

# CURRICULUM VITAE

## ISABELLE AUGENSTEIN

### PERSONAL DETAILS

---

G.6 5th Floor,  
One Euston Square, 40 Melton Street,  
London, NW1 2FD, UK  
Email: [i.augenstein@ucl.ac.uk](mailto:i.augenstein@ucl.ac.uk)  
WWW: <http://isabelleaugenstein.github.io/>  
Twitter: @IAugenstein  
Github: <https://github.com/isabelleaugenstein>

LinkedIn: <http://uk.linkedin.com/pub/isabelle-augenstein/7a/36b/824/>  
Google Scholar: <http://scholar.google.de/citations?user=DjJp0dcAAAAJ>  
Slideshare: <http://www.slideshare.net/isabelleaugenstein/>  
Videolectures: [http://videolectures.net/isabelle\\_augenstein/](http://videolectures.net/isabelle_augenstein/)

### RESEARCH INTERESTS

---

Statistical Natural Language Processing • Machine Learning • Information Extraction • Knowledge Base Population • Weakly Supervised Learning • Structured Prediction • Representation Learning

### RESEARCH EXPERIENCE

---

- since 03/2016    **Research Associate** in Machine Reading Group at Department of Computer Science, University College London
- 10/2015    –    **Research Associate** in Natural Language Processing Group at Department of Computer Science, The University of Sheffield
- 04 – 09/2015    **Research Associate (Part-Time)** in Natural Language Processing Group at Department of Computer Science, The University of Sheffield
- 10/2012    –    **PhD Student** at Department of Computer Science, The University of Sheffield
- 06/2010    –    **Research Assistant (Part-Time)** at AIFB, Karlsruhe Institute of Technology
- 09/2012
- 03 – 04/2010    **Research Intern** at AIFB, Karlsruhe Institute of Technology

## RESEARCH PROJECTS

---

- since 03/2016 Construction of Method and Algorithm Knowledge Graphs, funded by Elsevier  
*Researcher*
- 04/2015 – PHEME (Computing Veracity Across Media, Languages, and Social  
02/2016 Networks), EU FP7 ICT  
*Researcher*
- 10/2012 – LODIE (Linked Open Data for Information Extraction), EP/J019488/1  
11/2015 *PhD student associated with project*
- 10/2012 – Web Relation Extraction with Distant Supervision  
12/2015 *PhD topic*
- 2010 – 2012 LODifier - Generating Linked Data from Unstructured Text  
*Research topic at AIFB, BA thesis topic*

## ACADEMIC EDUCATION

---

- 10/2012 – Study of Computer Science (PhD)  
2016 **University of Sheffield**  
*Thesis topic:* Web Information Extraction using Linked Data  
*Supervisors:* Prof. Fabio Ciravegna, Dr. Diana Maynard  
PhD Thesis submitted: 12/2015, PhD awarded: 09/2016
- 2011 – 2012 Study of Computational Linguistics (M.A.)  
**Heidelberg University**
- 08/2011 B.A. Computational Linguistics and Psychology  
Final mark: *1.9*
- 2008 – 2011 Study of Computational Linguistics and Psychology (B.A.)  
**Heidelberg University**  
*Thesis topic:* LODifier - Generating RDF from Natural Language  
*Supervisors:* Prof. Sebastian Padó, Prof. Sebastian Rudolph

## SCHOOL EDUCATION

---

- July 2008 Abitur  
Final mark: *1.5*
- 2005 – 2008 Heinrich-Wieland-Schule, Technisches Gymnasium,  
Pforzheim
- 1999 – 2005 Theodor-Heuss-Gymnasium, Mühlacker
- 1995 – 1999 Ulrich-von-Dürrenz-Schule, Mühlacker

## OTHER EDUCATIONAL ACTIVITIES

---

- 07/2013            8th GATE Text Mining Summer Training Course, Sheffield, UK
- 01 - 05/2015      Machine Learning Online Course, Coursera, Tutor: Prof. Andrew Ng
- 07/2013            10th Summer School on Ontology Engineering and the Semantic Web (SSSW 2013), Cercedilla, Spain

## PUBLICATIONS

---

### SELECTED PUBLICATIONS

- **Isabelle Augenstein**, Tim Rocktäschel, Andreas Vlachos, Kalina Bontcheva. *Twitter Stance Detection with Bidirectional Conditional Encoding*. To appear in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), November 2016.
- George Spithourakis, **Isabelle Augenstein**, Sebastian Riedel. Numerically Grounded Language Models for Semantic Error Correction. To appear in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), November 2016.
- **Isabelle Augenstein**, Andreas Vlachos, Diana Maynard (2015). *Extracting Relations between Non-Standard Entities using Distant Supervision and Imitation Learning*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), September 2015.

