

**Djellel Difallah**

CONTACT INFORMATION	New York University Center for Data Science 60 5th Avenue, 7th Floor, Room 602 New York, NY 10011 USA	<i>Phone</i> +1-917-696-4069 <i>E-mail:</i> djellel@nyu.edu <i>WWW:</i> <a href="https://dedcode.github.io">https://dedcode.github.io</a>
RESEARCH INTERESTS	Data Science, Human Computation, Collaborative Knowledge Bases, Data Management, Knowledge Extraction	
EDUCATION		
	<b>University of Fribourg</b> , Fribourg, Switzerland	2011 to 2015
	Ph.D., Computer Science (Best CS Thesis Award) – June 2015	
	<ul style="list-style-type: none"><li>• Dissertation title: “Quality of Service in Crowd-Powered Systems”</li><li>• Adviser: Professor Philippe Cudré-Mauroux</li></ul>	
	<b>University of Louisiana</b> , Lafayette, USA	2009 to 2011
	M.S., Computer Science (with Honors) – May 2011	
	<ul style="list-style-type: none"><li>• Adviser: Professor Vijay V. Raghavan</li><li>• Topic: Action Rules Mining</li><li>• Project: Frequent Association Action Rules Mining Using FP-Trees</li></ul>	
	<b>University of Sciences and Technology H. Boumediene</b> , Algiers, Algeria	1999 to 2004
	Dipl.-Ing., Informatics – December 2004	
	<ul style="list-style-type: none"><li>• Track: Information Systems</li><li>• Project: Modeling a Medical Center Information System with UML</li></ul>	
PROFESSIONAL EMPLOYMENT		
	<b>Moore-Sloan Data Science Fellow</b> , New York University NYU Center for Data Science	June 2017 to present
	<ul style="list-style-type: none"><li>• Research in crowd population estimation and dynamics; crowd behavior analysis in Wikidata community; Wikidata-based data integration; Wikidata graph evolution prediction; Crowdsourcing sport video annotation</li></ul>	
	<b>Postdoctoral Researcher</b> , University of Fribourg (Switzerland)	September 2015 to May 2017
	eXascale Infolab	
	<ul style="list-style-type: none"><li>• Research in query expansion (tailored to job search), entity linking and co-reference resolution</li><li>• Supervision of PhD and Master students</li><li>• Supervisor: Professor Philippe Cudré-Mauroux</li></ul>	
	<b>Research Assistant</b> , University of Fribourg (Switzerland)	May 2011 to June 2015
	eXascale Infolab	
	<ul style="list-style-type: none"><li>• Research in crowd-powered systems: Human intelligence task scheduling, task routing, and worker retention.</li><li>• Research in crowdsourcing-based entity linking and instance matching</li><li>• Supervision of Bachelor and Master students</li><li>• Supervisor: Professor Philippe Cudré-Mauroux</li></ul>	
	<b>Intern</b> , Microsoft Research CISL, Mountain View, CA	May to August 2013
	Cloud and Information Services Lab	
	<ul style="list-style-type: none"><li>• Project: “Reservation-based scheduling with Hadoop YARN”</li><li>• Supervisor: Dr. Carlo Curino</li></ul>	
	<b>Student Developer</b> , Google Summer of Code Program	May to August 2010
	Drizzle DBMS	
	<ul style="list-style-type: none"><li>• Project: “A query cache plugin for Drizzle DBMS based on Memcached”</li><li>• Supervisor: Toru Maesaka</li></ul>	

