Masato Hagiwara

hagisan@gmail.com
http://masatohagiwara.net/

SKILLS

Natural language processing and machine/deep learning, with a focus on language education & assessment, machine translation/transliteration, and Asian language processing.

EXPERIENCE

Octanove Labs LLC, Pittsburgh, PA — *Owner & NLP/ML Engineer*

Feb 2019 - Present

- Consult early-to-mid stage startups in the US and Japan on their Artificial Intelligence strategies
- Build Machine Learning & Natural Language Processing solutions for AI startups
- Propose protocol & framework for integrated writing environments (https://www.teaspn.org/)

Duolingo, Inc., Pittsburgh, PA — *Senior Machine Learning Engineer / Scientist*

Feb 2015 - Feb 2019

- Built automatic grading technologies for Duolingo English Test writing and speaking questions using neural networks
- Led data creation and analysis for various research projects, including user behavior analysis and second language acquisition modeling (SLAM) shared task
- Led the content creation of Chinese, Japanese, and Korean from English courses

Rakuten Institute of Technology, New York, NY — *Lead Scientist*

Oct 2010 - Feb 2015

- Developed machine transliteration (NLP2011 paper award) and machine translation algorithms for the largest Japanese e-commerce website (Rakuten)
- Built a Chinese/Japanese word segmentation / morphological analyzer (RakutenMA)
- Developed a writing support system for English as a Second Language (ESL) learners

Baidu Japan, Inc, Shanghai / Beijing / Tokyo — *Research and Development Engineer*

Apr 2009 - Sep 2010

- Improved the ranking and page analysis algorithms including spam detection and emoticon search for Baidu mobile search
- Worked as a consultant on various NLP projects including Japanese Input Method BaiduType
- Led NLP data initiatives including Unnatural language processing contest and Baidu mobile corpus and timed corpus

Microsoft Research, Redmond, WA — *Research Intern*

Apr 2008 - Jul 2008

- Built a state-of-the-art method for Japanese query alteration for spelling correction and spelling/transliteration normalization
- Implemented the system using Visual C#, SQL Server, and Ruby, with tens of gigabytes of query log, which was integrated into Microsoft Live Search
- Published a research paper on the query alteration algorithm at NAACL 2009 and at the 3rd NLP Symposium for Young Researchers (Outstanding Presentation Award)

Google, Mountain View, CA — *Software Engineer Intern*

Aug 2005 - Sepl 2005

- Improved Japanese query suggestion, which is currently used as the basis for the query suggestion shown at the top and bottom of the Google search result
- Ran knowledge extraction algorithms on the distributed computation infrastructure (MapReduce and the Google's large network clusters)

EDUCATION

Nagoya University, Nagoya, Japan - *Ph.D., Information Engineering (Computer Science)*

Apr 2006 - Mar 2008

- Doctoral Thesis: "Modeling and Selection of Context for Better Synonym Acquisition"

Nagoya University, Nagoya, Japan - *Master's Degree., Information Engineering (Computer Science)*

Apr 2004 - Mar 2006

- Skipped a year in undergraduate due to the excellent academic performance. Overall GPA: 3.8
- Master's Thesis: "Utilization of Probabilistic Latent Semantics for Automatic Thesaurus Construction

Nagoya University, Nagoya, Japan - *Bachelor's Degree., Information Engineering (Computer Science)*

Apr 2001 - Mar 2004

- Overall GPA: 3.9

AWARDS AND PROFESSIONAL ACTIVITIES

- Co-organizer of the Workshop for Natural Language Processing Open Source Software (NLP-OSS), co-located at ACL 2018
- Invited keynote at the Optimizing Human Learning workshop co-located with ITS 2018 (Montréal, Canada, June 2018)
- Invited talk at CUNY NLP Seminar (hosted by Prof. Heng Ji) Title: Word Segmentation and Transliteration in Chinese and Japanese, April 2013
- 2011 Field Innovation Award from the Japanese Society for Artificial Intelligence: ANPI_NLP: Safety Information Confirmation Support using Natural Language Processing for The 2011 Tohoku Earthquake.

- Paper Award at NLP2011 “Latent Class Transliteration based on Source Language Origins” (the largest Japanese NLP academic conference)
- Best Paper Award at NLP2009 “Semantic Category Extraction from Unsegmented Text using Graph Kernels” (the largest Japanese NLP academic conference, chosen among 235 papers)
- Paper Award at the 3rd NLP Symposium for Young Researchers. Presentation: “A Unified Approach to Japanese Query Alteration based on Semantic Similarity

SELECTED PUBLICATIONS

BOOKS

- Masato Hagiwara. Real-World Natural Language Processing. Manning Publications, 2019.
- Yoh Okuno, Graham Neubig, Masato Hagiwara, Mamoru Komachi. Natural Language Processing: Basics and Technology (Shoeisha) (in Japanese). Shoeisha, 2016.
- Drew Conway, John Myles White. Masato Hagiwara, Yoh Okuno, Takaaki Mizuno, Tetsuya Kinoshita (translation). 入門 機械学習 (Machine Learning for Hackers). O'Reilly Japan, 2012.
- Steven Bird, Ewan Klein, Edward Loper. Masato Hagiwara, Takahiro Nakayama, Takaaki Mizuno (translation). 入門 自然言語処理 (Natural Language Processing with Python). O'Reilly Japan, 2010.

