Sina Ahmadi

The Insight Centre for Data Analytics
National University of Ireland
Galway, H91 AEX4 Ireland

email: ahmadi.sina@outlook.com
URL: http://sinaahmadi.github.io/

Research Interests

- Computational lexicography
- Natural language processing
- Machine learning
- Kurdish language processing

Education

April 2018 - present	**Ph.D. in Computer Science and Engineering** **Dissertation Advisor**: John P. McCrae, Ph.D. ABD – completed coursework requirements National University of Ireland Galway, Ireland College of Engineering and Informatics
Sep. 2016 - Sep. 2017	**Master of Science in Artificial Intelligence (Machine Learning for Data Mining)** **Thesis**: *"Attention-based Encoder-Decoder Networks for Spelling and Grammatical Error Correction"*, supervised by Joseph Le Roux, Ph.D. and Nadi Tomeh, Ph.D. Paris Descartes University, Paris, France Department of Mathematics and Computer Science
Sep. 2014 - Sep. 2016	**Master of Arts in Natural Language Processing** **Thesis**: *"From Text to Text: Natural Text Generation Using Patterns from Corpus"*, supervised by Patrick Paroubek, Ph.D. Sorbonne Nouvelle University, Paris, France Department of General and Applied Linguistics
Sep. 2010 - June 2014	**Bachelor of Science in Software Engineering** **Thesis**: *"Part-of-speech tagging for Kurdish using a statistical and rule-based approach"*, supervised by Kyumars Sheykh Esmaili, Ph.D. University of Kurdistan, Sanandaj, Iran Department of Computer and Electrical Engineering

Experience

June 2019	**Scientific visit,** Society for Danish Language and Literature & Centre for Language Technology in Copenhagen, Denmark A preliminary experiment was conducted on Danish lexicographic data linking, particularly senses in two monolingual dictionaries *Ordbog over det danske Sprog* (ODS) and *Den Danske Ordbog* (DDO).

April 2017 - Sept. 2017	**Research intern**, the Computer Science Laboratory of Paris-North University (LIPN) Developed neural network models for spelling and grammatical error correction. The performance of (Bidirectional) recurrent neural networks and encoder-decoder networks along with attention mechanism was evaluated.
Jan. 2016 - June 2016	**Research intern**, Computer Science Laboratory for Mechanics and Engineering Sciences (LIMSI-CNRS) Natural Language Generation for recommendation of commercial products. This internship was financed by an industrial partner.
June 2015 - Sep. 2015	**Research intern**, Laboratory of Informatics of Paris6 (LIP6), UPMC Co-reference resolution by representation learning using neural networks, restricted Boltzmann machine and deep Boltzmann network
Jan. 2013 - Dec. 2013	**Research volunteer**, Kurdish Language Processing Project, University of Kurdistan Under the supervision of Kyumars Sheykh Esmaili Implementation of fundamental tools for Kurdish processing, e.g. stemmer, lemmatizer, spell-checker and, resources such as KurdNet–the Kurdish WordNet

Publications

MAIN RESEARCH TOPICS

2020

1. (accepted) Sina Ahmadi, Sanni Nimb, Thomas Troelsgård, John P. McCrae, and Nicolai H. Sørensen. Towards Automatic Linking of Lexicographic Data: the case of a historical and a modern Danish dictionary. In *The XIX EURALEX International Congress*, Alexandroupolis, Greece, 2020

2. (accepted) Patricia Martín-Chozas, Sina Ahmadi, and Elena Montiel-Ponsoda. Defying Wikidata: Validation of Terminological Relations in the Web of Data. In *The 12th International Conference on Language Resources and Evaluation (LREC)*, Marseille, France, 2020

3. (accepted) Sina Ahmadi, John P. McCrae, Sanni Nimb, Fahad Khan, Monica Monachini, Bolette S. Pedersen, Thierry Declerck, Tanja Wissik, Andrea Bellandi, Irene Pisani, Thomas Troelsgård, Sussi Olsen, Simon Krek, Veronika Lipp, Tamás Váradi, László Simon, András Győrffy, Carole Tiberius, Tanneke Schoonheim, Yifat Ben Moshe, Maya Rudich, Raya Abu Ahmad, Dorielle Lonke, Kira Kovalenko, Margit Langemets, Jelena Kallas, Oksana Dereza, Theodorus Fransen, David Cillessen, David Lindemann, Mikel Alonso, Ana Salgado, José Luis Sancho, Rafael-J. Ureña-Ruiz, Kiril Simov, Petya Osenova, Zara Kancheva, Ivaylo Radev, Ranka Stanković, Andrej Perdih, and Dejan Gabrovšek. A Multilingual Evaluation Dataset for Monolingual Word Sense Alignment. In *Proceedings of the 12th Language Resource and Evaluation Conference (LREC 2020)*, Marseille, France, 2020

2019

1. Speranza Giulia, Carola Carlino, and Sina Ahmadi. Creating a multilingual terminological resource using linked data: the case of archaeological domain in the italian language. *The Sixth Italian Conference on Computational Linguistics - CLiC-it 2019*, 2019

2. John P. McCrae, Carole Tiberius, Anas Fahad Khan, Ilan Kernerman, Thierry Declerck, Simon Krek, Monica Monachini, and Sina Ahmadi. The elexis interface for interoperable lexical resources. *Proceedings of Sixth Biennial Conference on Electronic Lexicography (eLex 2019)*, 2019

