

# Analyse des points de vue politiques

**Durée:** de 2 à 5 mois

**Lieu:** Orsay, France

**Mots-clés:** les tâches pourront être adaptées au niveau de qualification du ou de la candidate dans les domaines suivants : **Traitement Automatique des Langues, Fouille de texte, Extraction d'information, Recherche d'information, Gestion de données sociales, Bases de données.**

La personne choisie travaillera sur un projet collaboratif impliquant l'équipe Oak de l'INRIA<sup>1</sup>, le LIMSI-CNRS<sup>2</sup> et le quotidien Le Monde<sup>3</sup>.

## 1 Contexte

L'arène politique française comprend de nombreux partis s'étendant de l'extrême-gauche à l'extrême-droite. De plus, comme dans toutes les démocraties, les idées et les valeurs s'entremêlent avec les tactiques de groupe, les ambitions personnelles et les stratégies de communication, conduisant parfois à des distorsions entre affiliations politiques et conviction personnelle.

Pour toutes ces raisons, décider quel candidat est le plus proche de nos propres opinions peut être complexe pour le citoyen, et mettre les choses en perspectives est un travail difficile pour le journaliste.

Le but de ce projet est de **construire automatiquement, à partir de différentes sources comme les articles de journaux, les flux Twitter, les bases de données structurées, etc.** des fils thématiques permettant d'organiser et de visualiser les **déclarations faites par les politiciens**, dans le but d'aider les journalistes et les citoyens à décoder la vie politique.

**Sources de données.** Nous utiliserons différentes sources textuelles et/ou structurées :

- Articles de journaux du journal Le Monde ;
- Textes en provenance des réseaux sociaux, ainsi que les métadonnées associées (auteur, date, retweets, etc.) ;
- La connaissance des affiliations politiques des personnalités, ainsi que leur position sur l'échiquier politique ;

**Domaines scientifiques.** Ce projet implique les domaines du **traitement automatique des langues, la fouille de texte, l'extraction d'information, la recherche d'information, la gestion de données sociales, les bases de données.**

## 2 Description

Selon le niveau de qualification et la durée du stage, le ou la candidate travaillera sur une ou plusieurs des tâches suivantes :

- Améliorer et évaluer un système d'extraction de déclarations des personnalités politiques dans les articles de journaux (**extraction d'information à base de règles**).

---

<sup>1</sup><https://team.inria.fr/oak/>

<sup>2</sup><https://www.limsi.fr/index.en.html>

<sup>3</sup><http://www.monde.fr>

- Construire automatiquement ou semi-automatiquement une base de connaissance contenant des noms de personnalités politiques, leur compte Twitter/Facebook, leur affiliation, leur bord politique et différentes variation de leur nom (**acquisition de connaissance, gestion de contenu**).
- Construire un système de collecte de déclarations en provenance des réseaux sociaux et des pages personnelles.
- Construire manuellement un corpus de déclarations politiques, pour une utilisation future par des systèmes d'apprentissage automatique (**annotation de corpus**).
- Implémenter et adapter des algorithmes non-supervisés tels que les “topic models” (comme LDA) sur différents types de déclarations (**fouille de textes non supervisée**).

### 3 Contacts

- Ioana Manolescu (ioana.manolescu@inria.fr)
- Xavier Tannier (xavier.tannier@limsi.fr)