Anca Dumitrache

http://ancad.ro | anca.dmtrch@gmail.com | +31-614-622-712

INTERESTS

Data Science, Human – Computer Interaction, Natural Language Processing, Machine Learning, Crowdsourcing, Linked Open Data.

PROFESSIONAL EXPERIENCE

Center for Advanced Studies (CAS), IBM, Amsterdam, Netherlands **02/2013 to 11/2018**

Research scientist

Implemented the CrowdTruth project (collecting gold standard data for the training, evaluation of machine learning systems) in the context of IBM Watson powered business solutions. Trained and evaluated question answering models from the IBM Bluemix stack for applications in the medical field, tourism and cultural heritage. Part-time position. Resulting publications:

- Dumitrache et al.: *Empirical Methodology for Crowdsourcing Ground Truth*. Semantic Web Journal (in publication).
- Dumitrache et al.: *Dr. Detective: combining gamification techniques and crowdsourcing to create a gold standard in medical text*. CrowdSem Workshop at ISWC 2013.

Google AI, New York, USA   **06/2016 to 09/2016**

Software engineering intern

Developed a model for relation classification from sentences in the open domain, using a convolutional neural network with word embeddings as features. Implemented in Python with Tensorflow. Resulting publications:

- Dumitrache, Aroyo, Welty: *Crowdsourcing Semantic Label Propagation in Relation Classification*. FeVeR Workshop at EMNLP 2018.
- Dumitrache, Aroyo, Welty: *False Positive and Cross-relation Signals in Distant Supervision*. AKBC Workshop at NIPS 2017.

Watson Research Group, IBM, New York, USA   **01/2014 to 06/2014**

Research intern

Trained and evaluated a model for relation extraction from sentences in the medical domain, showing that models trained with crowdsourcing annotations perform as well as those trained with expert annotations. Implemented in Java. Resulting publications:

- Dumitrache, Aroyo, Welty: *Crowdsourcing Ground Truth for Medical Relation Extraction*. ACM TiiS, 8(2), 12.
- Dumitrache, Aroyo, Welty: *CrowdTruth Measures for Language Ambiguity*. LD4IE Workshop at ISWC 2015. Best paper award.

Network Institute, Vrije Universiteit Amsterdam, Netherlands   **10/2012 to 08/2013**

Academy assistant

Analyzed online networks of scholarly publications for relevant indicators of researcher talent. Implemented in R. Resulting publication:

- Dumitrache, Groth, Besselaar: *Identifying research talent using web-centric databases*. ACM Web Science 2013.

Drupal   **05/2012 to 08/2012**

Google Summer of Code intern

Developed and tested patches for Schema.org microdata support for Drupal user-contributed modules (contrib field type). Implementation done in PHP.

Reasoning and Querying Unit, Digital Enterprise Research Institute (DERI), Galway, Ireland   **06/2010 to 09/2010**

Research intern

Developed an XML parser mapping XMPP stanzas to RDF, for the purpose of creating a server component for Jabber. Implemented in C/C++. Resulting publication:

- Dumitrache et al.: *Enabling privacy-preserving semantic presence in instant messaging systems*. Context 2011.

EDUCATION	**Vrije Universiteit Amsterdam**, Netherlands	**11/2013 to 11/2018**

Vrije Universiteit Amsterdam, Netherlands   **11/2013 to 11/2018**

PhD, User-Centric Data Science research group

Thesis: Modeling Ambiguity in Natural Language Processing with Crowdsourcing

Vrije Universiteit Amsterdam, Netherlands   **09/2011 to 08/2013**

MSc, Artificial Intelligence (cum laude)

Thesis: Combining Gamification Techniques and Crowdsourcing to Create a Gold Standard for Medical Text

Jacobs University Bremen, Germany   **09/2008 to 06/2011**

BSc, Computer Science

TECHNICAL SKILLS

Programming languages:
- Good: Python (Tensorflow, Gensim, NLTK, SciPy), R (tidyverse)
- Moderate: Java, PHP, C/C++

Software: Jupyter Notebooks, RStudio, Git, LaTeX, Gephi.

Databases: SQL, Mongo DB, HBase.

Distributed computing: Distributed TensorFlow, Spark, Map Reduce.

Operating systems: Unix/Linux.

WORKSHOPS & INVITED TALKS

Will organize tutorial on the *CrowdTruth Methodology for Crowdsourcing Ground Truth* at the International Semantic Web Conference (ISWC). Monterey, CA, USA. October, 2018.

Organized *Subjectivity, Ambiguity and Disagreement (SAD) Workshop* at the Human Computation (HCOMP) conference. Zürich, Switzerland. July 2018.

Tutorial on *CrowdTruth – Crowdsourcing Ground Truth with Disagreement Analysis* at the *Human Computation NL* symposium. Amsterdam, Netherlands. September, 2017.

Keynote talk on *Harnessing the Diversity in Human Annotation with CrowdTruth* at the ESSENCE Network *Conference on Computational Approaches to Diversity in Interaction and Meaning*. Venice, Italy. October 2017.

Talk on *Watson & Natural Language Processing* at the *Watson Experience MeetUp*. Utrecht, Netherlands. March 2016.

AWARDS & HONORS

IBM PhD Fellowship awards program, IBM, Netherlands, 2013 – 2016.

Best Poster in the Human and the Machine track, *2nd Prize* for best poster of the conference, ICT Open, Netherlands, 2016.

VU Fellowship Programme (VUFP) scholarship for academic excellence, Vrije Universiteit Amsterdam, Netherlands, 2011 – 2013.

President's List distinction for academic excellence, Jacobs University Bremen, Germany, 2010 – 2011 and 2008 – 2009.

SELECTED COURSES

Transylvanian Machine Learning Summer School (TMLSS) on machine & reinforcement learning. Cluj Napoca, Romania. July 2018.

Deep Learn summer school on deep learning. Bilbao, Spain. July 2017.

CBS Data Camp & Advanced Course on Managing Big Data and working with Spark. Enschede, Netherlands. December 2016.

Lisbon Machine Learning Summer School (LxMLS) on natural language processing. Lisbon, Portugal. July 2015.

Data Science in Society summer school. Southampton, UK. July 2014.

LANGUAGES

Romanian: native

English: proficient (C2)

French: intermediate (B2)

Dutch: basic (A2)