3. Sina Ahmadi* and Tobias Daudert*. CoFiF: A corpus of financial reports in French language. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, pages 21–26, Macao, China, 12 August 2019

4. Tobias Daudert and Sina Ahmadi. NUIG at the FinSBD task: Sentence boundary detection for noisy financial PDFs in English and French. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, pages 108–114, Macao, China, 12 August 2019

5. Mihael Arcan, Daniel Torregrosa, Sina Ahmadi, and John P McCrae. Inferring translation candidates for multilingual dictionary generation with multiway neural machine translation. In *Proceedings of the Translation Inference Across Dictionaries Workshop (TIAD 2019)*, Leipzig, Germany, 2019. Proceedings of the LDK 2019 Workshops

6. Mihael Arcan, Daniel Torregrosa, Sina Ahmadi, and John P McCrae. Tiad 2019 Shared Task: Leveraging knowledge graphs with neural machine translation for automatic multilingual dictionary generation. *Shared Task on Translation Inference Across Dictionaries*, 2019

7. Sina Ahmadi, Mihael Arcan, and John McCrae. Lexical sense alignment using weighted bipartite b-matching. In *Proceedings of the LDK 2019 Workshops*. 2nd Conference on Language, Data and Knowledge (LDK 2019), 2019

2018

1. Sina Ahmadi, Mihael Arcan, and John McCrae. On lexicographical networks. In *Workshop on eLexicography: Between Digital Humanities and Artificial Intelligence*. Workshop on eLexicography: Between Digital Humanities and Artificial Intelligence, 2018

2. Sina Ahmadi. Learning noun cases using sequential neural networks. *arXiv preprint arXiv:1810.03996*, 2018

3. Sina Ahmadi. Attention-based encoder-decoder networks for spelling and grammatical error correction. *arXiv preprint arXiv:1810.00660*, 2018

KURDISH LANGUAGE PROCESSING

2020

1. (accepted) Sina Ahmadi and Hossein Hassani. A Corpus of the Sorani Kurdish Folkloric Lyrics. In *Proceedings of the 1st Joint Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL) Workshop at the 12th International Conference on Language Resources and Evaluation (LREC)*, Marseille, France, 2020

2. (under review) Sina Ahmadi and Hossein Hassani. Towards Finite-State Morphology of Kurdish. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*

2019

1. Sina Ahmadi, Hossein Hassani, and John McCrae. Towards electronic lexicography for the Kurdish language. *Proceedings of Sixth Biennial Conference on Electronic Lexicography (eLex 2019)*, 2019

2. Roshna Abdulrahman, Hossein Hassani, and Sina Ahmadi. Developing a fine-grained corpus for a less-resourced language: the case of kurdish. *WiNLP ACL 2019*, 2019

3. Sina Ahmadi. A rule-based kurdish text transliteration system. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 18(2):18, 2019

2017

1. Shahin Salavati and Sina Ahmadi. Building a lemmatizer and a spell-checker for sorani kurdish. In *Proceedings of the 8th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan, Poland, 2017

2014

1. Purya Aliabadi, Sina Ahmadi, Shahin Salavati, and Kyumars Sheykh Esmaili. Towards building Kurdnet, the Kurdish Wordnet. In *Proceedings of the Seventh Global Wordnet Conference*, pages 1–6, 2014

Poster & Invited talks

December 2019	Workshop on eLexicography, Galway, Ireland
September 2019	Insight Student Conference, University College Dublin
June 2018	AI & the Smart Enterprise, Dublin City University, Dublin

Academic activities

Reviewing

PC member at 10th Global WordNet Conference 2019 (GWC2019)

Teaching Assistant

Spring 2019	Advanced Natural Language Processing (CT5121) at NUI Galway (laboratory sessions)
May 2019	Tutor at the 3rd Summer Datathon on Linguistic Linked Open Data, Dagstuhl, Germany
Spring 2020	Advanced Natural Language Processing (CT5121) at NUI Galway (laboratory sessions)
	Linked Data (DER5101) at NUI Galway (laboratory sessions)

Supervision

June-August 2019	Victor Troussel's internship at the Unit for Linguistic Data, Insight Centre for Data Analytics (Master's student from École Polytechnique de l'Université de Nantes)

Organizational

Spring 2020	Insight Student Conference 2020 (studentconference2020.insight-centre.org)

Honors & awards

February 2019   Top ranked team in Ireland at Google Hash Code 2019
2018-2022   ELEXIS grant for Ph.D. degree at NUI Galway
2011   University of Kurdistan's member in the 36th ACM International Collegiate Programming Contest (Asian Regional), Tehran
2008   Ranked 2nd in Kurdistan Regional High-school Competence of Programming, Iran, 2008

Technical skills

Programming Languages: *Expert in*: Python, Java, Perl, R - *Familiar with*: C++, Octave, C#
Tools and Frameworks: RStudio, Visual Studio, SciPy, Theano, DyNet, LaTeX
Database : MySQL, SPARQL
Web technologies: Jekyll, PHP, HTML (and CSS), XML technologies
Interpersonal Skills: Critical Thinking, Data Collection, Literature Reviews, Problem Solving, Self-Management, Confidentiality

Languages

- Kurdish (native)

- English, French, Persian, Arabic

- Modern Greek (B1), German (A1)

Personal interests

Writing, violin playing (2 public performances so far) - Reading - Travelling - Learning

References

Available on demand.