

Quanteda

Présentation d'un outil de recherche numérique innovant

Laurianne Martin

POL6078 - Outils numériques de recherche
Université Laval

5 décembre 2024

Plan de la présentation

1. Point d'observation
2. Arpentage et choix éditorial
3. Démonstration

Point d'observation

- QUantitative ANalysis of TExtual DAta
- Outil d'analyse des données/visualisation des données
- Créé par Kenneth Benoit, Kohei Wantanbe et collègues en 2018
- Financé par le Conseil européen de la recherche
- Pourquoi c'est un outil?
- Plusieurs fonctions et *work in progress* avec repo Github
- `quanteda`, `quanteda.textmodel`, `quanteda.textstats`, `quanteda.textplots`, `quanteda.sentiment`, `quanteda.tidy`

Arpentage et choix éditorial

Table: Résumé des principaux packages R en traitement de données textuelles

Critères	quanteda	tidytext	tm	text
Accessibilité	Oui	Oui	Oui	Oui
Existence d'une communauté d'utilisateurs	Grande	Très grande	Modérée	Croissante
Popularité dans le champ	Grande	Très grande	En décroissance	Variable
Compatibilité avec d'autres outils	Très bonne	Excellente	Bonne	Bonne
Transparence et répliquabilité	Oui	Oui	Oui	Oui
Adaptabilité et flexibilité	Très grande	Très grande	Grande	Grande

Démonstration

On code!

Références et pour aller plus loin

- **Nettoyage des données :**
<https://www.youtube.com/watch?v=OVXLyQKNNaY> (début du vidéo)
- **Text mining :** <https://www.youtube.com/watch?v=OVXLyQKNNaY> (fin du vidéo)
- **Autre chaîne Youtube avec tutoriels :**
https://www.youtube.com/playlist?list=PL-i7GM-A1wBZYRYTpem7hNVHK3hSV_1lt
- **Tutoriels écrits :** <https://tutorials.quanteda.io/>,
<https://joss.theoj.org/papers/10.21105/joss.00774>
- **Repo Github :** <https://github.com/quanteda>
- **Cheat sheet :**
<https://github.com/quanteda/quanteda/blob/master/tests/cheatsheet/quanteda-cheatsheet.pdf>
- **Attention, les tutoriels ne sont peut-être plus à jour, utiliser ChatGPT et de la patience pour faire fonctionner le code**