

Jacques Fize

DOCTEUR EN INFORMATIQUE

Villeurbanne, 69100, France

☎ (+33) 6 45 64 37 27 | ✉ jacques[dot]fize[at]gmail.com | 🏠 www.jacquesfize.com | 📷 Jacobe2169

Compétences

Recherche **Traitement Automatique du Langage Naturel (TALN)**
Data-mining, information retrieval, geoparsing, machine learning

Programmation Python, Java, C/C++, JAVA, R, PHP, LaTeX
SGBD MySQL, PostgreSQL (+PostGIS), Elasticsearch
Web HTML5/CSS, Javascript, PHP, Framework Web (Flask, Streamlit, Shiny, J2EE)
Langages Français (Natif), Anglais (Intermédiaire), Espagnol (Notions)

Formation

Doctorat en Informatique *Montpellier, France*
UNIVERSITÉ DE MONTPELLIER 2016 - 2019
• Sujet : Mise en correspondance de données textuelles hétérogènes fondée sur la dimension spatiale

Master en Informatique *Caen, France*
UNIVERSITÉ DE CAEN NORMANDIE 2014 - 2016
• Spécialisation Traitement Automatique des Langues et Traitement d'images

Licence en Informatique *Caen, France*
UNIVERSITÉ DE CAEN NORMANDIE 2010 - 2014

Baccalauréat Littéraire *Cherbourg-en-Cotentin, France*
JEAN-FRANCOIS MILLET HIGH SCHOOL 2010
• Spécialité Musique

Expérience

Université Lumière Lyon 2 *Lyon, France*
POST-DOCTORANT Dec. 2021 - Dec. 2022
Travail de recherche en détection de communautés dans le cadre du projet POIVRE coordonné par Julien Velcin.

Institut National des Sciences Appliquées (INSA) *Lyon, France*
POST-DOCTORANT Dec. 2020 - Mai 2021
Travail de recherche pour l'amélioration de la tâche de prédiction de lien dans les graphes.

Institut National des Sciences Appliquées (INSA) *Lyon, France*
POST-DOCTORANT Dec. 2019 - Dec. 2020
Développement de méthodes de Geoparsing (identification de lieux dans des textes) s'appuyant sur les avancées en Deep Learning.

Centre de coopération internationale en recherche agronomique pour le développement (CIRAD)
Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA) *Montpellier, France*
ALLOCATAIRE DE RECHERCHE Oct. 2016 - Oct. 2019
Sujet de thèse : Mise en correspondance de données textuelles hétérogènes fondée sur la dimension spatiale
L'objectif de la thèse est de proposer de nouvelles méthodes de mise en correspondance de données en se concentrant sur leur caractère hétérogène et massif. Les méthodes de mise en correspondance proposées reposent sur des représentations originales des documents ainsi que différentes mesures de similarité. Dans nos travaux, nous proposons une approche de mise en correspondance multidimensionnelle exploitant la dimension thématique et spatiale des données textuelles hétérogènes.

Orange Labs *Lannion, France*
STAGIAIRE Mars 2016 - Aout 2016
L'objectif du stage était de concevoir une base de connaissance générée à partir de différentes sources provenant du Linked Open Data (Wikidata, DBpedia, Geonames). Chaque entité stockée est associée avec une classe, des labels multilingues et d'autres attributs propres à sa catégorie (Personne, Organisation, Lieu). Enfin, pour interroger la base de connaissance, nous avons mis au point un langage spécifique (similaire au SQL).

L'objectif de ce stage était de développer une application web pour transférer des informations sensibles entre les services publics du département du Calvados.

Publications

REVUE

Matching heterogeneous textual data using spatial features

Journal

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2019

Intelligent Data Analysis Journal

Exploitation de l'hétérogénéité dans les données textuelles - Utilisation de données produites à Madagascar

Revue

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2017

J. Darmont, N. Grabar, O. Teste, Eds., About Variety in Humanities Big Data, Recherche d'information, document et web sémantique, Vol. 19, No. 1, ISTE OpenScience, London, UK, 2019.

ACTES DE CONFÉRENCES ET WORKSHOPS INTERNATIONAUX

Matching heterogeneous textual data using spatial features

Workshop

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2018

In the Proceedings of 13th International Workshop on Spatial and Spatiotemporal Data Mining (SSTD-18) - 2018 IEEE International Conference on Data Mining Workshops (ICDMW); p. 1389–1396

Cette article présente un processus de mise en correspondance de données textuelles hétérogènes selon la dimension spatiale.

Gemedoc : A Text Similarity Annotation Platform

Papier Démo

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2018

In the Proceedings of International Conference on Applications of Natural Language to Information Systems. LNCS, Springer, 2018; p. 333–336

Dans cet article, nous présentons une interface d'annotation de correspondance entre les documents d'un corpus.

GeoDict : an integrated gazetteer

Workshop

JACQUES FIZE, GAURAV SHRIVASTAVA

2018

In the Proceedings of Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017) : Association for Computational Linguistics; p. 31–41

Dans cet article, nous présentons un nouveau *gazetteer* (i.e. base de données géographique), Geodict, qui rassemble les données provenant de trois sources : Wikidata, Geonames, OpenStreetMap.

