

UTILISATION D'ONTOLOGIES POUR AMÉLIORER LES PERFORMANCES DES SYSTÈMES DE RECOMMANDATION *(English version below)*

Établissement : **IMT Mines Alès** (Ecole nationale supérieure des mines d'Alès)

Centre et équipe de recherche : KID (Knowledge and Image analysis for Decision making) / LGI2P (Laboratoire de Génie Informatique et d'Ingénierie de Production) – <https://kidknowledge.wp.imt.fr>

Localisation : Alès (1h de Montpellier, 45min de Nîmes)

Ecole doctorale : I2S Information Structures Systèmes – <http://www.edi2s.univ-montp2.fr>

Spécialité : Informatique (section CNU 27) / Computer Science

Encadrement : Sylvie Ranwez, Professeur à IMT Mines Alès, Vincent Ranwez, Professeur à Montpellier SupAgro, Nicolas Sutton Charani, Maître assistant à IMT Mines Alès.

Financement : Bourse école

Début de la thèse : Automne 2018

Date limite de candidature : 15 juin 2018

Contact : sylvie.ranwez@mines-ales.fr, 04 34 24 62 62

Résumé

Les évolutions technologiques reçoivent notre quotidien et suscitent de nouvelles pratiques, de nouvelles attentes. Les systèmes de recommandation se multiplient dans nos activités de loisir (suggestion d'articles à acheter, de vidéos à regarder, de restaurants), activités professionnelles (offres d'emploi, formations professionnelles, veille technologique) ou sociales (nouveaux contacts). Le sujet de thèse proposé envisage de coupler des méthodes d'apprentissage automatique avec des méthodes et des outils basés sur la sémantique (ontologies de domaine) pour favoriser la recherche d'informations pertinentes pour un individu, la recommandation. Une partie de ces travaux portera sur l'interaction avec l'utilisateur et la visualisation. Une application privilégiée concernera la recherche de formations en réponse à une réorientation de carrière avec en ligne de mire l'optimisation de tout le processus de formation (optimisation du remplissage des formations, suivi de carrière). Les résultats de recherche pourront être valorisés dans un contexte réel via la société Edotplus¹, basée à Nîmes et spécialisée dans la gestion de formations professionnelles. Ce partenariat permettra l'application des solutions proposées dans un contexte réel à grande échelle.

Mots-clés

Recommandation, Ontologies, Visualisation, Mesures sémantiques, Indexation, Recherche d'information

¹ <https://www.e.plus/>

Description du sujet de thèse

La recommandation consiste à rechercher et filtrer les informations pertinentes pour un utilisateur dans un contexte précis. La plupart des systèmes de recommandation utilisent le filtrage collaboratif qui consiste à baser la recommandation sur les choix, les goûts, d'utilisateurs *ressemblants*. Pour cela il est nécessaire de caractériser d'une part les entités manipulées par le système de recommandation et d'autre part de caractériser les utilisateurs. Il devient alors possible d'identifier des similitudes entre eux et ainsi identifier les propositions les plus pertinentes. Cette thèse propose d'aborder la recommandation en croisant les techniques d'apprentissage automatique et les approches sémantiques (basées sur des ontologies de domaine). Les ontologies sont envisagées comme support à ces traitements sémantiques car elles permettent non seulement de décrire la connaissance d'un domaine, mais surtout d'y appliquer des mesures indispensables pour l'indexation de ressources, la recherche d'information (appariement entre requêtes et ressources indexées, classification des utilisateurs...) et, in fine, la recommandation. Elles peuvent également servir lors de la restitution des résultats qui pourront être proposés au travers de cartes sémantiques.

Une application privilégiée de ces travaux concerne la recommandation de formations professionnelles. Un partenariat avec la société Edotplus, basée à Nîmes, est envisagé. Une ontologie de la formation professionnelle a été créée dans une collaboration antérieure et permet de caractériser l'offre de formation (plus de 70 000 gérées actuellement par Edotplus). L'objectif est de l'exploiter pour une analyse fine des parcours professionnels et ainsi recommander une ou plusieurs formations en réponse à une volonté d'adaptation ou de réorientation professionnelle.

