

## Offre de Stage Master2 Recherche

**Titre :** Fouille de patrons pour l'enrichissement semi-automatique d'ontologie à partir des textes

**Mots clés :** Extraction de patron, Construction semi-automatique d'ontologie à partir de textes  
TAL, Machine learning.

**Encadrants :** Giuseppe Berio, Nicolas Béchet (giusepppe.berio@univ-ubs.fr, nicolas.bechet@irisa.fr, EXPRESSION (expression.irisa.fr), IRISA, Vannes).  
Mounira Harzallah (mounira.harzallah@univ-nantes.fr, LINA, Duke (duke.univ-nantes.fr), Nantes)

**Durée :** 6 mois                    **Indemnité :** 508€/mois

**Compétences requises.** Des compétences en TAL, en apprentissage automatique ou en web sémantique sont souhaitées.

**Lieu du stage :** LINA, Nantes

**Poursuite en thèse** en relation avec le sujet est envisageable

**Contact.** Pour candidater, il faut envoyer votre CV, une lettre de motivation et les relevés des notes des deux dernières années à aux encadrants mentionnés ci-dessus.

**Sujet.** Avec l'accessibilité croissante des textes sur le web et leur richesse, l'intérêt pour la (semi)automatisation de la construction d'ontologie à partir de textes est devenu réel. Des outils utilisant des techniques spécifiques issues de domaines de recherches différents, notamment le TAL, l'apprentissage automatique ou le text mining, ont été développés afin de pouvoir extraire et organiser les connaissances explicites et/ou implicites, contenues dans le texte. Cependant, l'exploitation de ces techniques pour l'enrichissement d'une ontologie existante n'a pas été suffisamment explorée.

L'objectif du stage est d'étudier les techniques **d'analyse lexicale profonde** de textes et les **techniques d'apprentissage** afin de proposer une méthode et un outil : i) pour découvrir des nouveaux **patrons** (morpho-syntaxiques ou linguistiques) **d'extraction** des relations core préalablement définies dans une ontologie core ; et ii) d'utiliser ces patrons pour bien **classifier des nouveaux concepts**, extraits à partir de textes, dans une ontologie ayant cette **ontologie core**.

### Références

Buitelaar, P., Cimiano, P., and Magnini, B. (2005). Ontology Learning from Text: Methods, Applications and Evaluation, chapter Ontology learning from text: an overview, pages 3–12. IOS Press.

Chulyadyo R., M. Harzallah and G. Berio, Core Ontology Based Approach for Treating the Flatness of Automatically Built Ontology, KEOD 2013, Vilamoura, Portugal.

- Cimiano P. & Volker J. (2005). Text2onto - a framework for ontology learning and data-driven change discovery. In A. MONTYOYO, R. MUNOZ & E. MÉTAIS, Eds., 2nd Eur. Semantic Web Conf., volume 3513, p. 227–238.
- Deprès, S., and Szulman S. (2007) Merging of legal micro-ontologies from European Directives. In Artificial Intelligence and Law 15, 187–200.
- Ghadfi S., N. Béchet, G. Berio Building Ontologies from Textual Resources: A Pattern Based Improvement Using Deep Linguistic Information. In Proceedings of the 5th Workshop on Ontology and Semantic Web Patterns (WOP2014), Riva del Garda, Italy, October 19, 2014. Edited by: Victor de Boer, Aldo Gangemi, Krzysztof Janowicz, Agnieszka Lawrynowicz. CEUR-WS Vol1302, November 2014, pp14-25.
- Gherasim, T., M. Harzallah, G. Berio, P. Kuntz (2013). Methods and Tools for Automatic Construction of Ontologies from Textual Resources: A Framework for Comparison and Its Application. Advances in Knowledge Discovery and Management 3, 177-201.
- Snow R., D. Jurafsky, Andrew Y. Ng. 2004. “*Learning syntactic patterns for automatic hypernym discovery*”. In NIPS 2004.
- Zouaq A., D. Gasevic, M. Hatala. 2011. “*Towards open ontology learning and filtering*”. Information Systems, v.36 n.7, p.1064-1081, November, 2011.