ACTES DE CONFÉRENCES ET WORKSHOPS NATIONAUX

Le grand débat national, une aide pour prendre des décisions locales ?

Conference

JACQUES FIZE, LUCILE SAUTOT, MARTIN LENTSCHAT, LUDOVIC JOURNAUX, MOHAMED HILAL

2021

Le Grand Débat National, décidé par Emmanuel Macron début 2019 pour répondre au mouvement social des Gilets Jaunes, a permis de collecter les contributions de citoyens sur la transition écologique via une plateforme en ligne. Dans cet article, nous exploitons le corpus constitué par ces contributions pour identifier des zones où les participants demandent le développement de pistes cyclables et d'équipements ferroviaires. Pour cela, nous avons créé un modèle de classification permettant d'identifier les contributions traitant de la thématique du transport et proposé une méthode d'extraction de motifs traduisant les propositions des contributeurs.

Qui a peur du changement climatique ?

Conference

LUCILE SAUTOT, ERIC CHRAIBI, JACQUES FIZE, SÉBASTIEN PEILLET, LUDOVIC JOURNAUX, FLAVIE CERNESSON

2019

L'initiative française du "Grand Débat National" a été un événement politique majeur en 2019. Cet article présente les premiers résultats des méthodes de data-mining appliquées aux données traitant de la transition écologique. Différentes cartes ont été produites, par exemple la répartition des contributeurs sur le territoire français. En outre, nous proposons une analyse du discours basée sur des méthodes d'extraction de relations. Les résultats sont intégrés et visualisés à l'aide de réseaux.

Gemedoc : Un outil pour annoter les correspondances entre les documents

Workshop

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2018

In the Proceedings of EXCES workshop - EXtraction de Connaissances à partir de données Spatialisées, Spatial Analysis and GEOmatics 2017

Dans cet article, nous présentons une interface d'annotation de correspondance entre les documents d'un corpus.

In the Proceedings of Spatial Analysis and GEOmatics 2017 - SAGEO17; p. 15

Dans cet article, nous proposons une représentation de l'information spatiale contenue dans un document (i.e. données textuelles).

Autres productions scientifiques

LOGICIELS

Stance-detection browser

Python

JACQUES FIZE

2022-2023

Dans le cadre d'un contrat post-doctoral, nous avons mis en place une interface pour parcourir les résultats de détection de communautés dans notre approche. L'application est développée en Python 3.x avec le framework Dash et disponible sur Github :

https://github.com/Jacobe2169/stance_browser.

Intégration de BIOTEX en Python

Python

JACQUES FIZE

2021-2023

Biotex est un outil d'extraction automatique de terminologie (mots-clés) dans des corpus de textes. Cette approche est proposée dans : *Lossio-Ventura, J. A., Jonquet, C., Roche, M., and Teisseire, M. (2014). Biotex : A system for biomedical terminology extraction, ranking, and validation.* Le code officiel s'appuie sur un code Java et l'utilisation de l'outil TreeTagger codé en Perl. Afin de simplifier l'utilisation de cet outil, notamment pour des étudiants, je propose une librairie en Python reprenant les fonctionnalités de Biotex. Celle-ci est disponible sur GitHub :

<https://github.com/Jacobe2169/Biotex4Py>.

GMATCH4PY : a graph matching library

Python

JACQUES FIZE

2018-2019

GMATCH4PY est une librairie contenant différentes mesures de similarité de graphe proposées dans la littérature. GMATCH4PY est développée en Python 3.x et disponible sur Github : <https://github.com/Jacobe2169/GMatch4py>.

GEODICT

Python

JACQUES FIZE, GAURAV SHRIVASTAVA

2018-2019

GEODICT est un gazetier avec 12 millions d'entités spatiales référencées. Chaque entrée est associée avec un ensemble de données basiques mais précise comme les labels multilingues, le(s) polygone(s) des frontières, les coordonnées, etc. Les données renseignées dans GEODICT sont extraites à partir de jeux de données populaires : Geonames, Wikidata, OpenStreetMap.

GEODICT est accessible à cette adresse : [doi : 10.18167/DVN1/MWQQOQ](https://doi.org/10.18167/DVN1/MWQQOQ), [CiradDataverse](https://www.cirad.fr/dataverse)

GEMEDOC

Python

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2017

GEMEDOC est un outil permettant d'annoter la similarité inter- document pour un corpus. Il est accompagné d'un protocole d'annotation fondé sur la similarité entre document selon deux dimensions : la thématique et la spatialité. GEMEDOC est accessible à cette adresse :

<https://gitlab.irstea.fr/jacques.fize/gemedoc>

DATASETS

DONNÉES DU PROJET BVLAC

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2017-2019

Le projet SONGES a pour objectif l'organisation et la valorisation des ensembles de données dans leurs dimensions hétérogènes et massives. Parmi les données exploitées, nous travaillons sur un ensemble de données produites dans le cadre du projet BVLAC, un projet mené par le CIRAD qui promeut des techniques agricoles issues de l'agroécologie à Madagascar. Ces données sont accessibles à l'adresse suivante :

