



Le laboratoire LIUPPA, en collaboration avec l'équipe projet TERRE-ISTEX, organise la journée d'études intitulée
« VALORISATION et gestion des DONNÉES de la RECHERCHE »
Jeudi 9 mars 2017 de 13h à 16h15
Bâtiment Recherche IAE/Faculté Droit Economie Gestion,
Salle des thèses
LIUPPA, Université de Pau et des Pays de l'Adour

La disponibilité accrue des ressources numériques scientifiques via des bibliothèques numériques spécialisées (REVUES¹, HAL², THESES³, ISIDORE⁴, NAKALA⁵, ISTEEX⁶) offre de nouvelles et nombreuses opportunités d'usage. Les historiens et les sociologues des sciences peuvent ainsi analyser la genèse des disciplines, les conditions et facteurs d'émergence de concepts au sein de communautés scientifiques, leurs modalités de circulation et d'appropriation par d'autres communautés. Il est également possible d'observer l'évolution des fronts de recherche, les croisements disciplinaires ainsi que les modalités concrètes de recherche dans la mesure où ce gigantesque corpus scientifique rend compte des terrains, des méthodes et des cadres théoriques mobilisés.

Organisée dans le cadre du projet TERRE-ISTEX⁷ (projet Chantiers d'avenir ISTEEX, 2016 – 2018), cette journée d'études a pour objectif de créer un espace de rencontres et d'échanges sur les problématiques d'analyse, de gestion et de valorisation des données de la recherche qui sont volumineuses et hétérogènes. Ainsi, dans le contexte des humanités numériques, ces échanges viseront notamment la définition de données de recherche et les potentiels d'analyse portés par des chercheurs en Sciences Humaines et Sociales, en Sciences de l'information, en informatique ou encore en géomatique.

Programme :

13h - 13h50 : **Les données de la recherche dans les thèses de doctorat : Le projet D4Humanities**
Joachim Schöpfel - GERiICO, ANRT, Université de Lille Sciences Humaines et Sociales

Résumé : Depuis 2014, l'Université de Lille Sciences Humaines et Sociales a mené plusieurs études sur le domaine des données de la recherche, en particulier dans les thèses de doctorat. Avec le soutien de la Direction de la Recherche, de la MESHS, de l'Ecole Doctorale SHS et du Conseil Régional Hauts-de-France, une équipe du laboratoire GERiICO, du SCD et de l'ANRT ont réalisé plusieurs enquêtes sur les pratiques et besoins et sur la production des données de la recherche, organisé plusieurs séminaires et conférences et publié différentes études et textes, dont un livre blanc sur la mise en place d'une politique d'établissement adaptée et réaliste. La communication présentera une synthèse de ces travaux, en mettant l'accent sur la formation des doctorants, sur les aspects juridiques et sur le lien avec le TDM. Dans un 2^e temps, elle présentera le projet en cours *Deposit of Dissertation Data in Social Sciences and Humanities – A Project in Digital Humanities (D4Humanities)*, avec une enquête qualitative sur l'application des « FAIR principes » en SHS, le développement d'un

¹ <https://www.revues.org/>

² Hal.archives-ouvertes.fr/

³ Theses.fr/

⁴ Rechercheisidore.fr/

⁵ <https://www.nakala.fr/>

⁶ Istex.fr/

⁷ <https://terreistex.hypotheses.org>



workflow local pour la connexion avec la TGIR Huma-Num (NAKALA) et deux études sur les données SHS et l'évolution des thèses dans l'environnement de l'Open Science.

13h50 - 14h40 : **Apprentissage non supervisé pour l'extraction de relations d'hyponymie à partir de textes scientifiques**

Mouna Kamel, Elena manishina, Cassia Trojahn, Nathalie Aussenac - IRIT, Toulouse

Résumé : Les articles scientifiques, spécifiques d'un domaine, sont riches en connaissances dudit domaine. Ces connaissances non structurées nécessitent le développement des techniques d'extraction de connaissances efficaces, qui consistent à identifier les concepts et les relations qui les relient. Les modèles d'extraction basés sur des approches d'apprentissage supervisé nécessitent d'annoter les corpus, ce qui est coûteux en ressources. Des travaux récents montrent que les méthodes d'apprentissage non supervisé donnent des résultats équivalents lorsqu'elles sont appliquées à des corpus volumineux. Appliquer de telles approches à des corpus de taille modeste est un nouveau défi.

L'approche que nous proposons tente de relever ce défi, en détectant automatiquement les concepts du domaine et en mettant en oeuvre un algorithme de classification basé sur les réseaux de neurones, afin d'identifier les relations d'hyponymie entre concepts. Notre objectif est de construire un système indépendant du domaine, pour lequel le paradigme de sélection de traits permettrait d'ajuster les paramètres du système en fonction du domaine et de la relation ciblée. L'évaluation menée à ce jour sur les articles issus de la revue Nature du corpus ISTEEX montre que notre modèle donne de meilleurs résultats que l'algorithme K-Means, dans les mêmes conditions d'évaluation, ce qui apparaît comme une alternative aux approches supervisées.

14h40 - 15h00 : **Pause café**

15h - 15h50 : **Questionner le texte scientifique pour caractériser la science et l'innovation**

Guillaume Cabanac - IRIT, Toulouse

Résumé : Le projet ISTEEX vise à acquérir et valoriser de l'information scientifique et technique. Répondre à des requêtes formulées par des chercheurs ou un plus large public -- via des moteurs de recherche -- constitue un axe incontournable de valorisation. Cet exposé traitera cependant d'une voie complémentaire de valorisation. Il s'agit de questionner les textes scientifiques et les réseaux sous-jacents (lexique, références, auteurs, institutions, etc.) pour faire émerger des connaissances inédites et révéler les ressorts de la création et la diffusion des savoirs scientifiques. Le propos sera étayé par des résultats de recherche mobilisant l'informatique, la sociologie des sciences et la scientométrie (désignant l'étude quantitative de la science et de l'innovation). Je souhaite communiquer mon enthousiasme pour la recherche interdisciplinaire médiée par la scientométrie, tout en illustrant les opportunités qu'offre le corpus ISTEEX aux chercheurs en science du texte et des données.

15h50 - 16h15 : Discussions, conclusion de la journée d'études.

L'inscription est gratuite mais obligatoire par retour de mail auprès de Marie-Noelle Bessagnet (marie-noelle.bessagnet@univ-pau.fr) et Eric Kergosien (eric.kergosien@univ-lille3.fr).

Elle inclut la participation à la pause café de l'après midi.