TEACHING EXPERIENCE	<b>Information Management Engineer</b> , Schlumberger, Algiers, Agleria Schlumberger Information Solutions	2006 to 2009
	<ul style="list-style-type: none"> <li>• Oracle Database Administration, SQL reports and query tuning</li> <li>• Supervisors: Toufik Bessadat, Mounir Ferroudj</li> </ul>	
	<b>Software Developer</b> , EEPAD Internet Services Provider Authentication and Billing	2005 to 2006
	<ul style="list-style-type: none"> <li>• Development and integration of authentication and provisioning tools</li> <li>• Supervisor: Yacine Rezgui</li> </ul>	
	<b>University of Fribourg</b> , Fribourg, Switzerland	
	<i>Co-Instructor</i>	Fall 2016
	<ul style="list-style-type: none"> <li>• Big Data Infrastructures (Graduate) <ul style="list-style-type: none"> <li>• Co-Instructor: Prof. Philippe Cudré-Mauroux</li> <li>• Topics: Introduction to DBMS/SQL, Distributed Data Management, NOSQL, Hadoop, Spark</li> </ul> </li> </ul>	
	<i>Co-Instructor</i>	Fall 2015
	<ul style="list-style-type: none"> <li>• Data Science Seminar (Graduate) <ul style="list-style-type: none"> <li>• Co-Instructor: Dr. Mourad Khayati</li> <li>• Topics: clustering, compression and similarity techniques used for time series data and graphs</li> </ul> </li> </ul>	
	<i>Teaching Assistant</i>	Fall semesters from 2011 to 2014
	<ul style="list-style-type: none"> <li>• Social Computing (Graduate) <ul style="list-style-type: none"> <li>• Graduate-level course in animal behavior</li> <li>• Main instructors: Dr. Gianluca Demartini and Prof. Philippe Cudré-Mauroux</li> <li>• Topics: Semantic Web, Crowdsourcing, and Graphs'</li> </ul> </li> </ul>	
GRANTS AND AWARDS	<b>Lucerne University of Applied Sciences</b> , Lucerne, Switzerland	
	<i>Guest Speaker</i>	Summer 2016
	<ul style="list-style-type: none"> <li>• Introduction to Big Data</li> </ul>	
	<b>International Institute of Management in Technology</b> , Fribourg, Switzerland	
	<i>Guest Speaker</i>	Summer 2016
	<ul style="list-style-type: none"> <li>• Information Management and Decision Support</li> </ul>	
	<i>Senior Staff – European Project H2020</i>	2017
	<ul style="list-style-type: none"> <li>• Project: “FashionBrain: Understanding Europe’s Fashion Data Universe”</li> <li>• Project track: “Big Data PPP: Cross-sectorial and Cross-lingual Data Integration and Experimentation”</li> <li>• Grant value: 2.9M EUR total; University of Fribourg’s share: 690’000 EUR</li> </ul>	
	<i>Senior Staff – Swiss Commission of Technology and Innovation</i>	2016
	<ul style="list-style-type: none"> <li>• Project: “Query Expansion: Deep Learning and Crowdsourcing”</li> <li>• Grant value: 210’000 CHF</li> </ul>	
	<i>Best Computer Science thesis award – University of Fribourg</i>	February 2016
	<ul style="list-style-type: none"> <li>• Project: “Measuring Human Behaviour in Search Evaluation Micro Tasks”</li> <li>• Grant value: 3’000 CHF</li> </ul>	
	<i>Research Visit Grant – European Science Foundation</i>	Spring 2015
	<ul style="list-style-type: none"> <li>• Project: “Measuring Human Behaviour in Search Evaluation Micro Tasks”</li> <li>• Host: Dr. Gianluca Demartini, University of Sheffield, UK</li> <li>• Grant value: 1’350 EUR</li> </ul>	