Une partie de la recherche consistera à étendre des travaux précédents (c.f. <https://kidknowledge.wp.imt.fr>). OBIRS et OntoLex sont des outils dédiés à la recherche d'information conceptuelle, ils pourraient être étendus pour prendre en compte un profil utilisateur. USI est utilisé pour le clustering et l'indexation d'un ensemble d'éléments et pourra être étendu pour une visualisation personnalisée et synthétique des résultats. Les mesures sémantiques sont au cœur de ces outils.

Le travail de thèse concernera principalement la résolution des verrous suivants.

Verrou 1. Les mesures sémantiques *intrinsèques* utilisent le graphe de l'ontologie pour définir une mesure entre les concepts qui la composent. Ces mesures assument quasiment toutes que la distance entre un concept et son hyponyme direct (relation ISA) est toujours la même (pas de longueur de branche dans le graphe). Or en pratique ce n'est pas le cas et dans notre contexte cette approximation pourrait conduire à des suggestions ou des regroupements de formations non pertinents. Mais l'intégration de longueurs de branche est loin d'être triviale et impose un traitement automatique. D'une part parce que ces branches sont trop nombreuses pour être gérées manuellement. D'autre part parce que ces longueurs sont difficilement évaluables par un expert humain (quelle est la longueur de branche pertinente entre "LANGAGE DE PROGRAMMATION" et "JAVA" ?). Et enfin, il faut assurer la cohérence de ces longueurs de branches sur l'ensemble du graphe. Nous proposons d'étudier les propriétés théoriques d'une ontologie dont les branches seraient valuées ; d'identifier les propriétés qu'elle doit satisfaire pour apporter une plus-value dans un Système de Recherche d'Information et de fournir les algorithmes permettant de vérifier pour une ontologie si elle vérifie ou non ces propriétés. La valuation automatique des branches pourra se baser sur nos travaux précédents qui combinent traitement de la langue et ontologies (projet ONTOLEX).

Verrou 2. La difficulté à présenter visuellement à l'utilisateur un ensemble (potentiellement très vaste) de formations pertinentes, l'empêche de les comparer rapidement et d'identifier les plus pertinentes dans son contexte. Deux axes de recherche sont envisagés : i) proposer une classification des informations en fonction de la sémantique de leur contenu, de l'ontologie de domaine et d'un utilisateur ii) proposer une visualisation efficace des propositions basée sur cette classification (cartes sémantiques). La simplicité et l'expressivité de l'interface doivent masquer la complexité des traitements sémantiques sous-jacents. La classification d'éléments indexés en groupe d'éléments eux même indexés a de nombreuses applications : organisation automatiques de bookmarks, d'images, de mails, de documents en groupes sémantiquement cohérents. Ces travaux devraient aboutir à des développements algorithmiques intéressants (traitement en temps réel de très nombreux documents en exploitant des larges ontologies) pour un large panel d'applications.

Objectifs et résultats attendus

Les travaux de recherche envisagés dans le cadre de cette thèse visent à proposer des modèles théoriques qui répondent aux objectifs et verrous mentionnés plus haut. Ces modèles concernent à la

foi la connaissance des domaines considérés (exploitation des ontologies pour les calculs sémantiques) et les algorithmes utilisés pour l'indexation, la recherche, le filtrage et la visualisation des entités manipulées. Un prototype expérimental sera développé et servira pour la validation. Les résultats donneront lieu à des publications scientifiques dans des journaux ou des conférences de rang A.

