to make explicit the information (both hidden/implicit and explicit) carried on.

In this thesis, we are interested in extracting two information types from textual resources:

- 1) The ontological content mostly related to the subject/domain described by the texts (i.e. we are not interested in exactly representing meanings of sentences);
- 2) What we call “expressivity” i.e. any kind of variability used by the authors of the texts for expressing the ontological content (e.g. style, typical words and so on). Expressivity related information can be used for developing innovative applications requiring knowledge of humans (such as opinions, states and competencies) who are authors of texts.

The key point of the proposed work is to extract both information types at the same time. This is because whenever the ontological content is being extracted, much information gathered is clearly related to the expressivity as defined above.

We have already proposed a generic approach with some tools to extract the ontological content. This approach is based on linguistic patterns and comprises a validation framework. The thesis work will be performed in the context of the generic approach and contribute to 3 objectives:

- 1) To move from statically defined patterns to learned patterns (supervised/unsupervised); deep parse trees extracted from the texts and possibly rearranged are the data to be targeted by learning techniques for identifying common structures (clustering, clustering, pattern mining, graph mining); other information extracted from texts can be used, such as natural inferences as well as background knowledge, such as core and upper ontologies, for taking into account the partial knowledge domain and the poor structure and standardisation of texts on hand;
- 2) To perform an extensive validation based on the classification and procedure already proposed with the generic approach; the main work will be devoted to use the proposed framework for validating results obtained for various text types (e.g. technical texts, scientific texts, reference texts and so on);
- 3) To identify and implement expressivity related meta-data, linked to the extracted content ontology; expressivity is a new concept that needs to be explored; the work will be devoted to make an extensive analysis of existing concepts analogous/similar to expressivity; based on a this analysis the work should be focused on defining a meta-data model that can be

used with the ontology content extraction for collecting relevant information for representing the extracted expressivity.

Required qualifications

Only **top-rank candidates** with a degree equivalent to a “Master 2 Recherche” in France, in Computer Science or related areas with strong competencies both theoretical and practical in machine learning, data and text mining, with

Knowledge of NLP and Ontologies and their foundations (such as Description Logics),

Proficiency in Java programming (or equivalent) and

Be able to speak French or English and to write in English.

Supervisors of the work

Giuseppe Berio, Giuseppe.berio@univ-ubs.fr (IRISA, EXPRESSION team)

Nicolas Bechet, Nicolas.bechet@irisa.fr (IRISA, EXPRESSION team)

Mounira Harzallah, mounira.harzallah@univ-nantes.fr (LINA – Nantes, DUKE team, <http://duke.univ-nantes.fr/>)

Workplace

EXPRESSION team at IRISA – Vannes (<http://www-expression.irisa.fr/>)

Contract-Salary

Funding will be provided by Université de Bretagne Sud and PhD student signs a standard 3 years contract (fixed terms and standard salary). Expected salary is around 1684.93 euros gross/month. Standard additional funding is also possible (such as tutoring) when students fill requirements.

Selection procedure

Applications must be sent to Giuseppe.berio@univ-ubs.fr. We will receive applications till **3rd May 2015.**

Applications must include:

1. A **complete professional CV** (please do not put information about hobbies or similar),
2. **Official last 3 years notations and ranks,**
3. **Names and email address of 2 professors or researchers able to support your application,**
4. A **motivation letter,**

5. **One page describing in detail your proficiencies wrt the required ones.**

Applications fulfilling **all the above requirements** will be considered and applicants would be contacted for an **interview**. The unique successful applicant will be required to

- a) Formalize its candidature (details only given to the selected candidate) no later than **11th May**;
- b) Prepare and give a presentation (French or English) about the expected thesis work, to be presented on **18th May in Lorient** (presentations can be delivered with high quality visio conference systems). We will offer a support to the candidate for preparing a presentation with high quality scientific content. However, the candidate is expected to be autonomous for the other presentation aspects (styles, fonts, pictures, animations, words and sentences appropriateness) and speech.

Once all candidates have presented their own expected thesis work, the doctoral school provides an ordered list of candidates that will be used to select the successful candidates.

Bibliograhly (short)

M. Harzallah, G. Berio, A.L. Opdahl, "New Perspectives in Ontological Analysis: Guidelines and Rules for Incorporating Modelling Languages into UEML". In Information Systems. 37(5): 484–507. Elsevier, 2012.

A.L. Opdahl, G. Berio, M. Harzallah, R. Matulevicius, "Ontology for Enterprise and Information Systems Modelling". In Applied Ontology. Volume 7, Number 1 (2012): 49-92. IOS Press, 2012.

Rajani Chulyadyo, Mounira Harzallah and Giuseppe Berio Core Ontology based Approach for Treating the Flatness of Automatically Built Ontology. In Proceedings of KEOD: 316-323. Science and Technology Publications 2013.

Muhammad Aun Abbas and Giuseppe Berio. Creating Ontologies Using Ontology Mappings. Presented at KDO 2013, First Workshop on Knowledge Discovery in Ontologies. In Workshops proceedings of WI-IAT conference. IEEE, 2013.

S. Ghadfi, N. Béchet and G. Berio. Building Ontologies from Textual Resources: A Pattern Based Improvement Using Deep Linguistic Information. In Proceedings of Workshop on on Ontology and Semantic Web Patterns (WOP). A ISWC 2015 workshop. Vol 1302: 14-25. Ceur proceedings, 2014.

M. Harzallah, G. Berio. Towards an approach for configuring ontology validation, to appear in "Communications in Computer and Information Science" (CCIS), Springer-Verlag, 2015.

