

Master thesis : Extending Joint Entity and Event Coreference Resolution across Documents

Auteur : Nelissen, Louis

Promoteur(s) : Ittoo, Ashwin

Faculté : Faculté des Sciences appliquées

Diplôme : Master : ingénieur civil en science des données, à finalité spécialisée

Année académique : 2021-2022

URI/URL : <http://hdl.handle.net/2268.2/16441>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Extending Joint Entity and Event Coreference Resolution across Documents

Louis Nelissen

Supervisor: Prof. A. Ittoo

MSc in Data Science and Engineering

Academic year 2021-2022

Abstract

Detecting corefering events and entities in texts is an important task in NLP, where it plays a role in many other tasks and applications. In this work, we build on a joint approach of entity and event coreference resolution, pioneered by H. Lee, Recasens, et al. 2012 and matured by Barhom et al. 2019 using a neural architecture. In particular we look at coreference resolution across documents, more complicated and less researched than coreference resolution within documents. Using Barhom et al. 2019’s model, we propose a series of extensions to improve its results. This is done by increasing the amount of information provided to the model, in particular the joint nature of the modelling and by improving entity and event representation with the use of document embedding. As a secondary problem, we investigate ways to improve the model’s time performance through compressing the mention representations. Our results are compared with other works tackling the problem of cross document coreference resolution on the ECB+ dataset, the standard dataset for cross document entity and event coreference resolution.