Précision sur l'encadrement

Directeur de thèse : Sylvie Ranwez, Professeur, LGI2P/IMT Mines Alès
sylvie.ranwez@mines-ales.fr, 04 34 24 62 62

Co-directeur de thèse : Vincent Ranwez, Professeur, Montpellier SupAgro
vincent.ranwez@supagro.fr, 04 99 61 28 75

Encadrant de proximité : Nicolas Sutton-Charani
nicolas.sutton-charani@mines-ales.fr, 04 34 24 62 67

Sylvie Ranwez est une experte en représentation des connaissances et utilise les ontologies en particulier pour la recherche d'information, l'indexation et la visualisation de données via des cartes sémantiques. Vincent Ranwez est un collaborateur de longue date qui apporte son expertise en optimisation algorithmique et en théorie des graphes. Nicolas Sutton Charani est un membre récemment recruté de l'équipe KID avec une forte expertise dans les domaines de l'apprentissage automatique et des méthodes probabilistes. Leur complémentarité permet de couvrir l'ensemble des aspects scientifiques qui pourront être utile dans ce sujet de thèse ambitieux.

Profil du candidat

Master II en Informatique, ingénieur informaticien ou équivalent

Une bonne maîtrise des techniques classiquement utilisées dans les domaines de l'apprentissage automatique et de la fouille de données est souhaitée. Des connaissances en ingénierie des connaissances (ontologies), web sémantique, extraction d'information et algorithmique seraient fortement appréciées – un fort intérêt pour ces thématiques est attendu.

Un prototype sur des jeux de données à taille réelle est envisagé pour cette thèse avec application possible dans un contexte industriel. Des compétences en développement informatique et en particulier dans le développement de codes de calcul (algorithmes) sont requises. Le candidat aura un bon niveau en anglais (notamment à l'écrit), ainsi qu'une bonne capacité à travailler en équipe et à respecter les échéances.

Présentation de l'établissement et du centre d'accueil

L'institut Mines-Télécom (IMT) est un grand établissement public dédié à l'enseignement supérieur et la recherche pour l'innovation ; c'est le premier groupe d'écoles d'ingénieurs en France. Acteur majeur du croisement entre les sciences et les technologies du numérique et de l'ingénieur, l'IMT met les compétences de ses écoles en perspective dans les grands champs des transformations numériques, industrielles, énergétiques et éducatives. L'IMT fédère 11 écoles d'ingénieur publiques réparties sur le territoire national, qui forment 13 500 ingénieurs et docteurs. L'IMT emploie 4500 femmes et hommes et dispose d'un budget annuel de la recherche partenariale de 400M€ au sein de 55 centres de recherche rattachés à ses écoles. Il produit chaque année plus de 2000 publications et 60 brevets.

IMT Mines Alès est une des écoles de l'IMT. Forte de 180 ans d'histoire au service de la science et des entreprises, l'école dispose d'une solide culture scientifique et technique qu'elle met au service de l'enseignement, de la recherche et du transfert technologique. L'école emploie 350 personnes et forme plus de 1100 élèves ingénieurs et chercheurs ouverts sur le monde. Ses 3 centres de recherche développent une activité de haut niveau scientifique en partenariat avec les entreprises, dans les domaines de l'environnement, des risques, des matériaux, du génie civil, de l'intelligence artificielle et du génie industriel et numérique. Les valeurs promues à l'école sont l'audace, l'engagement, le partage et l'excellence.

Les thèmes de recherche du **centre de recherche LGI2P** (Laboratoire de Génie Informatique et d'Ingénierie de Production) relèvent essentiellement du domaine des sciences et technologies de l'information et de la communication avec en ligne de mire l'automatisation cognitive, ou comment rendre plus fiables et plus efficaces les activités d'un opérateur humain confronté à une situation cognitive complexe. Le centre est organisé en deux équipes de recherche :

- **L'Equipe ISOE** (Interoperable Systems and Organization Engineering), qui propose des méthodes et des développements pour l'appui conceptuel, méthodologique et technique à un collectif d'acteurs impliqués dans une ingénierie nécessairement pluridisciplinaire et collaborative de systèmes complexes.
- **L'Equipe KID** (Knowledge representation and Image analysis for Decision), qui développe des modèles, et des techniques pour l'analyse, la sélection, le filtrage et la visualisation de données volumineuses dans un contexte de prise de décision.