### JOURNAL ARTICLES

- **Isabelle Augenstein**, Diana Maynard, Fabio Ciravegna (2016). *Distantly Supervised Web Relation Extraction for Knowledge Base Population*. Semantic Web Journal, Volume 7, Number 4 / 2016.
- Ziqi Zhang, Anna Lisa Gentile, **Isabelle Augenstein**, Eva Blomqvist, Fabio Ciravegna (2016). *An Unsupervised Data-driven Method to Discover Equivalent Relations in Large Linked Datasets*. Semantic Web Journal, 2016, To appear.

### PAPERS IN CONFERENCE PROCEEDINGS

- **Isabelle Augenstein**, Tim Rocktäschel, Andreas Vlachos, Kalina Bontcheva. *Twitter Stance Detection with Bidirectional Conditional Encoding*. To appear in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), November 2016.
- George Spithourakis, **Isabelle Augenstein**, Sebastian Riedel. Numerically Grounded Language Models for Semantic Error Correction. To appear in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), November 2016.

- Piroska Lendvai, **Isabelle Augenstein**, Kalina Bontcheva, Thierry Declerck. Monolingual Social Media Datasets for Detecting Contradiction and Entailment. Language Resources and Evaluation Conference (LREC), May 2016.
- **Isabelle Augenstein**, Andreas Vlachos, Diana Maynard (2015). *Extracting Relations between Non-Standard Entities using Distant Supervision and Imitation Learning*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), September 2015.
- **Isabelle Augenstein**, Diana Maynard, Fabio Ciravegna (2014). *Relation Extraction from the Web using Distant Supervision*. Proceedings of the 19th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2014), November 2014. **Nominated for best research paper**.
- **Isabelle Augenstein** (2014). *Joint Information Extraction from the Web using Linked Data*. Proceedings of the 12th International Semantic Web Conference (ISWC 2014), October 2014.
- Ziqi Zhang, Anna Lisa Gentile, Eva Blomqvist, **Isabelle Augenstein**, Fabio Ciravegna (2013). *Statistical Knowledge Patterns: Identifying Synonymous Relations in Large Linked Datasets*. Proceedings of the 12th International Semantic Web Conference (ISWC 2013), October 2013.
- Ziqi Zhang, Anna Lisa Gentile, **Isabelle Augenstein**, Eva Blomqvist, Fabio Ciravegna (2013). *Mining Equivalent Relations from Linked Data*. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: Short Papers (ACL 2013), August 2013.
- Anna Lisa Gentile, Ziqi Zhang, **Isabelle Augenstein**, Fabio Ciravegna (2013). *Unsupervised Wrapper Induction using Linked Data*. Proceedings of the 8th International Conference on Knowledge Capture (K-CAP 2013), June 2013.
- **Isabelle Augenstein**, Anna Lisa Gentile, Barry Norton, Ziqi Zhang, Fabio Ciravegna (2013). *Mapping Keywords to Linked Data Resources for Automatic Query Expansion*. Proceedings of the 10th Extended Semantic Web Conference (ESWC 2013), May 2013. **Best Workshop Paper of the 2nd International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data (KNOW@LOD 2013)**.
- **Isabelle Augenstein**, Sebastian Padó and Sebastian Rudolph (2012). *LODifier: Generating Linked Data from Unstructured Text*. Proceedings of the 9th Extended Semantic Web Conference (ESWC 2012), May 2012.

#### PAPERS IN WORKSHOP PROCEEDINGS

- Ben Eisner, Tim Rocktäschel, **Isabelle Augenstein**, Matko Bosnak, Sebastian Riedel. *emoji2vec: Learning Emoji Representations from their Description*. To appear in Proceedings of the International Workshop on Natural Language Processing for Social Media at EMNLP 2016, November 2016.
- **Isabelle Augenstein**, Andreas Vlachos and Kalina Bontcheva. *USFD at SemEval-2016 Task 6: Any-Target Stance Detection on Twitter with Autoencoders*. Proceedings of the International Workshop on Semantic Evaluation (SemEval 2016), June 2016.

