

Analyse sémantique et spatialisée des sentiments exprimés dans les Cahiers citoyens

Lors du Grand débat national (GDN), des Cahiers de doléances renommés ensuite Cahiers citoyens (CC) ont été déposés dans les mairies et remplis librement par les personnes qui le souhaitent. Ces Cahiers représentent un corpus à la fois considérable et précieux car complémentaire des corpus issus des plates-formes numériques ouvertes à l'occasion du GDN. Ils ont été numérisés puis transcrits sous l'égide de la BNF et conservés aux Archives nationales ; leur analyse portant à la fois sur la dimension géographique et sémantique permettrait d'avoir un regard ancré dans un contexte géographique de la parole des citoyen.nes. Le stage proposé vise à faire un pas dans cette direction à l'aide de méthodes du traitement automatique des langues et de modèles de langues. Le travail portera sur le développement d'une analyse de sentiments des CC qui tiendra compte du bruit dans le texte issu de la transcription et d'une nécessité d'anonymisation. Les objectifs du stage sont particulièrement :

- faire l'état de l'art des approches pour l'analyse de sentiments (incluant perceptions, émotions et opinions),
- mettre en oeuvre une approche d'analyse de sentiments des contributions contenues dans les Cahiers,
- spatialiser les contributions à l'aide de leurs métadonnées et d'un modèle existant pour identifier les informations de localisation contenues dans les contributions (par ex. CamemBERT-NER)
- définir des consignes d'annotation en vue de constituer des jeux d'évaluation, et d'évaluer la pertinence de l'approche proposée.

Bibliographie (à titre indicatif)

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.

Ploux, S. Genay, M., Ploux-Chillès, L. (2021) Les mots du Grand Débat national : les réseaux lexicaux des contributions déposées sur trois plateformes, *Humanités numériques*. DOI:

<https://doi.org/10.4000/revuehn.2655>

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335-4385.

Dominguès, C. Jolivet, L. Brando, C. Cargill, M. (2019). Place and Sentiment-based Life story Analysis, *Revue française des sciences de l'information et de la communication* [DOI:

<https://doi.org/10.4000/rfsic.7228>

Compétences et aptitudes recherchées

Maîtrise du langage de programmation Python (obligatoire)

Connaissances en traitement automatique des langues et apprentissage automatique (obligatoire)

Appétence pour les sciences sociales

Gratification : environ 650€/mois (4,05€/h), 50 % d'un abonnement de transport (ex. Navigo) remboursés et accès au restaurant administratif à un tarif préférentiel (environ 2,5€ par repas).

Localisation : 54 boulevard Raspail, 75006 Paris et 73 avenue de Paris 94160 Saint-Mandé ; télétravail partiel possible.

Durée : 5 mois.

Comment candidater ?

Envoyer un CV et une lettre de motivation à Carmen Brando (carmen.brand@ehess.fr), Sabine Ploux (sabine.ploux@ehess.fr) et Catherine Dominguès (catherine.domingues@ign.fr)