HAL open science

Traitement du langage et modélisation de relations pour la représentation unifiée de documents narratifs

Arthur Amalvy, Vincent Labatut, Richard Dufour

To cite this version:

Arthur Amalvy, Vincent Labatut, Richard Dufour. Traitement du langage et modélisation de relations pour la représentation unifiée de documents narratifs. Meetup LIAvignon, Nov 2022, Avignon, France. 2022. hal-03832543

HAL Id: hal-03832543

https://hal.science/hal-03832543

Submitted on 2 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Traitement du langage et modélisation de relations pour la représentation unifiée de documents narratifs

Arthur Amalvy[1], Vincent Labatut[1], Richard Dufour[2]

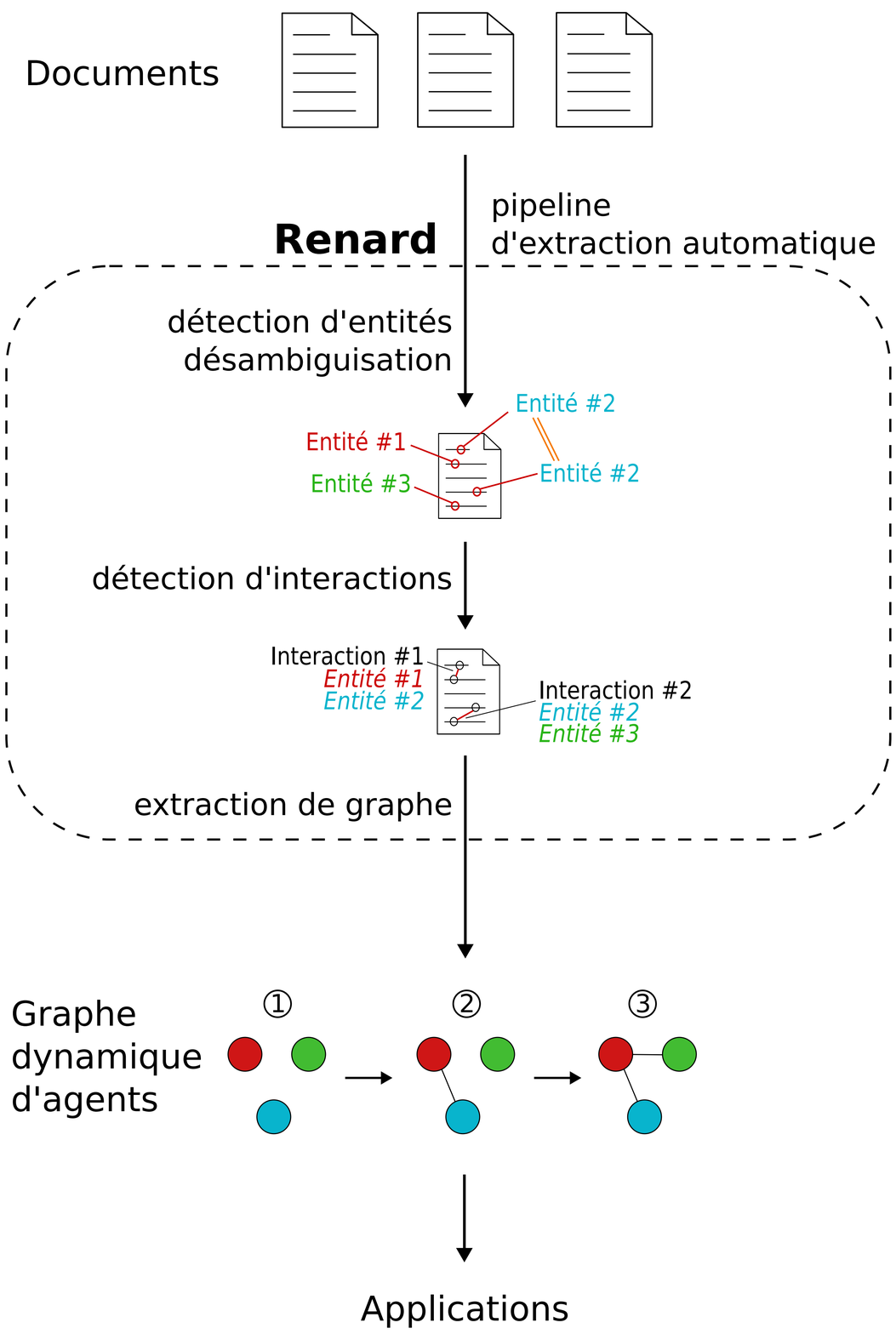
(1) Laboratoire Informatique d'Avignon (2) Laboratoire des Sciences du Numérique de Nantes

Comment représenter un document à grande échelle ?

Les approches récentes de plongements (Word2Vec, BERT, SentenceBERT...) permettent la représentation vectorielle de mots ou de phrases.

Comment représenter des documents (romans, articles journalistiques...) à grande échelle ?
→ Représentation par des graphes d'agents.

Développement de **Renard**, un pipeline automatique modulaire permettant l'extraction automatique de ces graphes.



Applications

La représentation sous forme de graphe peut être convertie en vecteur par plongement.