- Leon Derczynski, **Isabelle Augenstein**, Kalina Bontcheva (2015). *USFD: Twitter NER with Drift Compensation and Linked Data*. Proceedings of the ACL Workshop on Noisy User-generated Text (W-NUT 2015), July 2015.
- **Isabelle Augenstein** (2014). *Seed Selection for Distantly Supervised Web-Based Relation Extraction*. Proceedings of the COLING Workshop on Semantic Web and Information Extraction (SWAIE 2014), August 2014.
- Eva Blomqvist, Ziqi Zhang, Anna Lisa Gentile, **Isabelle Augenstein**, Fabio Civravegna (2013). *Statistical Knowledge Patterns for Characterizing Linked Data*. Proceedings of the 4th Workshop on Ontology and Semantic Web Patterns (WOP 2013), October 2013.

#### OTHER

- Piroska Lendvai, **Isabelle Augenstein**, Dominic Rout, Kalina Bontcheva, Thierry Declerck. *Algorithms for Detecting Disputed Information: Final Version*. FP7-ICT Collaborative Project ICT-2013-611233 PHEME Deliverable D4.2.2 (WP4), June 2016.
- Ziqi Zhang, Anna Lisa Gentile, **Isabelle Augenstein** (2014). *Linked Data as Background Knowledge for Information Extraction on the Web*. ACM SIGWEB Newsletter, July 2014.

#### DISSERTATIONS

- **Isabelle Augenstein** (2016). Web Information Extraction using Distant Supervision. PhD theses: University of Sheffield, Department of Computer Science, UK, July 2016.
- **Isabelle Augenstein** (2011). *LODifier - Generating RDF from Natural Language*. Heidelberg University, Department of Computational Linguistics, August 2011.

#### PARTICIPATION IN SHARED TASKS

---

- 6th place in SemEval 2016 Task 6 “Detecting Stance in Tweets” SubTask B (weakly supervised task)
- 2nd place in ACL 2015 W-NUT Twitter Named Entity Recognition no-types challenge

## TALKS

---

- *Machine Reading for Social Media NLP*, University of Cambridge, Cambridge, UK. Invited talk in Cambridge NLP Seminar Series, November 2016.
- *Weakly Supervised Machine Reading*, University of Sheffield, Sheffield, UK. Invited talk in University of Sheffield NLP Seminar Series, October 2016.
- *Weakly Supervised Machine Reading*, University of Copenhagen, Copenhagen, DK. Invited talk and paid 1-week research visit in University of Copenhagen Computer Science department, October 2016.
- *Distant Supervision with Imitation Learning*, Heriot-Watt University, Edinburgh, UK. Invited talk in University of Edinburgh Computer Science Seminar Series, November 2015.
- *Distant Supervision with Imitation Learning*, University of Sheffield, Sheffield, UK. Talk in University of Sheffield NLP Seminar Series, November 2015.
- *Web Information Extraction using Distant Supervision*, University of Sheffield, Sheffield, UK. Talk in University of Sheffield NLP Seminar Series, January 2015.

## LANGUAGE SKILLS

---

- German: Native
- English: Fluent
- Swedish: Moderate
- French, Spanish, Danish: Basic

## PROGRAMMING AND SCRIPTING LANGUAGES

---

- Python, Java: Very Good
- LaTeX, HTML: Good
- C++, Bash, XML, Javascript: Moderate
- C, Assembler, Visual Basic, PHP, R: Basic

## IT EXPERIENCE

---

- Natural Language Processing Software: GATE, StanfordNLP, OpenNLP, NLTK, WordNet
- Machine Learning Software: scikit-learn, TensorFlow, Weka, Mallet, CRFSuite
- Numerical Analysis Software: Octave, R, SPSS, gnuplot
- Other Software: Tomcat, Hadoop, Eclipse, PyCharm, Lucene, Jena, GraphDB (OWLIM), Amazon Web Services
- Operating systems: Linux, Mac OS, Windows
- Databases: MySQL

