

Proposition de stage Master 2 - période avril 2015

## Analyse de données semi-structurées dans un contexte de tandem homme-machine.

Clic and walk

December 2, 2014

### Contexte

La société Clic and Walk a développé un produit permettant de mettre en place de la collecte de données (image, texte fermé et/ou texte libre) dans un contexte géolocalisé. Le service repose sur une application mobile alimentée en missions via l'équipe marketing (au sein de Clic and walk) ou via un outil Saas mis à la disposition des clients finaux (entreprises, collectivités, etc...).

Une fois les missions réalisées par les 'ClicWalkers', un processus de validation et d'agrégation de données est nécessaire afin de produire des analyses et des rapports d'une forte valeur ajoutée. Ce processus de validation qui repose pour grande partie sur de la vérification humaine n'est pas viable dans le contexte d'une augmentation significative du volume des missions.

Ce stage interviendra donc au coeur d'un des axes critiques de la politique de développement de Clic and walk et au sein de l'équipe de R&D, actuellement composée d'un ingénieur et d'une doctorante en thèse Cifre.

### Objectifs du stage

L'objectif de ce stage est d'analyser, étudier et prototyper des solutions permettant d'améliorer les processus de validation et de consolidation de données qui sont au coeur du métier de Clic and walk. Plusieurs axes pourront être explorés dont certains potentiellement en collaboration avec l'équipe Magnet de L'Inria.

Le stage portera principalement sur les aspects suivants :

- Analyse de textes courts et proposition de tags via un moteur de recommandation answer-to-tag.
- Classification d'images à rejeter et détection d'objets avec l'aide du contexte textuel.
- Mise en place d'un processus de validation/rejet basé sur l'apprentissage par renforcement

Suivant le niveau et l'appétence de l'étudiant(e) pour les techniques d'apprentissage semi-supervisées, le sujet pourra être redirigé vers l'utilisation d'auto-encodeurs et des travaux basés sur l'apprentissage d'espaces latents pour la représentation de mots et/ou d'images (GloVe, Word2Vec et NIC).

## Compétences

Les domaines explorés par ce sujet de stage étant larges, la première compétence attendue est une capacité d'analyse et de veille technique aussi bien avant que pendant la période du stage. Il est attendu que l'étudiant(e) soit force de proposition et intègre dans son argumentaire la viabilité des solutions en terme de complexité et d'intégration dans la structure actuelle du système d'information de l'entreprise.

Les compétences souhaitées sont :

- La maîtrise d'un langage orienté objet et d'un langage scripté (de préférence Java et Python).
- Des connaissances pointues en analyse de données (supervisée et non supervisée).
- Une appétence pour la réalisation de prototypes orientés vers une approche Minimum Viable Product.

La connaissance préalable d'Elasticsearch, FlaskApi, Scikit-learn et/ou Mahout sont appréciés.

## Renseignements complémentaires

Merci de prendre contact avec Guillaume André et Thomas Pocreau  
guillaume/thomas at clicandwalk.com