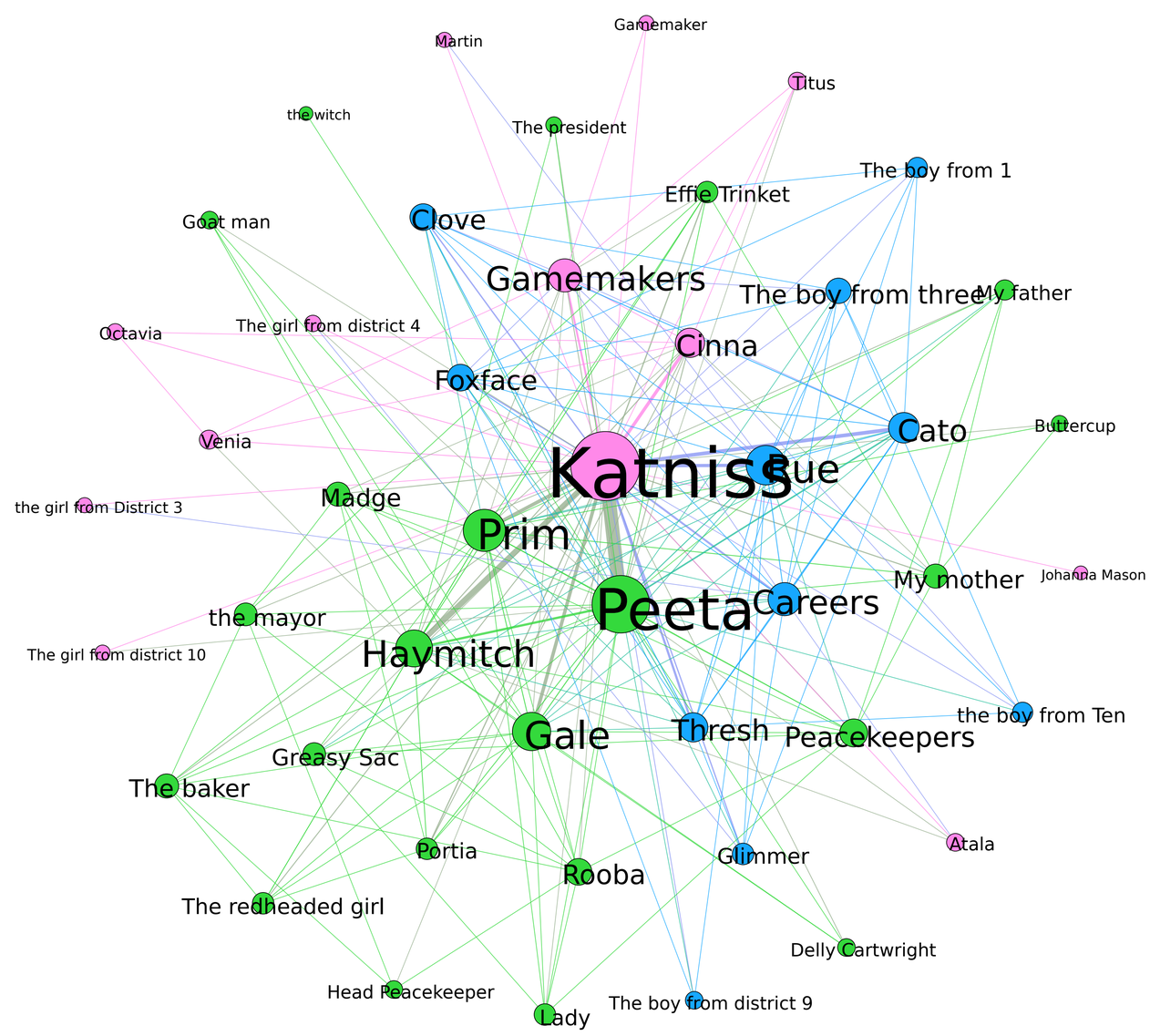
Utilisation de ce vecteur pour résoudre diverses tâches (classification, recommandation, segmentation, tagging...).

Qualité de la représentation

La qualité de la représentation finale peut être définie comme la performance qu'elle permet d'obtenir sur des tâches diverses.

Quelle est l'influence des erreurs de traitement du langage de **Renard** sur la qualité de la représentation ?

Quelles informations intégrer dans le graphe pour obtenir la meilleure qualité possible ?



Réseau extrait de *The Hunger Games* par **Renard**

Publications

- A. Amalvy, V. Labatut et R. Dufour. "*BERT meets d'Artagnan: Data Augmentation for Robust Character Detection in Novels*". *Workshop on Computational Methods in the Humanities.* 2022.

- A. Amalvy, V. Labatut et R. Dufour. "*Remplacement de mentions pour l'adaptation d'un corpus de reconnaissances d'entités nommées à un domaine cible*". *29e Conférence sur le Traitement Automatique des Langues Naturelles.* 2022, p197-204.

Arthur Amalvy

arthur.amalvy@univ-avignon.fr

Doctorant au Laboratoire Informatique d'Avignon depuis Septembre 2021

HAL