

Offre de thèse

Laboratoires : IRIT (Toulouse) et CEA Tech LIST/DIASI/LVIC

Directeur de thèse : Mohand BOUGHANEM – IRIT

Encadrement CEA Tech : Romaric BESANCON – CEA Tech LIST/DIASI/LVIC

Contact : boughanem@irit.fr

Sujet de thèse

Titre :

Analyse et synthèse d'opinions sur le web et les réseaux sociaux : intégration de caractérisations fines des opinions, de leur contexte spatio-temporel et de l'importance sociale de l'émetteur pour la mise en évidence de tendances d'opinion

Description :

Le web et les réseaux sociaux offrent une incroyable plate-forme d'échange d'information, largement plébiscitée par de nombreux utilisateurs. Cette quantité d'information est également une grande opportunité pour les systèmes automatiques de fouille de données et d'analyse de flux d'information, avec des applications dans le cadre, par exemple, de la cyber-sécurité ou du marketing. Le sujet de cette thèse concerne plus précisément le domaine de l'analyse d'opinion sur ces plateformes, pour détecter de façon automatique des tendances d'opinion sur des sujets donnés : dans le domaine de la cyber-sécurité, on s'intéressera par exemple aux réactions concernant des personnes ou des événements, avec une localisation spatiotemporelle ciblée, dans le domaine commerciale, aux opinions sur certaines marques ou produits.

L'analyse d'opinion est un domaine de recherche en plein essor et a fait l'objet de nombreuses études, intégrant à la fois des technologies de traitement automatique des langues pour la normalisation des textes, des techniques à base de lexiques d'opinion ou d'apprentissage automatique pour attribuer automatiquement une opinion à un texte [1]. Au delà de la simple attribution d'une polarité d'opinion à un segment de texte, une analyse plus fine est souvent nécessaire pour avoir une meilleure caractérisation de l'opinion (trouver sur quel objet ou sur quel trait précis de l'objet porte l'opinion). Même si plusieurs approches ont été testées avec des modèles à base de règles [2] ou statistiques [3], le sujet reste ouvert, en particulier pour les documents courts comme des messages ou des tweets.

De façon complémentaire, une autre caractéristique des réseaux sociaux est l'organisation structurée des relations entre les utilisateurs au sein des communautés (les « amis » sur Facebook, les « followers » et les « re-tweets » sur Twitter, les systèmes de commentaires sur différentes plate-formes). Cette structuration peut être exploitée pour trouver des informations sur l'importance sociale des émetteurs d'information, comme sa popularité ou sa réputation [4,5].

L'idée de cette thèse est de combiner des technologies d'analyse d'opinion sur les textes, en apportant des éléments innovants sur une analyse plus fine permettant de mieux caractériser l'opinion et des technologies spécifiques d'analyse des réseaux sociaux pour déterminer l'importance sociale de l'émetteur, de façon à produire un résumé (une synthèse) consolidé de tendances d'opinion sur des sujets ou personnes en fonction de leur contexte spatiotemporel.

Bibliographie :

- [1] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1-135, January 2008.
- [2] Stefan Gindl, Albert Weichselbraun, and Arno Scharl. Rule-based opinion target and aspect extraction to acquire affective knowledge. In *Proceedings of WWW '13 Companion*, 2013.
- [3] Kang Liu, Liheng Xu, and Jun Zhao. Opinion target extraction using word-based translation model. In *Proceedings of EMNLP-CoNLL '12*, Stroudsburg, PA, USA, 2012.
- [4] Min-Chul Yang, Jung-Tae Lee, Seung-Wook Lee, and Hae-Chang Rim. Finding interesting posts in twitter based on retweet graph analysis. In *Proceedings of ACM SIGIR*. ACM, 2012.
- [5] Ismail Badache and Mohand Boughanem. Social Priors to Estimate Relevance of a Resource (In *ACM Information Interaction in context (IliX)*). ACM, 2014.
- [6] Malik Muhammad Saad Missen, Mohand Boughanem, and Guillaume Cabanac. Opinion mining: reviewed from word to document level. *Social Network Analysis and Mining*, 3(1):107-125, 2013.

Profil / Compétences et formations requises :

- Titulaire d'un diplôme de Master (ou titre équivalent) en sciences informatiques
- Bonnes connaissances en programmation, notions TAL, mathématiques(statistiques)
- Compétences écrites et orales en anglais fortement souhaitées

Candidature :

Les candidats transmettront par email un CV, une lettre de motivation et le relevé de notes du Master ou de la dernière année d'école d'ingénieur à Mohand Boughanem (boughanem@irit.fr)

Accueil et encadrement :

Le doctorant sera inscrit à l'université Paul Sabatier de Toulouse. Le travail de thèse se déroulera principalement au laboratoire IRIT à Toulouse, tout en étant en lien avec les équipes CEA Tech (Saclay et Toulouse).

Financement : Financement région Contrat Laboratoire Entreprise CEA-IRIT

Durée : 36 mois

Début : 1 septembre 2015 ou 1 octobre 2015

Adresse du laboratoire IRIT :

Université Paul Sabatier-IRIT, 118 Route de Narbonne 31062