



Groupe de Travail DigiCosme Représentations Sémantiques Multilingues

Responsables : Marianna Apidianaki (LIMSI, CNRS, Université Paris-Saclay)
Olivier Ferret (CEA, LIST, LVIC)

RÉSUMÉ

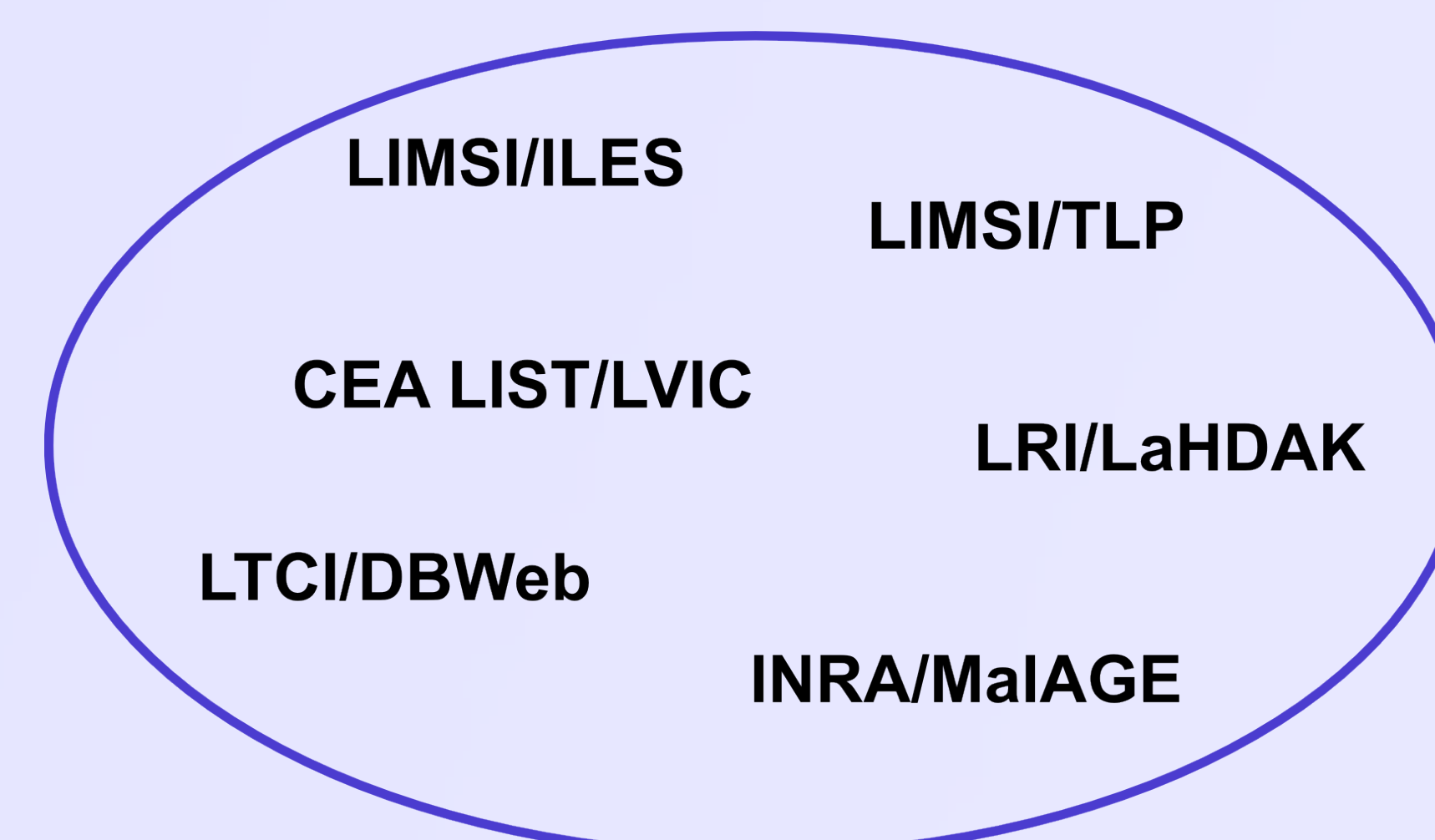
La représentation des connaissances sur les mots ou expressions d'un texte est un point clé pour de nombreuses tâches en Traitement Automatique des Langues (TAL). Le Groupe de Travail Représentations Sémantiques Multilingues se propose d'explorer différents aspects des représentations sémantiques pouvant améliorer la qualité des résultats des processus de TAL, en particulier pour des tâches d'extraction d'information. Ces aspects concernent spécifiquement :

