

Proposition de thèse
Enrichissement de bases de connaissances volumineuses
à partir de textes et d'ontologies : gestion des
inconsistances et renforcement de la confiance
– *Application aux approches thérapeutiques non conventionnelles* –

Knowledge bases enrichment from text analysis and
ontologies: management of inconsistencies and trust
strengthening

Début de la thèse: automne 2015

Thèse en Informatique (section CNU 27) de l'école doctorale I2S de l'université de Montpellier

Localisation : Centre de recherche LGI2P de l'école des mines d'Alès, Nîmes, France

Equipe de rattachement: Knowledge and Image Analysis for Decision making – KID

Direction de thèse: Sylvie Ranwez (HDR, LGI2P, école des mines d'Alès)

Encadrant de proximité: Isabelle Mougenot (UM Espace-Dev), Sébastien Harispe (LGI2P, école des mines d'Alès).

Mots-clés: Ontologies, extraction de connaissance, données liées, traitement du langage naturel (NLP), mesures sémantiques, incertitude

Financement : Proposition pour le concours de l'école doctorale I2S, Université de Montpellier.

Profil(s) de candidats souhaité(s) :

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, traitement automatique des langues, extraction d'information et algorithmique seraient fortement appréciées – un fort intérêt pour ces thématiques est cependant attendu.

Des compétences en développement informatique et en particulier dans le développement de codes de calcul (algorithmes) sont requises (de préférence en Java). 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.

Sujet de thèse

De nombreux travaux en Intelligence Artificielle illustrent les perspectives offertes par la définition d'approches hybrides couplant l'utilisation de représentations des connaissances (ontologies) et l'analyse sémantique de gros corpus de textes, notamment dans les domaines de la recherche d'information, de la recommandation, du question-réponse, et de l'aide à la décision. Dans ce contexte, cette thèse propose de s'intéresser à la question centrale de la véracité des ressources exploitées en étudiant cette problématique sous l'angle de l'extraction d'information, et plus particulièrement de relations entre entités, en vue d'enrichir une base de connaissances. Il s'agit d'étudier différentes approches permettant d'exploiter des techniques d'analyse de textes pour compléter des bases de connaissances avec de nouvelles relations, mais aussi d'étudier la définition de modèles permettant de d'évaluer le degré de confiance à accorder à de la connaissance ajoutée ou existante au regard des incohérences et contradictions détectées dans les textes et autres bases de connaissances. La connaissance n'étant pas immuable, mais plutôt en perpétuelle reconsidération,

ce projet abordera aussi la question de la prise en compte de la dimension temporelle et plus largement de son contexte. Cette thèse vient en complément de travaux de l'équipe encadrante portant sur la gestion des incertitudes dans l'extraction et l'inférence de connaissances. L'objectif final étant de renforcer la confiance dans des bases de connaissances, et dans les résultats qui découleront de son analyse dans un contexte de décision (e.g. inférences, recommandations).

Le domaine d'application concerne l'étude des approches thérapeutiques non conventionnelles (i.e. médecine parallèle : thérapies alternatives, ostéopathie, acuponcture, homéopathie, sophrologie, ...) et l'analyse portera sur des données ouvertes ou librement accessibles sur Internet (e.g., réseaux sociaux, forums dédiés, données liées).

Abstract

Numerous studies in Artificial Intelligence underline the interesting perspectives offered by the definition of new hybrid approaches mixing knowledge bases (ontologies) and semantic analyses of large text corpora – in particular for information retrieval, recommendation, question answering and decision making. In this context, this thesis proposes to focus on the important topic of knowledge sources veracity/trust. This central question will be tackled from the perspective of enriching knowledge bases using information extraction approaches, in particular relation extraction techniques, and inferences. The aim is to propose new techniques to enrich knowledge bases by considering models that question the degree of confidence an existing or newly added knowledge must have – according to inconsistencies and contradictions that can be detected from texts or knowledge bases analyses. Knowledge being in perpetual motion, this project will also study issues related to temporal and contextual dimensions of knowledge. This positioning is in line with works made by the supervising team related to uncertainty handling in the process of knowledge inference and extraction. The principal aim of the contributions that are expected in this thesis is to improve confidence on knowledge bases, as well as results that will be derived from their analysis in a decision making process (e.g. inferences, recommendations).

Contributions of this work will be applied to the study of alternative medicine (e.g. acupuncture, homeopathy, osteopathy); analyses will be made by considering open or freely available data exposed over the Internet (e.g. social networks, domain-specific forums, linked data).

Cette thèse est proposée en co-encadrement entre l'équipe KID (Knowledge and Image analysis for Decision making) du LGI2P de l'école des mines d'Alès (Laboratoire de Génie Informatique et d'Ingénierie de Production) situé à Nîmes sur le parc scientifique Georges Besse (lieu d'accueil de la thèse) et l'UMR Espace-Dev (Université Montpellier, IRD, Université de la Réunion, Université des Antilles et Université de la Guyane).

Partenariat industriel

Un partenariat est actuellement à l'étude avec une société montpelliéraine (à confirmer).

Correspondant/Contact :

Sylvie Ranwez (sylvie.ranwez@mines-ales.fr) 04 66 38 70 44

Plus de détail

- sur l'équipe KID : <http://kidknowledge.wp.mines-telecom.fr/>
- sur le LGI2P : <http://www.lgi2p.mines-ales.fr>
- sur l'école des mines d'Alès: <http://www.mines-ales.fr>