[doi : 10.18167/DVN1/8LIG1D](https://doi.org/10.18167/DVN1/8LIG1D), [CiradDataverse](https://www.cirad.fr/dataverse)

Relectures

Membre du Comité de Lecture (Conférence)

International Conference on Natural Language & Information Systems – NLDB 2017
Conférence en Recherche d'Informations et Applications - CORIA 2017
The Florida Artificial Intelligence Research Society - FLAIRS-31 - 2018
The 28th International Conference on Computational Linguistics - COLING 2020
SAC2020

Membre du Comité de Lecture (Journal)

International Journal of Geographical Information Science - IJGIS (2020)
International Journal of Data Science and Analytics JDSA (2021)

Présentations

16th Spatial Analysis and Geomatics Conference - SAGEO 2021

PRÉSENTATION D'ARTICLE

La Rochelle, France

2021

13th International Workshop on Spatial and Spatiotemporal Data Mining (SSTD-18) - IEEE ICDM'2018

PRÉSENTATION D'ARTICLE

Singapour

2018

International Conference on Applications of Natural Language to Information Systems (NLDB'2018)

PRÉSENTATION D'UN POSTER

Paris, France

2018

Journée "Variété des données SHS" - INFORSID - Action ADOC (Entrepôts et analyse de documents) du GDR MaDICS

PRÉSENTATION

Nantes, France

2018

Workshop "Agriculture numérique en Afrique"

PRÉSENTATION

Montpellier, France

2017

Spatial Analysis and GEOmatics 2017

PRÉSENTATION D'ARTICLE

Rouen, France

2017

Séminaire "Patrimoine Numérique Scientifique"

PRÉSENTATION

Montpellier, France

2017

Enseignements

Intégration de données dans un data ware-house

2ÈME ANNÉE IUT - UNIVERSITÉ LUMIÈRE LYON 2

Dans le cadre d'une S.A.E (Situation d'apprentissage et d'évaluation), l'objectif est de mettre à l'épreuve des étudiants face à une demande client concernant la mise en place d'une base de données.

Bron, France

2023

Programmation Avancée en Python

2ÈME ANNÉE IUT - UNIVERSITÉ LUMIÈRE LYON 2

Dans le cadre de la formation STID (Statistique et Informatique Décisionnelle), l'objectif du cours est de renforcer les connaissances des étudiants en Python. Le cours est divisé en trois parties. Dans la première partie, les étudiants apprennent à utiliser les outils de manipulation et visualisation de données Python : PANDAS, MATPLOTLIB et SEABORN. Dans la seconde partie, les étudiants appliquent les connaissances de la programmation orientée objets (acquises dans un autre cours) avec le langage Python. Dans la dernière partie, nous apprenons aux étudiants la programmation de jeux vidéo à l'aide de la librairie PYGAME¹.

Bron, France

2022-2023

Data-mining

2ÈME ANNÉE IUT - UNIVERSITÉ LUMIÈRE LYON 2

L'objectif de la matière est d'introduire les concepts de machine learning ainsi que plusieurs approches connues (clustering, classification).

Bron, France

2022

Méthodes avancées de la science de données (12 h)

MASTER 2 INFORMATIQUE - UNIVERSITÉ DE MONTPELLIER

Travaux Pratiques : Deep Learning (CNN), Word-Embedding

Projet : Développement et évaluation d'un système de classification d'occupation du sol utilisant des méthodes de machine learning.

Montpellier, France

2018

Gestion d'un projet de base de données (32 h)

3ÈME ANNÉE - EPSI MONTPELLIER

Cours + TP : Conception d'un modèle de données (MCD) répondant aux besoins du client, Utilisation d'une base de données Oracle, présentation du travail

Montpellier, France

2018

Extraction de connaissances avancées (11 h)

MASTER 2 INFORMATIQUE - UNIVERSITÉ DE MONTPELLIER

Travaux Pratiques : Deep Learning (CNN), Extraction de Motifs-séquentiels

Projet : Détection de levée de culture à partir d'images satellites. Travail sur des données réelles : données terrains fournis par un expert. Les étudiants ont présentés leurs travaux devant l'expert.

Montpellier, France

2017

Conception et exploitation d'une base de données (24 h)

3ÈME ANNÉE - EPSI MONTPELLIER

Cours + TP : Apprentissage des bases du langage SQL, Requêtes avancées avec PostgreSQL (Triggers, Procedure, etc.)

Montpellier, France

2017

Encadrement

Extraction d'entités spatiales dans des données textuelles

STAGE DE MASTER DE GAURAV SHRIVAASTAVA

Montpellier, France

2017

Tâches collectives

Doctorale de la Maison de la Télédétection

ÉVÈNEMENT DOCTORANT

Montpellier, France

2018-2019

Représentant élu des doctorants au conseil d'unité de l'UMR TETIS (Territoires, environnement, télédétection et information spatiale)

ACTIVITÉ LABORATOIRE

Montpellier, France

2018