L'équipe de recherche KID (Knowledge and Image Analysis) du LGI2P de l'IMT Mines Alès axe ses recherches autour de l'automatisation cognitive : comment rendre plus efficace les actions d'un opérateur humain dans un contexte de surcharge cognitive, en lui proposant des outils d'analyse, de synthèse, et de visualisation d'information. Les travaux proposés ici font clairement références à des thèses antérieures dont certains résultats constituent des bases solides de la recherche envisagée :

- Sur la recherche d'information conceptuelle – thèse de Mohameth François SY [2012]
- Sur l'indexation conceptuelle et le clustering sémantique – thèse de Nicolas Fiorini [2015]
- Sur les calculs de similarité/distances sémantiques – thèse de Sébastien Harispe [2014]
- Sur la visualisation – thèses de Mohameth François SY et Nicolas Fiorini, déjà citées.
- Traitements automatiques de la langue – thèse de Pierre-Antoine Jean [2017] (membre de la société Edotplus depuis la fin de sa thèse)

Références bibliographiques de l'équipe sur le sujet

Fast and reliable inference of semantic clusters. *Nicolas Fiorini, Sébastien Harispe, Sylvie Ranwez, Jacky Montmain, Vincent Ranwez.* Knowledge-Based Systems, Elsevier, Volume 111, 1 November 2016, Pages 133–143.

USI: a fast and accurate approach for conceptual document annotation. *Nicolas Fiorini, Sylvie Ranwez, Jacky Montmain, Vincent Ranwez.* BMC Bioinformatics, Volume 16, Issue 83, 14 March 2015.

Coping with imprecision during a semi-automatic conceptual indexing process. *Nicolas Fiorini, Sylvie Ranwez, Jacky Montmain, and Vincent Ranwez.* In Information Processing and Management of Uncertainty (part III), proceedings of IPMU 2014, 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Series: Communications in Computer and Information Science, Vol. 444, Springer, Laurent, A., Strauss, O., Bouchon-Meunier, B., Yager, R.R. (Eds.), ISBN: 978-3-319-08851-8, pp. 11-20, Montpellier, France, July 15-19 2014.

How ontology based information retrieval systems may benefit from lexical text analysis. *Sylvie Ranwez, Benjamin Duthil, Mohameth François Sy, Jacky Montmain, Patrick Augereau, Vincent Ranwez.* In « New Trends of Research in Ontologies and Lexical Resources », chapter 11, pp. 209-230, Series: Theory and Applications of Natural Language Processing, Oltramari, Alessandro; Vossen, Piek; Qin, Lu; Hovy, Eduard (Eds.), Springer, ISBN 978-3-642-31781-1, February 2013.

User Centered and Ontology Based Information Retrieval System for Life Sciences. *Mohameth-François Sy, Sylvie Ranwez, Jacky Montmain, Armelle Regnault, Michel Crampes, Vincent Ranwez.* In BMC Bioinformatics, 13(Suppl 1):S4, 2012.

Processus de candidature

Le dossier de candidature doit être envoyé par mail aux trois encadrants : sylvie.ranwez@mines-ales.fr, vincent.ranwez@supagro.fr et nicolas.sutton-charani@mines-ales.fr avant le 15 juin 2018. Il sera composé de : un CV détaillant en particulier les expériences dans les domaines en lien avec le sujet de thèse, une lettre de motivation, les relevés de note et les classements obtenus dans les dernières années de formation (master ou équivalent), des lettres de recommandation. Les candidats dont le dossier sera retenu seront invités à passer un entretien.

