

# Stage de master 2

## Retrouver la source des trending topics sur les réseaux sociaux ou sur le web

**Mots-clés :** traitement automatique des langues, extraction d'information, propagation d'information, graphes, Twitter, réseaux sociaux.

**Date de démarrage :** Mars/avril 2017

**Durée :** 4-6 months

**Encadrant :** Xavier Tannier (LIMSI-CNRS)

**Lieu :** LIMSI, Orsay, Univ. Paris-Saclay<sup>1</sup>

### Contexte

Les réseaux sociaux tels que Twitter ou Facebook, mais également certains sites web, sont devenus des sources d'information et de désinformation massive. De nombreuses rumeurs y sont lancées, des faits y sont déformés ou manipulés, et ce dans de nombreux domaines, dans le but de nuire à des personnes ou à des organisations, mais également de servir des idées politiques.

Le "fact-checking" est une discipline consistant à vérifier la véracité des déclarations faites par des personnalités publiques ou des rumeurs qui se propagent, notamment sur les réseaux sociaux. Une tâche importante est de remonter jusqu'à la source d'une information, pour vérifier si cette source est unique ou multiple et si elle est digne de confiance ou pas.

### Description

Le travail proposé consiste à réaliser un système semi-automatique permettant, à partir d'un thème ou d'un document fourni par l'utilisateur, de remonter dans

---

<sup>1</sup><https://www.limsi.fr/en/access>

la mesure du possible jusqu'à la source de cette information (un tweet, un blog, etc.).

Il s'agira donc de construire un graphe de citation et de référence basé sur le contenu textuel des documents (principalement, issus de réseaux sociaux et de pages web). Des outils déjà existants, comme un extracteur de citations et de sources dans les articles, pourront être utilisés.

L'automatisation complète d'un tel système étant un problème très complexe, nous considérons que la construction de ce graphe pourra être guidée par l'utilisateur, qui validera ou invalidera les propositions du système, guidant ainsi la progression vers la source supposée de l'information.

## Profil

Nous recherchons un(e) étudiant(e) intéressé(e) par le traitement de contenu en langage naturel et par la manipulation de données issue des réseaux sociaux. La personne retenue devra avoir des compétences solides en programmation et la volonté d'apprendre de nouveaux outils et de nouvelles approches. Elle manipulera les API d'interrogation Twitter et/ou Facebook ainsi que des outils de traitement automatique des langues. Les compétences en programmation ne sont cependant pas le seul critère, et la personne retenue devra également faire preuve de créativité et d'esprit d'analyse.

Les candidatures doivent comporter :

- Une lettre de motivation
- Le nom de deux personnes référentes
- Un curriculum vitae (CV)

Le stagiaire retenu recevra une "gratification" (qui était de 546,01 € en 2016) ainsi que le remboursement de la moitié du pass "Navigo" ou "Imagine R".

Pour toute question ou candidature: [Xavier.Tannier\[at\]limsi.fr](mailto:Xavier.Tannier@limsi.fr)