- la capacité à apprendre des représentations sémantiques de façon non supervisée à partir de grands corpus non annotés, en particulier dans le contexte des représentations apprises par les réseaux de neurones profonds (Deep Learning) ;
- l'intérêt de l'utilisation de ces représentations dans des systèmes d'apprentissage supervisé pour un nombre grandissant de tâches et l'articulation entre ces représentations et l'usage qui en est fait ;
- l'apprentissage de représentations communes à plusieurs langues.

Tâches concernées dans DigiCosme

- Axe DataSense : Tâches 2 (Making sense of complex, heterogeneous data and knowledge) et 3 (Machine learning)

MEMBRES



SÉMINAIRES 2014-2015

12 décembre 2014 : Slav Petrov, Google (NYC)

- Towards Universal Syntactic Processing of Natural Language

27 janvier 2015 : Stephan Gouws (invité du GT Deep Learning)

- Training Neural Word Embeddings for Transfer Learning and Translation

14 avril 2015 : Fabian Suchanec (LTCI, Paris-Saclay)

- A Hitchhiker's Guide to Ontology

19 juin 2015 : Ivan Vulić (KULeuven)

- Modeling Cross-Lingual Semantic Similarity based on Latent Cross-Lingual Concepts

10 novembre 2015 : Roberto Navigli (Sapienza University of Rome)

- Représentations vectorielles latentes et explicites pour la similarité de mots, la désambiguïsation et la liaison entre ressources

RÉUNIONS ET STAGES

Réunions

- 22 octobre 2014 (réunion de démarrage, LIMSI)
- 24 novembre 2014 (NanoInnov)
- 13 février 2015 (LIMSI)

Stage

- Sujet : Combinaison de méthode distributionnelle et d'extraction terminologique pour l'adaptation de ressources terminologiques
Stagiaire : Yunhe Wu
Stage co-encadré par Thierry Hamon (LIMSI/ILES) et Gaël de Chalendar (CEA LIST/LVIC)

RETOMBÉES SCIENTIFIQUES (2014-2015)

Dépôts de projets ANR

- projet franco-allemand ANR-DFG (GoAsq) sur la modélisation et la résolution de requêtes ontologiques sur des données médicales semi-structurées (LRI et LIMSI/ILES)
- pré-proposition de projet ANR ADDICTE (Analyse distributionnelle en domaine de spécialité) (CEA LIST, LIMSI/ILES, LINA (Nantes), CLLE-ERSS (Toulouse))

Autres

- dépôt d'un sujet de thèse sur le thème du GT (TermDist : CEA LIST/LIMSI)
- dépôt d'une proposition de GT (D2K, de la donnée à la connaissance, LRI/INRA) acceptée lors du dernier appel DigiCosme
- les séminaires, exposés et discussions ont aussi contribué aux travaux individuels des membres du groupe sur les thèmes du GT

ACTIVITÉS EN 2016

Séminaire du 25 mars 2016 : Marine Carpuat (University of Maryland)

- Modéliser et utiliser les divergences dans les corpus parallèles

Stages proposés

- Étude des représentations distribuées pour l'extraction de relations (LIMSI/ILES et CEA LIST)
- Normalisation : projection de termes médicaux vers un référentiel (INRA et LIMSI/ILES)