USE OF ONTOLOGIES TO IMPROVE THE PERFORMANCE OF RECOMMENDATION SYSTEMS

Host institution: **IMT Mines Alès** (Engineering School)

Laboratory and research team: KID (Knowledge and Image analysis for Decision making) / LGI2P (Center of Computer Science and Production Engineering)
<https://kidknowledge.wp.imt.fr>

Location: Alès, South of France (1h from Montpellier, 45min from Nîmes)

Doctoral school: I2S Information Structures Systems – <http://www.edi2s.univ-montp2.fr>

Discipline: Computer Science

Thesis supervisor: Sylvie Ranwez, Professor at IMT Mines Alès, Vincent Ranwez, Professor à Montpellier SupAgro, Nicolas Sutton Charani, Lecturer at IMT Mines Alès

Funding: IMT Mines Alès

Starting date: Autumn 2018

Application deadline: June, 15th 2018

Contact: sylvie.ranwez@mines-ales.fr, +33 (0) 434 246 262

Summary

Our daily lives evolve with technological developments and new practices and new expectations arise. Recommender systems are pervading in our activities being leisure ones (suggestion of items to purchase, videos to watch, restaurants), professional ones (job offers, professional training, technological survey) or social ones (new contacts). The proposed thesis aims at coupling automatic learning methods with semantic-based methods and tools (domain ontologies) to promote the search for relevant information (filtering) for a personalized recommendation. Part of this work will focus on user interaction and data visualization. One use case is the search of relevant corporate trainings in response to a career reorientation. The optimizing aim being at least twofold: providing the most relevant offer to end-user in a way that ensures optimal repartition of learners in the proposed trainings. The research results can be used in a real context via the Edotplus² company, based in Nîmes and specialized in professional training management. This partnership will enable the proposed solutions to be applied in a large-scale real-world context.

Keywords

Recommender systems, Ontologies, Visualisation, Semantic measures, Indexing, Information Retrieval

Description of the thesis project

Recommendation consists in searching and filtering the relevant information for a user in a precise context. Most recommendation systems use collaborative filtering which recommend items known to

² <https://www.e.plus/>

be appreciated by similar users. Therefore, it is necessary to characterize on the one hand the entities manipulated by the recommender system and on the other hand to characterize the users. It then becomes possible to identify similarities between users that have evaluated similar items and hence to propose to one user the items he may appreciate, since people with similar taste like them. This thesis proposes to approach the recommendation task by combining automatic learning techniques and semantic approaches (based on domain ontologies). Ontologies are considered as a pillar of these semantic treatments because they allow not only to describe the knowledge of a domain, but also to apply semantic measures for the indexing of resources, the search for information (matching between queries and indexed resources, classification of users...) and, *in fine*, the recommendation. They can also be used for displaying results through semantic maps.

One use case of this work concerns life-long learning recommendation. A partnership with Edotplus, based in Nîmes, is envisaged. An ontology of life-long learning domain was created in an earlier collaboration and makes it possible to characterize the training offer (more than 70,000 training offers are currently managed by Edotplus) and a detailed analysis of career paths and thus to recommend one or more training courses to someone which intend to adapt or reorient his/her career.

Some previous works of our team can be adapted to this recommendation context (see <https://kidknowledge.wp.imt.fr>). OBIRS and OntoLex tools are dedicated to the search for conceptually indexed information, they could be extended to take into account a user profile. USI is used for clustering and indexing a set of items and can be extended for a personalised and synthetic visualisation of the results. Semantic measures are at the core of these tools.

The proposed thesis will mainly focus on the resolution of the following locks.

Lock 1. Intrinsic semantic measures use the ontology graph to define a measure between its component concepts. Almost all these measures assume that the distance between a concept and its direct hyponym (ISA relationship) is always the same (no branch length in the graph). However, in practice this is not the case and in our context, this could lead to irrelevant suggestions of professional training courses. The integration of branch lengths is far from trivial and requires automatic processing. On the one hand, because these branches are too numerous to be managed manually. On the other hand, because such lengths are difficult to evaluate for a human expert (how long should the branch between "PROGRAMMING LANGUAGE" and "JAVA" be?). And finally, we must ensure the consistency of these branch lengths throughout the graph. We propose to study the theoretical properties of an ontology with branch lengths; to identify the properties that it must satisfy to bring an added value in an Information Retrieval System and to provide the algorithms allowing to verify for an ontology if it verifies or not these properties. This task of automatically assigning branch lengths to existing ontology can benefit from our previous OntoLex work that combines natural language processing and ontologies to improve information retrieval.

