

## Laboratoire Informatique d'Avignon

**Sujet :** Extraction et exploitation de l'information des CV

**Stagiaire :** Étudiant de M2

**Co-encadrement :**

- Juan-Manuel Torres-Moreno ([juan-manuel.torres@univ-avignon.fr](mailto:juan-manuel.torres@univ-avignon.fr))
- Marc El-Bèze ([marc.elbeze@univ-avignon.fr](mailto:marc.elbeze@univ-avignon.fr))
- Luis Adrián Cabrera Diego ([luis-adrian.cabrera-diego@alumni.univ-avignon.fr](mailto:luis-adrian.cabrera-diego@alumni.univ-avignon.fr))
- Barthélémy Durette ([durette@adoc-tm.com](mailto:durette@adoc-tm.com))

Les cabinets de recrutement doivent aujourd'hui traiter des volumes de données de plus en plus important (candidatures électroniques) et de moins en moins qualifié (variété des formats, des sources des candidatures, etc). Ceci rend évident l'intérêt majeur d'un système automatique offrant une aide efficace pour sélectionner dans des masses de candidatures celles qui méritent d'être retenues en vue d'un examen approfondi. Grâce à une convention CIFRE obtenue en 2005 et une autre en 2012, le LIA s'est déjà fortement impliqué dans des travaux de recherches visant à proposer des algorithmes permettant de mettre au point quelques-uns des composants d'un tel système [1-5]. Dans la prolongation de ces travaux, nous projetons de nous intéresser aux problèmes d'extraction et de valorisation de l'information provenant des curriculums vitæ.

La société Adoc Talent Management<sup>1</sup> et le Laboratoire d'Informatique d'Avignon (LIA), partenaires de ce stage de M2, proposent de traiter trois questions :

- Découper les curriculums vitæ pour identifier les différentes sections qui les composent : identité, formation, expérience professionnelle, compétences, etc.
- Extraire automatiquement les termes (mono ou multi-mots) qui ont contribué le plus au processus de recrutement.
- Représenter graphiquement l'information trouvée dans chaque curriculum vitæ pour qu'elle soit synthétique et compréhensible pour les recruteurs ou les candidats.

Ces questions seront abordées dans le contexte des activités de recrutement du cabinet Adoc talent Management et seront valorisés au travers de la mise en place d'outils testés sur le serveur du cabinet et utilisés par les consultants en recrutement et possiblement à plus large échelle.

Pour la réalisation de ce stage de M2, l'étudiant devra, a priori, utiliser les méthodes liées à l'apprentissage automatique, la fouille de texte (représentations des textes, classifications, distances, etc.), la création de graphes ou de nuages de mots. Parmi les outils envisagés à développer se trouvent : un segmenteur de CV pour trouver leur différentes sections et un extracteur de termes à partir des CV et d'offres d'emploi.

---

<sup>1</sup> Adoc Talent Management est une société de ressources humaines situé à Paris, qui travaille sur le recrutement et sélection du personnel, la formation de doctorants et le R&D.

La liste des termes tels qu'ils ont été extraits et ordonnés automatiquement devra pouvoir être modifiée manuellement de telle sorte que l'on se donne la possibilité d'étudier l'impact de cette nouvelle sélection dans le processus de *relevance feedback* [1,4].

Les outils pourront être implémentés dans différents langages, notamment Python, Perl, R, HTML et PHP. La partie recherche de ce stage est susceptible de donner lieu à des publications dans un congrès international.

Ce stage se déroulera entre le LIA, à Avignon, et Adoc Talent Management, à Paris, au sein de son pôle R&D. Ainsi, il existe la possibilité de continuer en thèse ces travaux dans le cadre d'une convention CIFRE, en fonction des résultats obtenus.

## Références

- [1] R. Kessler, N. Béchet, M. Roche, J.-M. Torres-Moreno, and M. El-Bèze, 2012. A hybrid approach to managing job offers and candidates. *Information Processing & Management* 48(6), 1124–1135.
- [2] R. Kessler, N. Béchet, J.-M. Torres-Moreno, M. Roche, and M. El-Bèze, 2009. Job offer management: how improve the ranking of candidates. In *Foundations of Intelligent Systems*, 431–441. Springer Berlin Heidelberg.
- [3] R. Kessler, J. M. Torres-Moreno, and M. El-Bèze, 2008a. E-gen: Profilage automatique de candidatures. *TALN 2008*, Avignon, France, 370–379.
- [4] L. A. Cabrera-Diego, B. Durette, M. Lafon, J.-M. Torres-Moreno, M. El-Bèze, 2015. How Can We Measure the Similarity Between Résumés of Selected Candidates for a Job? *Proceedings of the 11th International Conference on Data Mining (DMIN'15)*, Las Vegas, 99–106.
- [5] L. A. Cabrera-Diego, J.-M. Torres-Moreno, M. El-Bèze, 2013. SegCV : traitement efficace de CV avec analyse et correction d'erreurs. *TALN 2013*, Les Sables d'Olonne, France, 707–714