

Extraction de relations

Plan de présentation

- I/ Introduction et définitions
- II/ Domaine d'utilisation
- III/ Méthodes d'extraction
- IV/ But de l'application
- V/ Architecture de l'application
- VI/ Résultats et méthodes d'évaluations
- VII/ Démonstration
- VIII/ Conclusion

Introduction et définitions

- Qu'est que signifie « extraction de relations » ?
- Domaine plus large: « Extraction d'informations »
- Plusieurs difficultés:
 - Grammaire complexe
 - Ambiguïtés (mots polysémiques par exemple)
 - Relations implicites (des non-dits, etc, ...)

Domaine d'utilisation

- Application bibliographique (extraire des « pdf » pertinents en fonction des données utilisateurs)
- WordNet (réseau lexical)
- Génération de résumé (il faut savoir analyser les relations extraites pour en titrer les relations pertinentes d'un texte)
- Moteur de recherche (synonymie de mots clés, etc)

Méthodes d'extraction

- Extraction de motif (pattern) qu'on appelle aussi « Is-a »
- Permet de relier des termes très simplement par l'utilisation d'expressions régulières

Hearst's Patterns for extracting IS-A relations

Hearst pattern	Example occurrences
X and other Y	...temples, treasures, and other important civic buildings.
X or other Y	Bruises, wounds, broken bones or other injuries...
Y such as X	The bow lute, such as the Bambara ndang...
Such Y as X	... such authors as Herrick, Goldsmith, and Shakespeare.
Y including X	...common-law countries, including Canada and England...
Y, especially X	European countries, especially France, England, and Spain...

Méthodes d'extraction

- Cause à effets (Que se passe-t-il ? Pourquoi est-ce arrivé et que va-t'il se passer ensuite ?)
- Décrire des événements en partant d'un motif (exemple: A [déclenché par] B provoque C)
- Méthode à partir de la classification supervisée

Méthodes d'extraction

- Réseaux de neurones (exemple Graphrel)
- Dédire des relations à partir des réseaux de neurones
- Graphrel utilise ce principe (introduction de pondération entre les relations, qualité d'une relation)
- Avantage: du deep learning pourra détecter des relations contextuelles

But de l'application

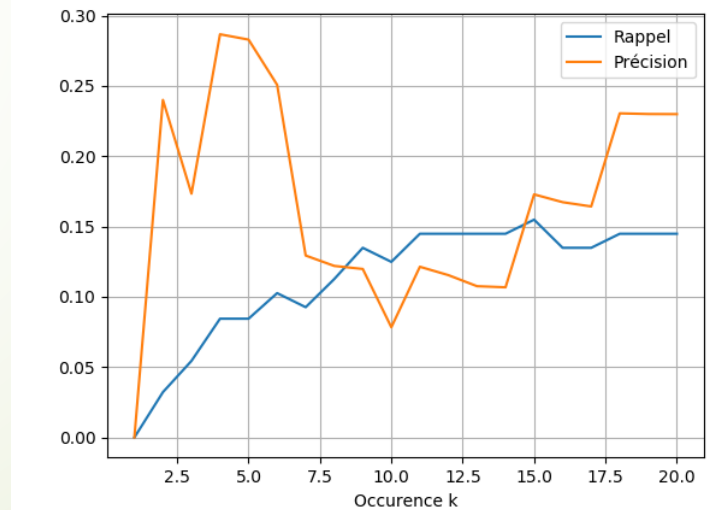
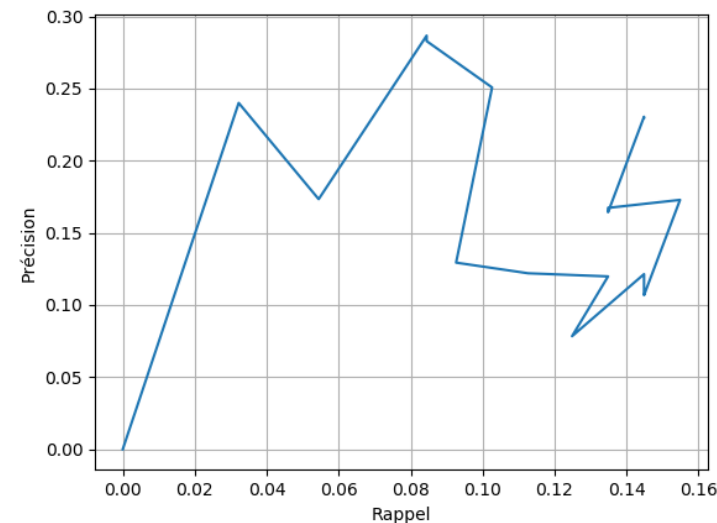
- Faire une liste de tous les personnages
- Appliquer la recherche de coréférences par réseau de neurones
- Recherches de relations entre les personnages à partir d'un fichier qui contient les catégories de relation (champs lexicaux)

Architecture de l'application

- Méthode de recherche: recherche d'un motif « IS-A »
- Technologies: spacy et neuralcoref (Python 3.7)
- Extraire les personnages
 - Par entités nommées
 - Utilisation des coréférences pour davantage de données
 - Éviter les doublons / noms nominatifs (« Mr » Henry) et privilégier nom complet
- Système de motifs par décalage de « k-mots » avec identification d'un terme d'un champ lexical
- Exemple simple: « Alice est la sœur de Bob ».

Résultats et méthodes d'évaluations

- Résultats sur 5 corpus (debug inclus)
- Occurrences décaler de 20 (prend beaucoup de temps)
- Graphes pas assez pertinents pour analyser leur nature (étude sur 5 corpus seulement). Cependant, on peut constater que le nombre d'occurrence ne doit pas être trop grand, auquel on perd une grande précision.



Démonstration

- Démonstration sur spyder
- Utilisation de textes extraits du projet Gutenberg et annotés manuellement
- Exécution de la coréférence lente

Démonstration

Limites de l'application

- Problèmes de formatages des noms propres – Problèmes de détection des personnages (doublons, noms composés avec ou sans accent, ...) ou problème de différenciation des noms
- Coréférences non contrôlées (des erreurs apparaissent) et temps d'attente trop longues
- Extraction de relations purement lexicales. (ne peut pas déterminer de relations à partir d'un contexte implicite ou des non-dits) => nécessité d'utiliser du deep learning avancé.
- Pauvreté des corpus (annotée personnellement et nombres insuffisants)

Conclusion