Lock 2. The difficulty of visually presenting the user with a (potentially very large) set of relevant items prevents him/her from comparing them quickly and identifying the most relevant in his/her context. Two research axes are envisaged: i) to propose a classification of retrieved items according to the semantics of their content, the domain ontology and the end-user profile; ii) to propose an efficient visualization of those retrieved items based on this classification (e.g. using semantic maps). The simplicity and expressiveness of the interface must mask the complexity of the underlying semantic processing to the end-user. The automatic classification of indexed elements into (automatically) semantically indexed groups has many applications: automatic organization of bookmarks in named, images, emails, documents into semantically coherent groups. This work should lead to interesting algorithmic developments (to allow real-time processing of a large number of documents using large ontologies) for a wide range of applications.

Objectives and expected results

This thesis should provide theoretical results and algorithms to tackle above mentioned challenges. Expected results may concern the knowledge modeling (use of ontologies for semantic measures calculus) as well as the algorithms used to index, retrieve, filter and visualize relevant resources. Implementing these solutions in an integrated platform will be part of the evaluation. Results are expected to be published in international journals and/or conferences.

Information on the supervisory staff

Head: Sylvie Ranwez, Professor, LGI2P/IMT Mines Alès
sylvie.ranwez@mines-ales.fr, +33 (0) 434 246 262

Co-supervisor: Vincent Ranwez, Professor, Montpellier SupAgro

vincent.ranwez@supagro.fr, +33 (0) 499 612 875

Co-supervisor : Nicolas Sutton-Charani, Lecturer, LGI2P/IMT Mines Alès

nicolas.sutton-charani@mines-ales.fr, +33 (0) 434 246 267

Sylvie Ranwez is an expert on knowledge representation and ontology with a strong expertise on information retrieval and data visualization based on semantic maps. Vincent Ranwez is a long term collaborator, bringing his expertise on algorithmic optimization and graph theory. Nicolas Sutton Charani is a recently recruited member of the KID teams with a strong expertise on machine learning technics and probabilistic approaches. Their complementarity allow to cover the whole set of scientific fields related to this ambitious thesis.

Candidate profile

Master II in Computer Science, computer engineer or equivalent

A good mastering of the techniques classically used in the fields of automatic learning and data mining is required. Background in knowledge engineering (ontologies), semantic web, information extraction and algorithmics would be highly appreciated – a strong interest in these themes is expected.

A prototype on full-scale data sets is planned during this thesis with possible application in an industrial context. Skills in computer development and in particular in the development of calculation codes (algorithms) are required. The candidate will have a good level in English (especially written), as well as a good ability to work in a team and meet deadlines.

Presentation of the institution and the host laboratory

The **Institute Mines-Télécom (IMT)** is a French public establishment dedicated to higher education and innovative research and, as it represents the ministries of industry and digital technology, it is the largest group of engineering schools in France. Being a major actor in the crossover of science, digital technologies and engineering, the IMT harnesses the competencies of its member schools and allows them to excel in the fields of the transformation of digital technologies, the industry, energy and education. The IMT brings 11 public engineering across France together. Collectively, they train 13,500 engineers and doctoral students as well as employing 4,500 women and men and manage a budget of €400m within 55 research centers connected to the schools. IMT publishes 2000 publications each year and registers 60 patents.

IMT Mines Alès is one of the schools of IMT and, thanks to its 180 years of history of service to science and industry, the school thrives due to its solid culture of science and technology that it puts at the disposal of training, research and technology transfer. The school employs 350 people and trains more than 1100 students, engineers and researchers. Its three teaching and research poles contribute the development of high level scientific research in the areas of risk environment, materials, civil engineering, industrial engineering and digital technology. The school proposes rich and varied career paths as teacher/researchers have job mobility in the different schools of IMT and can equally take up, if they wish, positions of responsibility within the functional management of the school during a share of their time. Examples include dean, research director, international relations or economic development positions. The values promoted at the school are boldness, commitment, sharing and excellence.

