

Jacques Fize

PHD IN COMPUTER SCIENCE

LIRIS, Bâtiment Blaise Pascal, 14 Rue de la Physique, 69100 Villeurbanne, France

☎ (+33) 4 67 55 86 27 | ✉ jacques[dot]fize[at]gmail.com | 🏠 www.jacquesfize.com | 📱 Jacobe2169

Skills

Research Data-Mining, Natural Language Processing
Text-Matching, Information Retrieval, Spatial feature extraction in text (*geoparsing*), Text Geocoding

Programming Python, Java, C/C++, JAVA, R, PHP, LaTeX
DBMS MySQL, PostgreSQL (+PostGIS), Elasticsearch
Web HTML5/CSS, Javascript, PHP, Web Application (Flask, Shiny, J2EE)
Languages French (native), English (intermediate)

Education

Master Degree in Computer Science

UNIVERSITY OF CAEN NORMANDY

- Image Processing and Natural Language Processing

Caen, France

2015 - 2016

Bachelor Degree in Computer Science

UNIVERSITY OF CAEN NORMANDY

Caen, France

2010 - 2014

Literary Baccaureat

JEAN-FRANCOIS MILLET HIGH SCHOOL

- Music Speciality

Cherbourg-en-Cotentin, France

2010

Experience

Institut National des Sciences Appliquées (INSA)

POST-DOCTORAL RESEARCHER

Research conducted to improve the link prediction task in graphs.

Lyon, France

Dec. 2020 - Mai 2020

Institut National des Sciences Appliquées (INSA)

POST-DOCTORAL RESEARCHER

Development of Geoparsing methods (identification of places in texts) based on Deep Learning.

Lyon, France

Dec. 2019 - Dec. 2020

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)

PH.D STUDENT

Thesis: Matching massive and heterogeneous textual data

The objective of this thesis is to propose new methods of data matching with a focus on their heterogeneous and massive nature. The proposed matching methods are based on original representations of the documents and different measures of similarity. In our work, we propose a multidimensional mapping approach that exploits the thematic and spatial dimension of heterogeneous textual data.

Montpellier, France

Oct. 2016 - Oct. 2019

Orange Labs

INTERNSHIP

The objective of this internship was to design a knowledge base generated from different sources from Linked Open Data (Wikidata, DBpedia, Geonames). Each stored entity is associated with a class, multilingual labels and other attributes specific to its category (Person, Organization, Location). Finally, to query the knowledge base, we have developed a specific query language (similar to SQL).

Lannion, France

March 2016 - August 2016

IT DEPARTMENT OF THE GENERAL COUNCIL OF CALVADOS

INTERNSHIP

The objective of this internship was to develop a web application to transfer sensitive information between the public services of the Calvados department.

Caen, France

March 2014 - May 2014

JOURNAL

Matching heterogeneous textual data using spatial features

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

Intelligent Data Analysis Journal (To Appear)

Journal

2019

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

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

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.

Journal

2017

INTERNATIONAL CONFERENCE AND WORKSHOPS PROCEEDINGS

Mapping Heterogeneous Textual Data: a Multidimensional Approach based on Spatiality and Theme

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

In the Proceedings of 6th International Conference on INTERNET SCIENCE (INSCI2019), Perpignan, France, December, 2-5, 2019 (To Appear)

Conference

2019

Matching heterogeneous textual data using spatial features

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

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

This paper presents a process for matching heterogeneous textual data according to the spatial dimension.

Workshop

2018

Gemedoc : A Text Similarity Annotation Platform

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

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

In this article, we present an interface for annotating correspondence between documents in a corpus.

Demo Paper

2018

GeoDict : an integrated gazetteer

JACQUES FIZE, GAURAV SHRIVASTAVA

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

In this article, we present a new *gazetteer* (i.e. geographic database), Geodict, which compiles data from three sources: Wikidata, Geonames, OpenStreetMap.

Workshop

2018

NATIONAL CONFERENCE AND WORKSHOP PROCEEDINGS

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

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

The Great National Debate, decided by Emmanuel Macron at the beginning of 2019 to respond to the Yellow Vests social movement, allowed the collection of citizens' contributions on the ecological transition via an online platform. In this article, we use the corpus constituted by these contributions to identify areas where participants are asking for the development of bicycle paths and railway facilities. For this purpose, we have created a classification model to identify contributions dealing with the theme of transportation and proposed a method for extracting patterns that reflect the contributors' proposals.

Conference

2019

Qui a peur du changement climatique ?

