

Co-extraction d'ontologie du contenu et expressivité dans les ressources textuelles (concours CDE)

Contexte

Le **sujet de thèse** porte en premier lieu sur le développement des méthodes et outils pour **l'extraction semi-automatique d'ontologies** du contenu (ou du domaine) à partir de ressources textuelles (hétérogènes). En deuxième lieu, le sujet propose la définition et l'expérimentation d'une représentation de **l'expressivité** (à savoir toute sorte de variabilité pour exprimer un même contenu, par exemple, terminologie particulière ou style du document), utilisable lors de l'extraction de l'ontologie. Un premier intérêt d'extraire conjointement l'ontologie et des formes d'expressivité est de pouvoir établir les liens et la frontière entre elles et les spécificités de chacune (par exemple, pour un texte donné, il s'agit de mettre en évidence ce qui relève de l'ontologie et ce qui relève de l'expressivité) ; il s'agit aussi de pouvoir réutiliser de l'information intéressante, collectée pour l'extraction de l'ontologie mais qui n'est pas représentée par l'ontologie. Un deuxième intérêt est de pouvoir fouiller les informations d'expressivité pour extraire ensuite des points de vue, attitudes, états des humains auteurs des textes ; des applications innovantes utilisent ses informations pour reconnaître des opinions, des situations, et des capacités de ses humains.

Le travail de thèse se fera dans le contexte de l'approche générique **approche générique** d'extraction semi-automatique d'ontologie à partir des ressources textuelles que nous développons.

Contributions attendues

Dans le cadre de la thèse, nous proposons de contribuer aux **3 objectifs** suivants qui portent à la fois sur l'amélioration de l'approche générique mentionnée ci-dessus et sur la liaison ontologie-expressivité.

Objectif 1. Le travail de thèse abordera l'apprentissage de **patrons linguistiques** à l'aide de techniques de fouille (motifs séquentiels, arbres, graphes) afin de permettre une extraction semi-automatique de ces patrons.

Objectif 2. L'amélioration et expérimentation de l'approche de validation de l'ontologie.

Objectif 3. Nous proposons aussi de contribuer à la définition d'une première représentation formalisée de **métadonnées** liées à l'ontologie du contenu, portant sur l'expressivité au sein du texte et à la mise en œuvre des méthodes pour identifier cette expressivité au sein de ressources textuelles.

Qualification

Seulement les **candidats classés au meilleur niveau** d'un « Master 2 Recherche » en France ou équivalent à l'étranger, spécialité Informatique ou similaire, ayant acquis des fortes

compétences pratiques et théoriques en apprentissage automatique, fouille de données et de textes,

possédant des connaissances en Ontologies et Traitement automatique de la langue naturelle,

une excellente maîtrise de la programmation Java (ou équivalent) et

une excellent communication en Français ou en Anglais et une excellente maîtrise de l'Anglais écrit.

Encadrants

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

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

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

Lieu de travail

Equipe EXPRESSION de l'IRISA – Vannes (<http://www-expression.irisa.fr/>)

Contrat et salaire

Il s'agit d'un un contrat type standard (3 ans). Le doctorant sera payé directement par l'Université de Bretagne Sud avec un salaire attendu de 1684.93 euros bruts par mois. En fonction des besoins de l'établissement et du respect des conditions requises, des heures de vacation ou de monitorat pourraient être allouées.

Procédure de sélection

Toute demande de candidature ou d'information complémentaire doit être envoyée à giuseppe.berio@univ-ubs.fr.

Pour postuler, il faut envoyer à Giuseppe BERIO (giuseppe.berio@univ-ubs.fr) avant le **3 Mai 2015** :

1. Un **CV complet** portant sur vos diplômes, expériences et compétences,
2. **Les bulletins des notes et les classements pour les 3 dernières années,**
3. **Noms et adresses e-mail de 2 chercheurs ou enseignants-chercheurs** pouvant répondre à des questions sur vos expériences et niveaux,
4. Une **lettre de motivation,**
5. **Une page à part décrivant en détail vos compétences/expériences par rapport aux compétences requises.**

Pour une sélection des demandes, des interviews avec les encadrants pourront être mises en place. Une seule demande de candidature sera retenue et le candidat devra

- a) Formaliser sa candidature (les détails seront données au candidat sélectionné) au plus tard le **11 Mai** ;
- b) Préparer et faire une présentation (en Anglais ou en Français) pour la session du **18 Mai à Lorient** de l'École Doctorale SICMA (il est envisageable d'utiliser la visioconférence mais il faut disposer d'un très bon système).

Une fois que tous les candidats auront été auditionnés, l'école doctorale établira un classement des candidats. Seulement ce classement permettra de retenir le candidat pour la signature d'un contrat.

Bibliographie (courte)

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.