The research themes of the **LGI2P research center** (Computer Engineering and Production Engineering Laboratory) are essentially in the field of information and communication sciences and technologies with a focus on cognitive automation, or how to make more reliable and efficient the activities of a human operator faced with a complex cognitive situation. It is organized into two research teams:

- The **ISOE Team** (Interoperable Systems and Organization Engineering), which proposes methods and developments for conceptual, methodological and technical support to a collective of actors involved in a necessarily multidisciplinary and collaborative engineering of complex systems.

- The KID Team (Knowledge representation and Image analysis for Decision), which develops models and techniques for the analysis, selection, filtering and visualization of voluminous data in a decision-making context.

The KID (Knowledge and Image Analysis) research team of LGI2P at ITM Mines Alès focuses its research on cognitive automation: how to make a human operator's actions more efficient in a context of cognitive overload, by proposing analysis, synthesis and information visualization tools. The work proposed here clearly refers to previous research works, some of whose results form a solid basis for the proposed research:

- On the search for conceptual information - thesis by Mohameth François SY[2012]
- On conceptual indexing and semantic clustering - thesis by Nicolas Fiorini[2015]
- On semantic similarity/distances calculations - thesis by Sébastien Harispe[2014]
- On visualization - theses by Mohameth François SY and Nicolas Fiorini, already cited.
- Automatic language processing - thesis of Pierre-Antoine Jean[2017] (member of Edotplus since the end of his thesis)

Personal references related with the subject

Fast and reliable inference of semantic clusters. *Nicolas Fiorini, Sébastien Harispe, Sylvie Ranwez, Jacky Montmain, Vincent Ranwez.* Knowledge-Based Systems, Elsevier, Volume 111, 1 November 2016, Pages 133–143.

USI: a fast and accurate approach for conceptual document annotation. *Nicolas Fiorini, Sylvie Ranwez, Jacky Montmain, Vincent Ranwez.* BMC Bioinformatics, Volume 16, Issue 83, 14 March 2015.

Coping with imprecision during a semi-automatic conceptual indexing process. *Nicolas Fiorini, Sylvie Ranwez, Jacky Montmain, and Vincent Ranwez.* In Information Processing and Management of Uncertainty (part III), proceedings of IPMU 2014, 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Series: Communications in Computer and Information Science, Vol. 444, Springer, Laurent, A., Strauss, O., Bouchon-Meunier, B., Yager, R.R. (Eds.), ISBN: 978-3-319-08851-8, pp. 11-20, Montpellier, France, July 15-19 2014.

How ontology based information retrieval systems may benefit from lexical text analysis. *Sylvie Ranwez, Benjamin Duthil, Mohameth François Sy, Jacky Montmain, Patrick Augereau, Vincent Ranwez.* In « New Trends of Research in Ontologies and Lexical Resources », chapter 11, pp. 209-230, Series: Theory and Applications of Natural Language Processing, Oltramari, Alessandro; Vossen, Piek; Qin, Lu; Hovy, Eduard (Eds.), Springer, ISBN 978-3-642-31781-1, February 2013.

User Centered and Ontology Based Information Retrieval System for Life Sciences. *Mohameth-François Sy, Sylvie Ranwez, Jacky Montmain, Armelle Regnault, Michel Crampes, Vincent Ranwez.* In BMC Bioinformatics, 13(Suppl 1):S4, 2012.

Application process

Applications must be sent by June, the 15th 2018, conjointly to each supervisor: sylvie.ranwez@mines-ales.fr, vincent.ranwez@supagro.fr et nicolas.sutton-charani@mines-ales.fr. The application file must include a CV, a motivation letter, marks and ranking of the student during at least the two last years (master degree), reference letters. Successful applicants will be contacted to plan an interview.