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

In the Proceedings of Spatial Analysis and GEOmatics 2019, (SAGEO 2019)

The French "Grand Débat National" (Great National Debate) initiative was an important political event in 2019. This article presents early results of data-mining methods applied to the data dealing with the ecological transition. Different maps were produced, e.g. the distribution of contributors on the French territory. In addition, we propose a discourse analysis based on relation extraction methods. Results are integrated and visualized using networks.

Conference

2019

Gemedoc : Un outil pour annoter les correspondances entre les documents

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

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

In this article, we present an interface for annotating correspondences between documents in a corpus.

Workshop

2018

Spatial Textual Representation (STR) ou comment représenter la spatialité des données textuelles

Conference

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2017

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

In this article, we propose a representation for the spatial information contained in a document (i.e. textual data).

Other scientific productions

SOFTWARE AND LIBRARY

GMATCH4PY : a graph matching library

Python

JACQUES FIZE

2018-2019

GMATCH4PY is a Python library that implements different graph matching algorithms proposed in the literature. GMATCH4PY is developed in Python 3.x and available on Github at this address: <https://github.com/Jacobe2169/GMatch4py>.

GEODICT

Python

JACQUES FIZE, GAURAV SHRIVASTAVA

2018-2019

GEODICT is a gazetteer with more than 12 million referenced space entities. Each entry is associated with basic but precise data such as multilingual labels, border polygon(s), coordinates, etc. The data in GEODICT is extracted from three popular data sets: Geonames, Wikidata, OpenStreetMap.

GEODICT is available at this address: [doi:10.18167/DVN1/MWQQOQ](https://doi.org/10.18167/DVN1/MWQQOQ), [CiradDataverse](https://www.cirad.fr/dataset/cirad-dataverse)

GEMEDOC

Python

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2017

GEMEDOC is a tool to annotate the inter-document similarity for a corpus. It is accompanied by an annotation protocol based on the similarity between documents according to two dimensions: the theme and spatiality. GEMEDOC is available at this address:

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

DATASETS

DATA FROM THE BVLAC PROJECT

JACQUES FIZE, MATHIEU ROCHE, MAGUELONNE TEISSEIRE

2017-2019

The SONGES project aims to organize and valorize datasets in their heterogeneous and massive dimensions. Among the data used, we are working on a dataset produced as part of the BVLAC project, a project led by CIRAD that promotes agricultural techniques derived from agroecology in Madagascar. These data are available at the following address:

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

Reviews

Member of the reviewing committees (Conference)

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

Member of the reviewing committees (Journal)

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

Presentations

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

Singapour

PAPER PRESENTATION

2018

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

Paris, France

POSTER

2018

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

Nantes, France

PRESENTATION

2018

Workshop "Agriculture numérique en Afrique"

PRESENTATION

Montpellier, France

2017

Spatial Analysis and GEOmatics 2017

PAPER PRESENTATION

Rouen, France

2017

Seminar "Patrimoine Numérique Scientifique"

PRESENTATION

Montpellier, France

2017

Teachings

Advanced methods in Data Science (12 h)

2ND YEAR OF MASTER DEGREE IN COMPUTER SCIENCE - MONTPELLIER UNIVERSITY

Practical work: Deep Learning (CNN), Word-Embedding

Project: Development and evaluation of a land use classification system using machine learning methods.

Montpellier, France

2018

Database project management (32 h)

3RD YEAR - EPSI MONTPELLIER

Course + TP: Design of a data model (MCD) that meets a client's needs. Teaching with Oracle database.

Montpellier, France

2018

Extraction of advanced knowledge (11 h)

2ND YEAR OF MASTER DEGREE IN COMPUTER SCIENCE - MONTPELLIER UNIVERSITY

Practical Work: Deep Learning (CNN), Sequential Pattern Extraction

Project: Detection of crop emergence from satellite images. Project on real data: field data provided by an expert. The students presented their work to the expert.

Montpellier, France

2017

Design and exploitation of a database (24 h)

3RD YEAR - EPSI MONTPELLIER

Course + TP: Learning the basics of SQL language and advanced queries with PostgreSQL (Triggers, Procedure, etc.)

Montpellier, France

2017

Supervision

Extraction of spatial entities from textual data

MASTER'S INTERNSHIP OF GAURAV SHRIVAASTAVA

Montpellier, France

2017

Collective

"Doctoriales" - Maison de la Télédétection

PH.D EVENT

Montpellier, France

2018-2019

PhD students representative in the Mixed Research Unit TETIS council

LABORATORY ACTIVITY

Montpellier, France

2018