

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIS-2018-009**
(à rappeler dans toute correspondance)

Lieu : Palaiseau

Département/Dir./Serv. :

Tél. : 01 80 38 65 65

DTIS/PSR

Responsable du stage : Valentina Dragos

Email : valentina.dragos@onera.fr

DESCRIPTION DU STAGE

Domaine d'étude : Knowledge engineering, semantic annotation, ontology

Type de stage Fin d'études bac+5 Master 2 recherche Bac+2 à bac+4

Intitulé : Semantic annotation for opinion mining in social media

Sujet : Semantic annotation [4] is the process of attaching meta data or additional information to paragraphs, documents or any other type of content by linking background information to concepts unambiguously defined by formal models, such as ontologies. When the content is annotated it becomes a resource that is easier to analyze, combine and reuse by algorithms and automatic processing chains.

The goal of this stage is to develop, implement and evaluate an approach allowing the semantic annotation of corpora gathered from social media in order to detect opinions. Annotations will not be embedded into documents, but instead will be built in the form of quadruple sets highlighting the subject (paragraphs, document or the whole corpora), predicate (e.g. ontological relations), object (e. g. topic, concept) and context (author and time of annotation). The annotation base will be further used to evaluate the precision of algorithms developed for automatic opinion detection for PRACTICES (H2020) project.

Opinion mining is a research direction developed as an increasing number of online platforms allow users, companies or organizations to freely express their own views and to create, share or view information. Such data become a valuable resource for various applications (e.g. political trend analysis [5]) and tools were developed to browse the amount of data [6]. Methods rely on both natural language processing methods and semantics [1] because the task does not require in-depth understanding of each sentence or document, but rather seeks to identify some aspects conveyed: entities, areas and topics mentioned, along with a positive or negative attitude expressed by the author. While semantic annotating improves the overall analysis [3], developing effective tools still faces several challenges [2], among which the assessment of results by using reliable ground truth is an important one. The work to be done will use several domain and opinion-oriented ontologies in order to build an annotation base and has the following milestones:

- State of art on methods and tools for social data annotation
- Text mining techniques for concept recognition from text
- Generation of annotations and construction of annotation base
- Proof of concept and experiments for opinion detection

References

- [1]Bontcheva, K., & Rout, D. (2014). Making sense of social media streams through semantics: a survey. *Semantic Web*, 5(5), 373-403.
- [2]Maynard, D., Bontcheva, K., & Rout, D. (2012). Challenges in developing opinion mining tools for social media. *Proceedings of the @ NLP can u tag# usergeneratedcontent*, 15-22.
- [3]Maynard, D., Greenwood, M. A., Roberts, I., Windsor, G., & Bontcheva, K. (2015, June). Real-time Social Media Analytics through Semantic Annotation and Linked Open Data. In *Proceedings of the ACM Web Science Conference* (p. 46). ACM.
- [4]Bontcheva, K., & Cunningham, H. (2011). Semantic annotations and retrieval: Manual, semiautomatic, and automatic generation. In *Handbook of semantic web technologies* (pp. 77-116). Springer Berlin Heidelberg.
- [5]Maynard, D., & Funk, A. (2011, May). Automatic detection of political opinions in tweets. In *Extended Semantic Web Conference* (pp. 88-99). Springer, Berlin, Heidelberg.
- [6]Faridani, S., Bitton, E., Ryokai, K., & Goldberg, K. (2010, April). Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1175-1184). ACM.

Est-il possible d'envisager un travail en binôme ? NON

Méthodes à mettre en oeuvre :

- | | |
|---|---|
| <input type="checkbox"/> Recherche théorique | <input checked="" type="checkbox"/> Travail de synthèse |
| <input checked="" type="checkbox"/> Recherche appliquée | <input type="checkbox"/> Travail de documentation |
| <input type="checkbox"/> Recherche expérimentale | <input checked="" type="checkbox"/> Participation à une réalisation |

Possibilité de prolongation en thèse : NON

Durée du stage : Minimum : 4 mois Maximum : 6 mois

Période souhaitée : march – august 2018

PROFIL DU STAGIAIRE

Connaissances et niveau requis :
JAVA, OWL, Knowledge modeling

Ecoles ou établissements souhaités :
Université, grande école