## ORGANISING COMMITTEES

---

2016	11th Workshop on Semantic Evaluation (SemEval 2017), Primary organiser of Task 10, ScienceIE: Extracting Keyphrases and Relations from Scientific Publications
2015	5th European Semantic Web Summer School (ESWCSS 2015), Community Manager
2014	2nd University of Sheffield Engineering Symposium (USES 2014), Artificial Intelligence & Virtual Reality Track Chair
2013 - 2014	President of USES, the University of Sheffield Postgraduate Society
2008 - 2012	Member of the student council at the Department of Computational Linguistics, Heidelberg University

## PROGRAMME COMMITTEES & REVIEWING

---

- Reviewing for Journals: Transactions of the Association for Computational Linguistics (TACL), Information Processing & Management (IP&M)
- Programme Committee Member for Conferences: EACL 2017 (Information Extraction Track), COLING 2016 (Applications Track), EMNLP 2016 (Machine Learning Track), EMNLP 2015 (Information Extraction Track), ISWC 2015 (Research Track), SEMAPRO 2015, ESWC 2015 (Poster & Demo Track), WWW 2015 (Research Track), ISWC 2014 (Research Track, In Use Track), SEMAPRO 2014
- Programme Committee Member for Workshops: SemEval 2016, KET at WWW 2015, AHA!-Workshop at COLING 2014, LD4IE at ISWC 2014, LD4IE at ISWC 2013, NLP&LOD 2013 at RANLP 2013
- Subreviewer for Conferences: ACL 2015, ESWC 2014 (Poster & Demo Track)

## TEACHING EXPERIENCE

---

### LECTURES & TUTORIALS

- |             |  |
|-------------|--|
| 2016        | Guest Lecturer at Department of Computer Science,<br>University College London<br><i>Module:</i> Statistical Natural Language Processing (Dr. Sebastian Riedel). |
| 09/2015     | Bridges Summer School 2016, Tutor, Tutorial on Practical Machine Learning for Social Media Analysis  |
| 09/2015     | ESWC Summer School 2015, Tutor, Hands-On session on NLP for the Semantic Web   |
| 09/2014     | ESWC Summer School 2014, Tutor, Hands-On session on Linked Data for NLP with GATE, Secondary tutor for two student projects                                      |
| 2013 – 2014 | Demonstrator at Department of Computer Science,<br>University of Sheffield<br><i>Modules:</i> Web Technologies, The Intelligent Web (Prof. Fabio Ciravegna)      |

### STUDENT SUPERVISION

- George Spithourakis. *Multimodal distributional semantics for understanding clinical language*. PhD in Computer Science Dissertation, expected submission in 2017, co-supervised with Sebastian Riedel, Steffen Petersen.
- Danish Sheikh. *Extracting Keyphrases and Relations from Scientific Publications*. MEng in Computer Science Dissertation, to submit in 2017.
- Edward Collins. *Generating Automatic Summaries of Scientific Papers*. MEng in Computer Science Dissertation, to submit in 2017.
- Rupert Chaplin. *End-to-end Differentiable and Interpretable Deep Learning for Natural Language Programming*. MSc in Machine Learning Dissertation, 2016, co-supervised with Tim Rocktäschel.
- Dhruv Ghulati. *Distant Supervision and Cost Sensitive Classification for Weakly Supervised Claim Detection*. MSc in Computer Science Dissertation, 2016, co-supervised with Sebastian Riedel, George Spithourakis.



## AWARDS, SCHOLARSHIPS, GRANTS, RESEARCH VISITS

---

- 2016 Invited to 1-week **Research Visit** in Computer Science Department, University of Copenhagen, paid for flights, accommodation and meals
- 2015 Invited to 4-day **Google NLP PhD Summit** in Zurich, paid for flights, accommodation and meals
- 2014 Nominated for EKAW 2014 best research paper award
- 2014 **ISWC 2014 student travel grant**  
€800 for registration and accommodation
- 2013 Best workshop paper of the 2nd International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data (KNOW@LOD)
- 2012 **Department of Computer Science, University of Sheffield PhD Studentship**  
Stipend for a full 3-year research scholarship, plus tuition fees
- 2008 **Literary Society Scheffelbund, Scheffel prize**  
<http://web1.karlsruhe.de/Kultur/MLO/>  
Prize for the best exam results in German obtained by school leavers at grammar schools in the states of Baden-Württemberg  
€80, 5-year membership in Scheffelbund and book subscription

## REFERENCES

---

On request

London, 11 October 2016