Djellel Difallah

Address: 888 Main Street Apt. 112
New York, NY 10044 – USA
Email: djellel@nyu.edu, Phone: +1 (917) 696-4069
Webpage: https://dedcode.github.io/

Research Interests

My main research interests revolve around human computation and its application to data management, data processing, machine learning, natural language processing and knowledge extraction. Specifically, I investigate novel human machine learning techniques, Bayesian models for truth inference, task scheduling optimization problems and analysis of crowdsourcing marketplaces.

Education

2011–2015	PhD in Computer Science, University of Fribourg – Switzerland. – Dissertation title: "Quality of Service in Crowd-Powered Systems". – Received the best thesis award in Computer Science (University of Fribourg 2015).
2009–2011	MSc in Computer Science, University of Louisiana at Lafayette – USA. – Fulbright Scholar. – Received honors for maintaining a GPA of 4.0 four semesters.
1999–2004	Dipl.-Ing. in Informatics, University of Science and Technology of Algiers – Algeria.

Employment history

Since 2017	**Moore-Sloan Data Science Fellow (New York University).** – Current research focuses on studying crowdsourcing platforms, their dynamics and evolution both from a worker and economic perspective and based on public activity logs. – Pursue a research plan on combining machine learning and crowdsourcing in the context of data science.
2015–05/2017	Senior Research Scientist at eXascale Infolab (University of Fribourg). – Postdoctoral research on active learning with the crowd, and truth inference from crowdsourced data; applied to data analysis and information retrieval projects. – Contribute to research projects on entity linking, entity resolution, and query expansion. – Lecturer for the course "Big Data Infrastructures" and the seminar "Data Science".
2011–2015	Research Assistant at eXascale Infolab (University of Fribourg). – Contributed to various projects in the group e.g., anomaly detection (stream processing), array data management (SciDB), human memory-based information systems (TransactiveDB), entity linking and coreference resolution (semantic Web and NLP). – Teaching assistant for the social computing class. – Supervise master students working on stream data processing solutions for anomaly detection in Water Distribution Networks.
2013	Internship at Cloud and Information Services Lab (Microsoft Research). – Project: "Reservation based scheduling with Hadoop YARN". – Repo: https://issues.apache.org/jira/browse/YARN-1051
2010	Student Developer at Google Summer of Code Program. – Project: "A query cache plugin for Drizzle DBMS based on Memcached".

	– Repo: `https://code.launchpad.net/~dedzone/drizzle/query-cache-soc`
2006–2009	Information Management Engineer at Schlumberger. – Database administration (Oracle). – Dashboard and KPI development. – SQL reports and query optimization.
2005–2006	Engineer at EEPAD Internet Services Provider. – System administration of the authentication system platform (RADIUS) for both line and wireless clients. – Developed an automatic provisioning solution to synchronize the information system and the deployed ADSL infrastructure using SNMP.

Teaching Experience

Lecturer	*(Swiss Joint Master of Science in Computer Science `http://mcs.unibnf.ch/`)* Big Data Infrastructures (Graduate – Fall 2016). – Main Topics: NOSQL systems (Dynamo, Cassandra, MongoDB), Hadoop Map/Reduce, YARN, Spark and Kafka. – Co-Instructor: Philippe Cudré-Mauroux.
Lecturer	Data Science Seminar (Graduate – Fall 2015). – Topics: Selected papers focusing on time series analysis and crowdsourcing. – Co-Instructor: Mourad Khayati.
TA	Social Computing (Graduate – Fall semester from 2011 to 2014). – Instructors: Philippe Cudré-Mauroux and Gianluca Demartini.

Professional Activities

PC member	– WWW(2017), CIKM(2017), KDD(2017), CHI(2017), ICWE(2016), ESWC(2015,2016).
Reviewer	– I routinely review journal publications for TKDE, TSC, VLDBJ, SWJ as well as conference papers for WWW, VLDB, SIGMOD, ICDE, SoCC, ISWC.

Grants and Other Funding

2016	EU H2020 ICT-14. – Project track: "Big Data PPP: Cross-sectorial and Cross-lingual Data Integration and Experimentation". – Project title: "FashionBrain: Understanding Europe's Fashion Data Universe". – Three academic institutions and three industry partners. – Grant value: 2.9M EUR total cost; University of Fribourg share: 690'000 EUR.
2016	Swiss Commission of Technology and Innovation. – Project title: "Query Expansion: Deep Learning and Crowdsourcing". – A technology transfer project with one industry partner (job aggregator). – Grant value: 210'000 CHF.
2015	European Science Fundation (Research Visit Grant). – Host: Dr. Gianluca Demartini University of Sheffield. – Theme: "Measuring Human Behaviour in Search Evaluation Micro Tasks". – Grant value: 1'350 EUR.
2014	HCOMP Doctoral Consortium (Travel Grant). – Theme: "Scalable Human-based Computation". – Travel grant value: 700 USD.

Talks and Tutorials

Introduction to Big Data (CAS Business Intelligence and Analytics).
– At Lucerne University of Applied Sciences. Lucerne, Switzerland – August 2016.

Using Crowdsourcing Effectively for Social Media Research (tutorial).
– At ICWSM. Cologne, Germany – May 2016.
– At WebSci. Hannover, Germany – May 2016.
– Together With Ujwal Gadiraju, Gianluca Demartini, and Michele Catasta.

