

Titre: Comprendre et structurer les informations venant du site de fact-checking Full Fact

Encadrement:

Konstantin Todorov, todorov@lirmm.fr,

Résumé: Les fausses nouvelles (fake news) sont devenues un problème de plus en plus important, tant du point de vue de la société que de celui de la recherche. Le LIRMM en collaboration avec 2 équipes allemandes a construit et mis à disposition la base de connaissance ClaimsKG qui recueille les informations et méta-données provenant d'un grand nombre de sites journalistiques de fact checking. Le sujet de TER consiste en l'enrichissement de cette base de connaissances avec des nouvelles données provenant du site FullFact, qui a une structure particulière et nécessite pour cette raison un traitement à part.

Sujet: Les fausses nouvelles (fake news) sont devenues un problème de plus en plus important, tant du point de vue de la société que de celui de la recherche. De nombreuses approches récentes dans diverses communautés scientifique portent sur des problèmes tels que la vérification des faits, la détection de la pertinence ou de point de vue des documents par rapport à des revendications particulières.

Le LIRMM en collaboration avec 2 équipes allemandes (L3S Hannover et l'institut de sciences sociologiques GESIS à Cologne), a construit et mise à disposition la base de connaissance ClaimsKG¹ [1] qui recueille les informations et méta-données provenant d'un grand nombre de sites journalistiques de fact checking, tels que Politifact² ou Snopes³. ClaimsKG est un graphe de connaissances d'assertions annotées et liées qui facilite la création de requêtes structurées sur les assertions, leurs valeurs de vérité (True, Mostly False, etc.), leurs auteurs, date de publication, etc. ClaimsKG est généré par un pipeline entièrement automatisé qui collecte des assertions et des métadonnées à partir des sites de fact-checking, transforme les données en graphes de connaissances selon un modèle établis, et annote les assertions avec des entités DBpedia (Wikipedia). La base actuelle comprend plus de 32,000 assertions publiées depuis 1996 et est mis à jour régulièrement.

¹ <https://github.com/claimskg>

² <https://www.politifact.com>

³ <https://www.snopes.com>

Le sujet de TER consiste à intégrer à ClaimsKG le site de fact-checking britannique Full Fact⁴. Ce site a une structure particulière, où les assertions sont regroupés selon des “histoires”. Le but du TER sera d’identifier les assertions individuelles dans chaque histoire, ainsi que leur label de véracité et par la suite identifier les relations entre les assertion dans chaque histoire, ainsi que les relations entre histoire. Les données produites par cette étude seront intégrées à la base de connaissance ClaimsKG.

Références:

[1] Tchechmedjiev, A., Fafalios, P., Boland, K., Gasquet, M., ... & Todorov, K. (2019). ClaimsKG: A Knowledge Graph of Fact-Checked Claims. In *International Semantic Web Conference* (pp. 309-324). Springer.

⁴ <https://fullfact.org/>