

01

Named Entity Recognition (NER)

Présenté par:
KARMOUT Ismail

Date : le 19 mai 2023

PLAN

- INTRODUCTION
- LES DIFFÉRENTES APPROCHES DU NER
- ÉTAPES DU PROCESSUS DE NER
- EXEMPLES D'APPLICATIONS DU NER
- LES OUTILS POUR LE NER
- LES LIMITES DU NER
- IMPLÉMENTATION
- CONCLUSION

Introduction

Le Named Entity Recognition (NER) est une technique de traitement automatique du langage naturel (TALN) qui vise à extraire des informations structurées à partir de textes non structurés en identifiant et en classant les entités nommées telles que les noms de personnes, les noms de lieux, les noms d'organisations, les dates, les montants d'argent, etc.

Le but du NER est d'extraire ces informations pertinentes à partir de textes afin de les rendre utilisables pour une analyse ultérieure.

Les différentes approches du NER

RÈGLES ET EXPRESSIONS RÉGULIÈRES

Les règles et les expressions régulières sont l'une des approches les plus anciennes et les plus simples pour le NER. Cette approche consiste à définir des règles grammaticales et des expressions régulières pour extraire les entités nommées à partir de texte non structuré.

STATISTIQUES ET APPRENTISSAGE AUTOMATIQUE

L'approche statistique et d'apprentissage automatique est l'une des approches les plus populaires pour résoudre le problème de la reconnaissance d'entités nommées. Cette approche consiste à entraîner un modèle à partir de données étiquetées et à l'utiliser pour identifier les entités nommées dans de nouveaux textes.

Étapes du processus de NER

LA SEGMENTATION DE TEXTE

La segmentation de texte est la première étape du processus de NER. Elle consiste à diviser un texte en phrases ou en segments plus petits afin de faciliter l'identification des entités nommées. Cette étape est importante car les entités nommées peuvent souvent être identifiées au niveau de la phrase ou du segment plutôt que du texte complet.

LE PART-OF-SPEECH TAGGING

La deuxième étape du processus de NER est le part-of-speech (POS) tagging, qui consiste à associer à chaque mot de la phrase une étiquette grammaticale telle que nom, verbe, adjectif, pronom, etc. Cette étape est importante car elle aide à comprendre la structure grammaticale de la phrase et fournit des informations sur les relations entre les mots.

LA RECONNAISSANCE DES ENTITÉS NOMMÉES

La reconnaissance des entités nommées (NER) est la dernière étape du processus de NER. Cette étape consiste à identifier et extraire les entités nommées dans le texte segmenté et étiqueté en parties du discours (POS-tagged). Les entités nommées peuvent être des noms de personnes, d'organisations, de lieux, de dates, d'heures, de montants d'argent, de pourcentages, etc.

Exemples d'applications du NER

L'importance du NER est qu'il permet de structurer automatiquement des informations non structurées, qui peuvent être utilisées pour diverses applications, notamment :



RÉSUMÉ
DE TEXTE



ANALYSE DE
SENTIMENT



RECHERCHE
D'INFORMATIONS



SURVEILLANCE DES
MÉDIAS SOCIAUX



TRADUCTION
AUTOMATIQUE



RECHERCHE D'INFORMATIONS

La reconnaissance d'entités nommées est également très utile dans la recherche d'informations, où elle peut être utilisée pour extraire des informations pertinentes à partir de grands ensembles de données non structurées. Par exemple, Dans le domaine des médias sociaux, le NER peut être utilisé pour extraire des informations sur les personnes, les lieux et les événements à partir de tweets ou de messages Facebook. Cela peut aider à identifier les tendances, à surveiller les réactions du public à un événement spécifique ou à détecter les discussions sur un sujet particulier.



LE TRAITEMENT DU LANGAGE NATUREL (NLP)

le Traitement du langage naturel (NLP) est un domaine d'application important pour le NER. Le NER peut être utilisé pour extraire des informations précieuses à partir de documents textuels, tels que des entités nommées telles que des noms de personnes, des organisations, des lieux, des dates, etc.

Les outils pour le NER

● SPACY

Spacy est une bibliothèque NLP en Python qui fournit des modèles pré-entraînés pour l'analyse syntaxique, le part-of-speech tagging, et la reconnaissance des entités nommées. Elle dispose également d'outils pour entraîner des modèles personnalisés sur des données spécifiques.

● NLTK

NLTK (Natural Language Toolkit) est une bibliothèque NLP en Python qui fournit des outils pour la tokenization, le part-of-speech tagging, la lemmatization, et la reconnaissance des entités nommées. Elle est souvent utilisée pour l'apprentissage automatique appliqué au traitement du langage naturel.

● FLAIR

Flair est une bibliothèque de traitement du langage naturel en Python qui se concentre sur l'apprentissage profond pour la reconnaissance des entités nommées. Elle fournit des modèles pré-entraînés pour une variété de tâches de traitement du langage naturel, y compris la reconnaissance des entités nommées.

Les limites du NER

LES DIFFICULTÉS LIÉES À LA LANGUE FRANÇAISE

Les difficultés liées à la langue française sont l'une des limites importantes de NER. En effet, la langue française possède de nombreuses particularités qui la rendent complexe pour la reconnaissance des entités nommées. Par exemple, le français est une langue très riche en termes d'expressions et de structures grammaticales, ce qui rend la détection des entités nommées plus difficile. De plus, le français possède de nombreux homonymes et polysémyes, ce qui peut rendre la tâche de détermination de l'entité nommée plus difficile.

Implémentation

Conclusion



En conclusion, la reconnaissance des entités nommées est une tâche clé en traitement automatique des langues. Elle a de nombreuses applications dans divers domaines. Il existe plusieurs approches et bibliothèques Python pour effectuer cette tâche, mais il reste des limites et des enjeux éthiques à prendre en compte. Néanmoins, NER continue à être un domaine de recherche en évolution constante et est appelé à jouer un rôle de plus en plus important dans le traitement de l'information en langage naturel.

MERCi

THANK YOU FOR YOUR PATIENCE
FOR YOUR WORK