# HASSAN S. SHAVARANI

Contact Information		Research Interests
Email: Website: LinkedIn: Nationality:	sshavara@sfu.ca https://shavarani.github.io www.linkedin.com/in/shavarani Canadian	<ul> <li>Neural Entity Linking</li> <li>Information Extraction</li> <li>Natural Language Processing (esp. Med.</li> <li>Deep Learning</li> </ul>

## BIO

I am a PhD student at the Natural Language Processing Lab in the School of Computing Science at Simon Fraser University in Vancouver, Canada. My research uses pre-trained language models and structured prediction based fine-tuning methods. My published work includes using structured prediction to extract linguistically useful representation learning from pre-trained models for neural machine translation, work on policy generation for simultaneous neural machine translation, and more recently my work involves novel pre-training, fine-tuning and structured prediction inference for entity linking to ontologies like Wikipedia.

# EDUCATION

Ph.D. in Computing Science Sep. 2017 - Present GPA: 4.25/4.00 Simon Fraser University M.Sc. in Computing Science Sep. 2014 - May 2016 GPA: 3.93/4.00 Simon Fraser University **B.Sc.** in Information Technology Sep. 2009 - Apr. 2014 Amirkabir University of Technology - Tehran Polytechnic GPA: 17.69/20

# **Research Experience**

Ph.D. Research Assistant

Simon Fraser University

Sep. 2017 - Present Advisor: Dr. Anoop Sarkar

- Investigated structured prediction for simplifying and improving various NLP tasks relying on pretrained language models as part of a Ph.D. dissertation.
- Conducted research on neural entity linking, neural machine translation and multi-lingual document classification.
- Presented research findings at EMNLP'23, EACL'21, and LREC'20.
- Developed SPEL, the SoTA entity linking framework using structured prediction, and SFUTrans*late*, an academic neural machine translation toolkit with a focus on readability and changeability.
- Created *AIDA/testc*, a new challenging test set to evaluate modern entity linking models, containing 131 annotated news articles.
- Created SHINRA-5LDS, a large multi-lingual and multi-labeled set of annotated Wikipedia articles in Japanese, English, French, German, and Farsi using Extended Named Entity (ENE) tag set.
- Published 8 research papers in the course of the Ph.D. program.
- Supervised 2 research projects in the course of the Ph.D. program.
- Collaborated with Language Information Access Technology Team at Riken Center for Advanced Intelligence Project, Tokyo, Japan.

#### **Research and Development Associate** Emtelligent Inc.

Aug. 2020 - Present Supervisor: Dr. Anoop Sarkar

- Designed and implemented deep learning models for entity linking in medical NLP.
- Conducted research on the design of large-scale unsupervised neural models for multi-task learning in medical NLP.

dical)

# Data Science Intern

Traction on Demand

Jan. 2019 - Apr. 2019 Supervisor: Vince Liu

May 2018 - Aug. 2018

- Conducted research on negative opinion mining in slack messages from employees.
- Performed data anonymization to remove traces of the employees.
- Performed *sentiment analysis* using OpenAI's sentiment neuron and conducted exploratory data analysis to identify trends and patterns in collected messages.
- Presented insights and findings to team members and stakeholders through clear and concise reports.

#### Research Intern

Riken Center for Advanced Intelligence Project

- Developed a graph-based algorithm to expand the SHINRA Japanese dataset to English, French, German, and Farsi using the Wikipedia language links.
- Benchmarked the created dataset over the multi-class multilingual classification techniques existing at the time.

#### **Research Intern**

TextIQ Labs Canada Inc.

Jun. 2016 - Sep. 2016 Supervisor: Dr. Apoorv Agarwal

Supervisor: Dr. Satoshi Sekine

- Designed and implemented an attention-based deep learning model for information extraction from unstructured email text.
- Created a data visualization tool based on brat rapid annotation tool to demonstrate the extracted information.

#### M.Sc. Research Assistant

Simon Fraser University

- Conducted research on simultaneous text-to-text machine translation.
- Conducted experiments and analyzed data, published and presented research findings at IWSLT'15.
- Developed expertise in statistical NLP with a special focus on statistical machine translation.

## **B.Sc.** Research Assistant

Amirkabir University of Technology

May 2013 - Apr. 2014 Advisor: Dr. Shahram Khadivi

Sep. 2014 - May 2016

Advisor: Dr. Anoop Sarkar

- Conducted research on a multi-hop Farsi news document classification technique.
- Conducted experiments on using machine translation to translate Farsi articles into English and perform unsupervised document classification through keyword extraction and semantic similarity between document class labels and the document keywords.

## PUBLICATIONS

## **Conference Papers:**

- 1. Hassan S. Shavarani, and Anoop Sarkar. "SPEL: Structured Prediction for Entity Linking" In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023.
- Ashkan Alinejad, Hassan S. Shavarani, and Anoop Sarkar. "Translation-based supervision for policy generation in simultaneous neural machine translation." In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 1734-1744. 2021.
- 3. Hassan S. Shavarani, and Anoop Sarkar. "Better Neural Machine Translation by Extracting Linguistic Information from BERT." In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pp. 2772-2783. 2021.
- 4. Hassan S. Shavarani, and Satoshi Sekine. "Multi-class Multilingual Classification of Wikipedia Articles Using Extended Named Entity Tag Set." In Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 1197-1201. 2020.
- 5. Jetic Gū, Hassan S. Shavarani, and Anoop Sarkar. "Top-down Tree Structured Decoding with Syntactic Connections for Neural Machine Translation and Parsing." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 401-413. 2018.
- Maryam Siahbani, Hassan S. Shavarani, Ashkan Alinejad, and Anoop Sarkar. "Simultaneous translation using optimized segmentation." In Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track), pp. 154-167. 2018.

7. Hassan S. Shavarani, Maryam Siahbani, Ramtin Mehdizadeh Seraj, and Anoop Sarkar. "Learning segmentations that balance latency versus quality in spoken language translation." In Proceedings of the 12th International Workshop on Spoken Language Translation: Papers, pp. 217-224. 2015.

# Preprint and ArXiv Papers:

- 1. Nasrin Seifi, and **Hassan S. Shavarani**. "Predicting Idiosyncratic Volatility from Stock Market Trade Records: A Machine Learning Approach." (2023).
- 2. Jetic Gū, Hassan S. Shavarani, and Anoop Sarkar. "Pointer-based fusion of bilingual lexicons into neural machine translation." arXiv preprint arXiv:1909.07907 (2019).

# Awards and Scholarships

Upper Bound Talent Bursary Alberta Machine Intelligence Institute (Amii) Clark, Wilson Graduate Scholarship - Ebco Eppich Competition	Summer 2023
Simon Fraser University Helmut and Hugo Eppich Family Graduate Scholarship - Ebco Ep	Spring 2022 Opich Competition
Simon Fraser University Faculty of Applied Sciences Graduate Fellowship Award	Spring 2020 & Spring 2021
Simon Fraser University MITACS Research Training Award	Fall 2020
Mitacs Computing Science Travel Award	Summer 2020
Simon Fraser University Graduate Fellowship Award	Summer 2020
Simon Fraser University Century 21 Charlwood Family Award - Ebco Eppich Competition	2015-2020
Simon Fraser University	Spring 2019

## SERVICE

• Peer reviewed in CanAI 2019, ACL 2020, EMNLP 2021, ACL 2022 - ARR, ARR 2023