JOURNAL PAPERS

- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Supervised Synonym Acquisition Using Distributional Features and Syntactic Patterns. Journal of Natural Language Processing, Vol. 16, Num. 2, pp. 59-83, 2009.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. A Comparative Study on Effective Context Selection for Distributional Similarity. Journal of Natural Language Processing, Vol. 5, Num. 5, pp. 119-150, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Effective Use of Indirect Dependency for Distributional Similarity. Journal of Natural Language Processing, Vol. 15, Num. 4, pp. 19-42, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Bootstrapping-based Extraction of Dictionary Terms from Unsegmented Legal Text. New Frontiers in Artificial Intelligence: JSAI 2008 Conference and Workshops, Revised Selected papers, Lecture Notes in Computer Science, Vol. 5447, pp. 213-227, 2009.

CONFERENCE PAPERS

- Masato Hagiwara, Takumi Ito, Tatsuki Kuribayashi, Jun Suzuki and Kentaro Inui. TEASPN: Framework and Protocol for Integrated Writing Assistance Environments. EMNLP (system demonstrations), 2019 (to appear).
- Burr Settles, Chris Brust, Erin Gustafson, Masato Hagiwara, Nitin Madnani. Second Language Acquisition Modeling. BEA 2018, 2018.
- Ayah Zirikly, Masato Hagiwara. Cross-lingual Transfer of Named Entity Recognizers without Parallel Corpora. ACL 2015, pp. 390-396, 2015.

- Masato Hagiwara, Satoshi Sekine. Lightweight Client-Side Chinese/Japanese Morphological Analyzer Based on Online Learning. COLING 2014 system demonstration, pp. 39-43, 2014.
- Haibo Li, Masato Hagiwara, Qi Li, Heng Ji. Comparison of the Impact of Word Segmentation on Name Tagging for Chinese and Japanese, LREC 2014, pp.2532-2536, 2014.
- Masato Hagiwara, Satoshi Sekine. Accurate Word Segmentation using Transliteration and Language Model Projection, ACL 2013, pp 183-189.
- Masato Hagiwara, Soh Masuko. KooSHO: Japanese Text Input Environment based on Aerial Hand Writing. NAACL/HLT 2013, demo session, pp. 24-27. 2013.
- Yuta Hayashibe, Masato Hagiwara, Satoshi Sekine. phloat : Integrated Writing Environment for ESL learners, Second Workshop on Advances in Text Input Methods (WTIM 2012), 2012.
- Masato Hagiwara, Satoshi Sekine. Latent Semantic Transliteration using Dirichlet Mixture. NEWS 2012 (the 4th Named Entities Workshop), pp. 30-37, 2012.
- Graham Neubig, Yuichiroh Matsubayashi, Masato Hagiwara, Koji Murakami. Safety Information Mining — What can NLP do in a disaster —, Proc. of IJCNLP 2011.
- Masato Hagiwara and Satoshi Sekine. Latent Class Transliteration based on Source Language Origins. Proc. of ACL-HLT 2011, pp. 53-57, 2011.
- Masato Hagiwara and Hisami Suzuki. Japanese Query Alteration Based on Lexical Semantic Similarity. Proc. of NAACL HLT 2009, pp. 191-199, 2009.
- Nobuyuki Shimizu, Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama and Hiroshi Nakagawa. Metric learning for synonym acquisition. Proc. of COLING 2008, pp. 793-800, 2008.
- Masato Hagiwara. A Supervised Learning Approach to Automatic Synonym Identification based on Distributional Features. Proc. of ACL 2008 Student Research Workshop, pp. 1-6, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Context Feature Selection for Distributional Similarity. Proc. of IJCNLP 2008, pp. 553-560, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Effective Proximity Distance for Word-Based Context. Proc. of SNLP 2007, pp. 105-110, 2007.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Effectiveness of Indirect Dependency for Automatic Synonym Acquisition. Proc. of CoSMo 2007, pp. 1 - 8, 2007.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Selection of Effective Contextual Information for Automatic Synonym Acquisition. COLING/ACL 2006, pp. 353 - 360, 2006.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. PLSI Utilization for Automatic Thesaurus Construction. Proc. of IJCNLP 2005, pp. 334 - 345, 2005.

PRESS

- The best time of day to learn a new language, according to Duolingo data (Feb. 2018, Quartz)
- 3 habits of successful language learners (Mar. 2017, TechCrunch)

LANGUAGES

Japanese (native), English (fluent), Mandarin Chinese (fluent), Korean (intermediate)