	<i>Travel Grant – HCOMP Doctoral Consortium</i> <ul style="list-style-type: none"> <li>• Project: “Scalable Human-based Computation”</li> <li>• Grant value: 700 USD</li> </ul> <i>University of Louisiana at Lafayette Honors, for maintaining a GPA of 4.0</i> <i>Fulbright Exchange Student Grant</i>	October 2014   2011 2009-2011
INVITED TALKS	<ul style="list-style-type: none"> <li>• Djellel Difallah. “The Dynamics of Crowdsourcing Platforms: The Case of Amazon Mechanical Turk and Wikidata” (Seminar) NYU Center for Data Science. NY, USA</li> <li>• Ujwal Gadiraju, Gianluca Demartini, Djellel Difallah and Michele Catasta. “Using Crowdsourcing Effectively for Social Media Research” (Tutorial) ICWSM’16 Cologne, Germany</li> <li>• Ujwal Gadiraju, Gianluca Demartini, Djellel Difallah and Michele Catasta. “Using Crowdsourcing Effectively for Social Media Research” (Tutorial) WebSci’16 Hannover, Germany</li> <li>• Djellel Difallah. “Quality and Performance Optimizations in Microtask Crowdsourcing” MIT CSAIL, Boston, MA</li> <li>• Djellel Difallah. “Quality and Performance Optimizations in Microtask Crowdsourcing” Parallel Data Lab, CMU, Pittsburgh, PA</li> <li>• Djellel Difallah. “Extending The OLTP-Bench Framework for Big Data Systems” (Workshop) Workshop on Big Data Benchmarking, Postdam, Germany</li> <li>• Djellel Difallah. “Resource Management and Planning for Hadoop” Microsoft Research. Mountain View, CA</li> <li>• Djellel Difallah. “Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms” (Workshop) CrowdSearch 2012 workshop at WWW, Lyon, France</li> </ul>	September 2017  May 2016  May 2016  November 2014  October 2014  August 2014  August 2013  April 2012
PROFESSIONAL SERVICE	<b>Program Committee Member</b> <ul style="list-style-type: none"> <li>• 28<sup>th</sup> International World Wide Web Conference (WWW 2019)</li> <li>• 27<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2018)</li> <li>• IEEE International Conference on Data Mining (ICDM 2018)</li> <li>• 24<sup>th</sup> ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2018)</li> <li>• 12<sup>th</sup> International AAAI Conference On WEB and Social Media (ICWSM 2018)</li> <li>• 27<sup>th</sup> International World Wide Web Conference (WWW 2018)</li> <li>• 23<sup>rd</sup> ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2017)</li> <li>• 26<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2017)</li> <li>• 26<sup>th</sup> International World Wide Web Conference (WWW 2017)</li> <li>• 16<sup>th</sup> International Conference on Web Engineering (ICWE 2016)</li> </ul> <b>External Reviewer</b> <ul style="list-style-type: none"> <li>• IEEE International Conference on Data Engineering – ICDE (2016)</li> <li>• ACM Special Interest Group on Management of Data – SIGMOD (2016)</li> <li>• ACM Conference on Web Science – WebSci (2016)</li> <li>• ACM Symposium on Cloud Computing – SoCC (2016)</li> <li>• Very Large Data Bases– VLDB (2014, 2016)</li> <li>• International Conference on Extending Database Technology – EDBT (2013)</li> <li>• Association for the Advancement of Artificial Intelligence – AAAI (2014, 2015)</li> <li>• IEEE BigData (2013)</li> </ul>	
OPEN SOURCE SERVICE	<i>OLTP-Bench</i> (Lead Contributor 2012-2015) <ul style="list-style-type: none"> <li>• Database Benchmarking Framework</li> <li>• <a href="https://github.com/oltpbenchmark">https://github.com/oltpbenchmark</a></li> </ul>	

*OpenTurk* (Lead Contributor 2013-2014)

- Chrome extension for Mturk HIT notification
- <https://github.com/openturk>

*Apache Hadoop YARN* (Contributor 2013)

- Research and development on the scheduling system for Hadoop 2.0
- Code Repository: <https://issues.apache.org/jira/browse/YARN-1051>
- Various bug fixes in the scheduler

*SciDB* (Contributor 2011-2013)

- Array Database Management System
- Implementation of a science benchmark (SSDB)

*Drizzle* (Contributor 2010)

- Database Management System with a micro-kernel architecture
- Code Repository: <https://code.launchpad.net/~dedzone/drizzle/query-cache-soc>

*Apache Mahout* (Contributor 2010)

- Scalable machine-learning algorithms
- Bug fixes in K-Means implementation

#### ADVISING AND MENTORING

##### University Students

- **Alexander Striffeler**, Master Student, Computer Science 2015  
Thesis topic: “Benchmarking Big Data Infrastructures”  
Co-advised by: Philippe Cudré-Mauroux and Mourad Khayati
- **Marwa Bouzeyane**, Master Student, Computer Science 2015  
Thesis topic: “Grammatical Errors Detection Based On N-gram Analysis”  
Co-advised by: Philippe Cudré-Mauroux and Mourad Khayati
- **Stefan Nüesch**, Master Student, Computer Science 2014  
Thesis topic: “Real-Time Anomaly Detection in Water Distribution Networks using Spark Streaming”  
Co-advised by: Philippe Cudré-Mauroux
- **Phokham Nonava**, Master Student, Computer Science 2014  
Thesis topic: “HDFS Blocks Placement Strategy”  
Co-advised by: Philippe Cudré-Mauroux, Benoit Perroud and Martin Grund
- **Simpal Kumar**, Master Student, Computer Science 2014  
Thesis topic: “Real Time Data Analysis for Water Distribution Network using Storm”  
Co-advised by: Philippe Cudré-Mauroux
- **Dani Rotzetter**, Master Student, Computer Science 2014  
Thesis topic: “Crowd-Flow Designer: An Open-Source Toolkit to Design and Run Complex Crowd-Sourced Tasks”  
Co-advised by: Philippe Cudré-Mauroux and Gianluca Demartini
- **Victor Felder**, Bachelor Student, Computer Science 2013  
Thesis topic: “Openturk: An Implementation of the Pick-A-Crowd Architecture”

##### High School Students

- **Job-Info Day Fribourg**. Introduction to the computer science career track for high school students (2013-2017).
- **Cyber Camp**. Instructor at the computer science summer camp for high school students (2012-2014).