Quality and Performance Optimizations in Microtask Crowdsourcing.
– At MIT CSAIL. Boston, MA – November 2014.
– At CMU. Pittsburgh, PA – October 2014.

Extending The OLTP-Bench Framework for Big Data Systems.
– At the Workshop on Big Data Benchmarking. Postdam, Germany – August 2014.

Resource Management and Planning for Hadoop.
– At Microsoft Cloud and Information Services Lab. Mountain View, CA – August 2013.

Opensource Projects

Lead
– OLTP-Bench: `https://github.com/oltpbenchmark`
– Openturk: `https://github.com/openturk`

Contributor
– Apache Hadoop YARN (Bug fixes and reservation based scheduling PoC).
– SciDB (I implemented the *SSDB* astronomy benchmark).
– Drizzle (I implemented the query cache plugin).
– Apache Mahout (Bug fixes).

Spoken Languages

Fluent in French, English, and Arabic.

List of Publications

1. Prokofyev, R., Luggen, M., **Difallah, D. E.** & Cudré-Mauroux, P. *SwissLink: High-Precision, Context-Free Entity Linking Exploiting Unambiguous Labels* in *Proceedings of the 13th International Conference on Semantic Systems* (2017).

2. **Difallah, D. E.**, Demartini, G. & Cudré-Mauroux, P. *Scheduling Human Intelligence Tasks in Multi-Tenant Crowd-Powered Systems* in *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016* (2016), 855–865.

3. Tonon, A., Felder, V., **Difallah, D. E.** & Cudré-Mauroux, P. *VoldemortKG: Mapping schema.org and Web Entities to Linked Open Data* in *The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part II* (2016), 220–228.

4. Aken, D. V., **Difallah, D. E.**, Pavlo, A., Curino, C. & Cudré-Mauroux, P. *BenchPress: Dynamic Workload Control in the OLTP-Bench Testbed* in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015* (2015), 1069–1073.

5. Prokofyev, R., Tonon, A., Luggen, M., Vouilloz, L., **Difallah, D. E.** & Cudré-Mauroux, P. *SANAPHOR: Ontology-Based Coreference Resolution* in *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part I* (2015), 458–473.

6. **Difallah, D. E.**, Catasta, M., Demartini, G., Ipeirotis, P. G. & Cudré-Mauroux, P. *The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk* in *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22, 2015* (2015), 238–247.

7. Catasta, M., Tonon, A., **Difallah, D. E.**, Demartini, G., Aberer, K. & Cudré-Mauroux, P. *Hippocampus: answering memory queries using transactive search* in *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume* (2014), 535–540.

8. Catasta, M., Tonon, A., **Difallah, D. E.**, Demartini, G., Aberer, K. & Cudré-Mauroux, P. TransactiveDB: Tapping into Collective Human Memories. *PVLDB* **7,** 1977–1980 (2014).

9. Curino, C., **Difallah, D. E.**, Douglas, C., Krishnan, S., Ramakrishnan, R. & Rao, S. *Reservation-based Scheduling: If You're Late Don't Blame Us!* in *Proceedings of the ACM Symposium on Cloud Computing, Seattle, WA, USA, November 03 - 05, 2014* (2014), 2:1–2:14.

10. **Difallah, D. E.**, Catasta, M., Demartini, G. & Cudré-Mauroux, P. *Scaling-Up the Crowd: Micro-Task Pricing Schemes for Worker Retention and Latency Improvement* in *Proceedings of the Seconf AAAI Conference on Human Computation and Crowdsourcing, HCOMP 2014, November 2-4, 2014, Pittsburgh, Pennsylvania, USA* (2014).

11. Demartini, G., **Difallah, D. E.** & Cudré-Mauroux, P. Large-scale linked data integration using probabilistic reasoning and crowdsourcing. *VLDB J.* **22,** 665–687 (2013).

12. **Difallah, D. E.**, Pavlo, A., Curino, C. & Cudré-Mauroux, P. OLTP-Bench: An Extensible Testbed for Benchmarking Relational Databases. *PVLDB* **7,** 277–288 (2013).

13. **Difallah, D. E.**, Demartini, G. & Cudré-Mauroux, P. *Pick-a-crowd: tell me what you like, and i'll tell you what to do* in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013* (2013), 367–374.

14. **Difallah, D. E.**, Cudré-Mauroux, P. & McKenna, S. A. Scalable Anomaly Detection for Smart City Infrastructure Networks. *IEEE Internet Computing* **17,** 39–47 (2013).

15. Curino, C., **Difallah, D. E.**, Pavlo, A. & Cudré-Mauroux, P. *Benchmarking OLTP/web databases in the cloud: the OLTP-bench framework* in *Proceedings of the Fourth International Workshop on Cloud Data Management, CloudDB 2012, Maui, HI, USA, October 29, 2012* (2012), 17–20.

16. Demartini, G., **Difallah, D. E.** & Cudré-Mauroux, P. *ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking* in *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012* (2012), 469–478.

17. **Difallah, D. E.**, Demartini, G. & Cudré-Mauroux, P. *Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms* in *Proceedings of the First International Workshop on Crowdsourcing Web Search, Lyon, France, April 17, 2012* (2012), 26–30.

18. **Difallah, D. E.**, Benton, R. G., Raghavan, V. V. & Johnsten, T. *FAARM: Frequent Association Action Rules Mining Using FP-Tree* in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, Vancouver, BC, Canada, December 11, 2011* (2011), 398–404.