

# Yoshinari Fujinuma

## Office Address

Virtual

## Contact

fujinumay@gmail.com

<http://akkikiki.github.io>

Twitter: @akkikiki

## Education

University of Colorado Boulder, USA

Dec 2021

Ph.D. in Computer Science

Advisors

- Katharina Kann(2020-2021), Jordan Boyd-Graber(-2021), Michael J. Paul (-2019)

University of Tokyo, Japan

Sep 2014

M.S. in Information Science and Technology; Advisor: Akiko Aizawa

International Christian University, Japan

Mar 2012

B.A. in Computer Science and Mathematics; Advisor: Grant Pogosyan

## Professional Experience

**Applied Scientist**, Amazon Web Services Inc, USA

Sep 2021 - Present

**Part-Time Instructor**, Univ. of Colorado Boulder, USA

Jan 2021 - May 2021

- CSCI 4622 Machine Learning

- Link: <https://github.com/akkikiki/CSCI-4622-Machine-Learning-sp21>

**Teaching Assistant**, Univ. of Colorado Boulder, USA

Jan-Dec 2020, May-Jul 2021

- CSCI 2270 Data Structures

**Applied Scientist Intern**, Amazon Web Services Inc, USA

May 2020 - Aug 2020

**Research Assistant**, Univ. of Colorado Boulder, USA

Aug 2016 - Dec 2019

**Applied Scientist Intern**, Amazon.com, USA

May 2018 - Aug 2018

**Software Development Engineer**, Amazon/A9.com, Japan

Oct 2014 - Aug 2016

**Software Engineer Intern**, Amazon/A9.com, Japan

Nov 2013 - Feb 2014

**Part-time Engineer**, Atilika, Japan

Aug- Nov 2013, Apr- Sep 2014

**Software Engineer Intern**, Cookpad, Japan

July 2013 (one month)

## Publications

- Yoshinari Fujinuma, Masato Hagiwara: “Semi-Supervised Joint Estimation of Word and Document Readability”, TextGraphs-15@NAACL (short) 2021
- Mozhi Zhang\*, Yoshinari Fujinuma\*, Michael J. Paul, Jordan Boyd-Graber: “Why Overfitting Isn’t Always Bad: Retrofitting Cross-Lingual Word Embeddings to Dictionaries”, ACL (short) 2020
- Mozhi Zhang, Yoshinari Fujinuma, Jordan Boyd-Graber: “Exploiting Cross-Lingual Subword Similarities in Low-Resource Document Classification”, AAAI 2020
- Yoshinari Fujinuma, Jordan Boyd-Graber, Michael J. Paul: “A Resource-Free Evaluation Metric for Cross-Lingual Word Embeddings based on Graph Modularity”, ACL (long) 2019

- Dasha Pruss, Yoshinari Fujinuma, Ashlynn R. Daughton, Michael J. Paul, Brad Arnot, Danielle Albers Szafr, Jordan Boyd-Graber: “Zika discourse in the Americas: A multilingual topic analysis of Twitter”, PLOS ONE 2019
- Yoshinari Fujinuma, Alvin Grissom II: “Substring Frequency Features for Segmentation of Japanese Katakana Words with Unlabeled Corpora”, IJCNLP (short) 2017
- Yoshinari Fujinuma, Hikaru Yokono, Pascual Martínez-Gómez, Akiko Aizawa: “Distant-supervised Language Model for Detecting Emotional Upsurge on Twitter”, PACLIC (long) 2015

\*denotes equal contribution

<b>Selected Projects</b>	<b>Analysis on Multilingual Pretraining</b>	2020 - 2021
	<ul style="list-style-type: none"> <li>• Investigating the effect of using different pretraining languages</li> </ul>	
	<b>Intrinsic Evaluation Measure for Cross-Lingual Embeddings</b>	2017 - 2019
	<ul style="list-style-type: none"> <li>• Developed a graph-based intrinsic measure to evaluate the quality of cross-lingual word embeddings.</li> <li>• Paper link: <a href="https://www.aclweb.org/anthology/P19-1489">https://www.aclweb.org/anthology/P19-1489</a></li> </ul>	
	<b>Finite State Transducer (FST) for Kuromoji</b>	2015
	<ul style="list-style-type: none"> <li>• Replaced a double-array trie to an FST to build a dictionary for Kuromoji, a java-based Japanese tokenizer used in Lucene, Solr, and Elastic Search.</li> <li>• Code available at <a href="https://github.com/atilika/fst">https://github.com/atilika/fst</a></li> </ul>	
<b>Academic Service</b>	<ul style="list-style-type: none"> <li>• Program Committee/Reviewer: <ul style="list-style-type: none"> <li>– 2022 AAAI, ACL Rolling Review</li> <li>– 2021 NAACL, ACL, EMNLP, CoNLL, ACL SRW, W-NUT, Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)</li> <li>– 2020 EMNLP, AACL SRW, ACL-SRW, W-NUT</li> <li>– 2019 Workshop on Noisy User-Generated Text (W-NUT)</li> </ul> </li> <li>• Secondary Reviewer: <ul style="list-style-type: none"> <li>– 2019 ACL</li> <li>– 2017 EMNLP, WWW</li> </ul> </li> </ul>	
<b>Academic Honors</b>	• Travel Grant (CU Boulder)	2017
	• Dean’s Fellowship (CU Boulder)	2016
	• Best Bachelor thesis in CS and Math (Gödel Foundation Prize)	2012
<b>Computer and Language Skills</b>	<u>Languages:</u>	Proficient: Python; Intermediate: C++, Java
	<u>Software:</u>	PyTorch, Git, Vim, L <sup>A</sup> T <sub>E</sub> X, MySQL
	<u>English:</u>	TOEFL iBT 101 (2015)
	<u>Domain-specific:</u>	machine learning, natural language processing