

Proposition de stage recherche en laboratoire 2018-2019

Titre : Annotation automatique et sémantique de documents de formation basée sur les ontologies de compétences et métiers du marché de l'emploi

Mots clés : Semantic Labeling, Natural Language Processing, Ontology, Machine Learning

Description du sujet :

La révolution numérique a transformé le marché du travail dont les acteurs principaux sont aujourd'hui les sites emploi. De manière à faire correspondre offres d'emplois et candidats, les outils de ces sites reposent majoritairement sur le concept clé de compétence. Ainsi, les offres sont décrites en termes de compétences requises, les candidats sont invités à mettre en avant leurs compétences acquises, et les outils informatiques sont chargés d'en optimiser la mise en correspondance. Dans la grande majorité des cas, pour que la mise en correspondance offre/candidat puisse être réalisée, il est nécessaire que les descriptions des compétences requises/acquises soient exprimées dans un modèle bien défini. Ce besoin s'est traduit par l'apparition d'ontologies des compétences et métiers telles que ESCO [1] qui peuvent être utilisées par les sites pour le calcul d'appariement sémantique entre les définitions des métiers, les compétences associées, les offres d'emplois, et les spécificités des candidats.

Une grande partie des compétences des candidats est acquise dans l'accomplissement de parcours de formations universitaires qui sont structurés en modules. Chaque module est défini au sein d'un syllabus qui décrit en termes de compétences les objectifs visés. Il est naturel de penser que l'acquisition d'un module correspond à l'acquisition des compétences spécifiées dans son syllabus, et que ces compétences feront partie de celles mises en avant sur les sites de recherche de stage ou d'emploi.

Toutefois, l'écriture des syllabus est aujourd'hui entièrement réalisée « manuellement » par des enseignants et responsables de formation, en langage naturel, et sans lien sémantique direct avec les définitions des compétences utilisées dans les sites emploi. En d'autres termes, il n'y a aujourd'hui aucun moyen pour les candidats d'extraire les compétences qu'ils ont acquises dans leur parcours en les traduisant automatiquement dans les termes « officiels » utilisés par les outils numériques du marché du travail.

Afin de pallier ce manque, le but de ce stage est d'identifier les moyens permettant de relier automatiquement les modules de formations (au travers de leurs syllabus, documents de cours associés, etc.) aux compétences telles qu'elles sont définies dans une ontologie officielle des métiers et compétences.

Il s'agira donc d'explorer les techniques d'**Intelligence Artificielle** dédiées à l'annotation automatique et sémantique de documents basée sur les ontologies, ceci dans le but de normaliser les compétences référencées dans les documents de formations en les projetant vers un modèle ontologique comme ESCO. Pour ce faire, le stagiaire devra réaliser un état de l'art du domaine de recherche lié au **Semantic Labeling**, en lien avec les techniques du **Natural Language Processing** pour le traitement des documents, les **Ontologies** pour la sémantique, et les travaux en **Machine Learning** pour la classification [2][3][4].

Cet état de l'art sera mis à profit dans le traitement et l'annotation sémantique (manuelle ou semi-automatique) d'un échantillon de documents de formations donné et destiné à fournir une base de connaissances et de tests qui pourra par la suite servir aux algorithmes d'apprentissage et/ou à l'évaluation de différentes méthodes pour une automatiser de l'annotation sémantique basée sur les ontologies.

Références :

- [1] <https://ec.europa.eu/esco/portal/home>
- [2] Celine Alec (2016). Enrichissement et peuplement d'ontologie à partir de textes et de données du LOD: Application à l'annotation automatique de documents. Intelligence artificielle [cs.AI]. Université Paris-Saclay, 2016.
- [3] Pham, M., Alse, S., Knoblock, C.A., & Szekely, P.A. (2016). Semantic Labeling: A Domain-Independent Approach. International Semantic Web Conference.
- [4] Zhao, M., Javed, F., Jacob, F., & McNair, M. (2015). SKILL: A System for Skill Identification and Normalization. AAAI.

Encadrant(s) : Grégory Bourguin, Arnaud Lewandowski

email : gregory.bourguin@univ-littoral.fr