

Passionné pour l'informatique, NLP, Machine Learning, et Deep Learning

ZIJIAN WANG

Master NLP & Master Informatique & Ingénieur

EDUCATION

Sorbonne Université, Master Langue et Informatique

2018-2020

- Informatique, Linguistique, NLP, Machine Learning, Deep Learning

Sorbonne Université, Master Langue Française Appliquée

2016-2018

EXPERIENCES

Laboratoire STIH, Sorbonne Université / Stage NLP Entités Nommées

05/2020 - maintenant

- Détecter par des outils NLP des Entités Nommées dans les corpus multilingues.
- Réentraîner des modèles de langues, évaluer et visualiser des résultats.
- <https://projet.liris.cnrs.fr/aphns-magis/AtelierAssisesMAGIS2020.html>

Shanghai Data Exchange Center / Data Scientist et Java Développeur

07/2019 - 08/2019

- Management des bases de données avec Spring
- Test Logiciel
- Github: <https://github.com/WANGZijian1994/fuxing-tools>

PROJETS IMPORTANTS

Comparaison Sémantique et Phonétique (Thèse pour Master 1)

- Création des algorithmes pour trouver des paires minimales (phonologie).
- Découverte du WordNet Français pour trouver hyponymie et hyperonymie
- Création des services pour chercher des paires minimales par Spring.
- Youtube: <https://www.youtube.com/watch?v=fQY9d4JoKxE>

Solr Research

- Nettoyage des données de Ligue 1 pour 5 ans avec pandas
- Annotation et indexation du corpus
- Creation des services pour chercher par Solr Interface et par Spring

Analyse des sentiments des commentaires d'Amazon

- Scrapy et nettoyage des données, avec nltk et pandas
- Vectorisation et Word2Vec pour créer des modèles, avec tensorflow et sklearn
- Classification des sentiments avec modèles de Machine Learning
- Trouver 10 'hot topics' dans certains commentaires par l'apprentissage non-supervisé.

Reconnaissance des paroles

- Collection des données orales par Praat et Formatage des données par Python
- Entraînement d'un modèle HMM pour identifier des mots isolés

Algorithmes et Structure des données

- Participation à un projet OpenSource des exercices LeetCode
- Github: <https://github.com/AlgoStudyGroup/Leetcode>

CERTIFICATIONS

- Coursera SQL pour Data Science
- Coursera C++
- GRE scores: Verbal 152/170 Quantity 170/170 Writing 4/6

Paris, France

+33 07 69 22 16 16

wangzijian1994@hotmail.com

[https://www.linkedin.com/in/](https://www.linkedin.com/in/wang-zijian-260442165/)

[wang-zijian-260442165/](https://github.com/WANGZijian1994)

<https://github.com/WANGZijian1994>

Techniques

- Python
- Pandas
- NumPy
- Nltk
- Spacy
- Data Visualization
- Data Cleaning
- Tensorflow
- PyTorch
- selenium
- Bash
- Git
- MySQL
- neo4j cypher
- Solr
- Java
- Maven and Spring
- StanfordNLP
- C++
- JavaScript
- HTML CSS

Langues

- Français DALF C1 obtenu
- Anglais ILETS 6.5
- Chinois Langue maternelle
- Allemand, Espagnole, et Japonais -Notion