Masato Hagiwara

CONTACT INFORMATION	New York, NY, 11201 USA *E-mail:* hagisan@gmail.com *Web*: http://masatohagiwara.net/

WORK EXPERIENCE

Oct. 2010 – Present: **Lead Scientist – Rakuten Institute of Technology** (in New York / USA)

    * Working on machine transliteration (NLP2011 paper award) and machine translation for e-commerce

    * Chinese/Japanese word segmentation, morphological analysis, named entity extraction

     (Released Rakuten MA https://github.com/rakuten-nlp/rakutenma - word segmentor + PoS Tagger for Chinese and Japanese written purely in JavaScript.)

    * Lexical knowledge acquisition and information extraction from the Web, such as set expansion

    * Writing support system for English as a Second Language (ESL) learners

Apr. 2009 – Sep. 2010: **Research and Development Engineer - Baidu Japan, Inc.** (in China / Tokyo)

    * Planned and acted as a lead developer in projects including Unnatural language processing contest, Baidu Mobile Corpus and Timed Corpus.

    * Worked on the ranking and page analytical algorithms including spam detection for Baidu mobile search. Also worked on the mobile emoticon search using various NLP semantic analysis techniques.

    * Also worked on various NLP topics including – word / sentence analysis technologies, synonym mining and dictionary construction, proper noun detection, Japanese Input Method BaiduType, etc.

Apr. 2008 – Jul. 2008:   **Research Intern - Microsoft Research, WA, USA.** (Mentor: Hisami Suzuki)

    * Proposed a state-of-the-art method for Japanese query alteration, which corrects misspellings and normalizes spelling/transliteration variants, with higher accuracy than conventional systems.

    * Implemented the system using Visual C#, SQL Server, and Ruby, with tens of gigabytes of query log. This system is being integrated into Microsoft Live Search (http://www.live.com/).

    * Developed a method to automatically and efficiently generate query re-writing pairs from session log.

    * Presented the project at the 3rd NLP Symposium for Young Researchers and was awarded the outstanding presentation award. International conference papers are being submitted as well.

Nov. 2006 – Aug. 2007:   **Developer – Information Technology Promotion Agency (IPA), JAPAN: Exploratory Software Project**. (Project Manager: David J. Farber)

    * Accepted as the Exploratory Software Project "Serendi: A Location-Aware Social Networking Platform" (http://serendi.org/), a location-aware meta social networking service targeted at mobile devices with GPS.

    * Developed the "compatibility" analysis module, which recommends users in real time based on natural language processing and network analysis. Used PHP, JavaScript, Ruby, MySQL, and ActiveRecord.

    * Conducted an extensive user test with more than 50 users and confirmed the reliability of the system.

Apr. 2006 – March. 2007:   **Research Assistant -- Nagoya University**

    * Worked on some research projects related to the 21st Century COE Program "Intelligent Media Integration for Social Information Infrastructure" at Nagoya University.

    * Developed improved methods for lexical similarity calculation and thesaurus construction

Sep. 2005 – Mar. 2006, Sep. 2006 – Mar. 2007:   **Teaching Assistant -- Nagoya University**

    * Taught "Linear Algebra" and "Automata and Formal Language Theory" to undergraduate students.

Aug. 2005 – Sep. 2005:   **Intern (Software Engineer) -- Google Inc., CA, USA.**

    (Mentors: Dekang Lin and Jun Wu)

    * Worked on 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.

    * Fully used the parallel distributed computation algorithms such as MapReduce and the large network cluster infrastructure which Google offers.

    * Participated in the two-month internship program, as one of the few interns chosen from Japan, as it was only the second year since the internship was started.

OTHER PROJECTS, ACTIVITIES	**NLTK Japanese Corpora**
	* Introduction and corpus readers for freely available Japanese corpora for NLTK
	Rakuten MA https://github.com/rakuten-nlp/rakutenma
	* Word segmentor + PoS Tagger for Chinese and Japanese written purely in JavaScript.

EDUCATION	May. 2014: **Artificial Intelligence** (CS188x; provided by Pieter Abbeel (UC Berkeley) on edX) Completed all the homework and assignments.
	Dec. 2012: **Probabilistic Graphical Model** (provided by Daphne Koller (Stanford University) on Coursera) Score: 92.8% (with distinction)
	Apr. 2006 – Mar. 2009: **Ph.D., Department of Information Engineering,** Graduate School of Information Science, Nagoya University Doctoral Thesis: "Modeling and Selection of Context for Better Synonym Acquisition"
	Apr. 2004 - Mar. 2006: **Master's Program in Department of Information Engineering,** Graduate School of Information Science, Nagoya University Overall GPA: 3.8
	* Skipped a year in undergraduate and admitted to the graduate school based on the grade-skipping system due to the excellent academic performance. Graduate project: "Automatic Construction of Multilingual Thesaurus for Cross Lingual Information Retrieval"
	Master's Thesis: "Utilization of Probabilistic Latent Semantics for Automatic Thesaurus Construction"
	Apr. 2001 - Mar. 2004: **Information Engineering Course, School of Engineering,** Nagoya University, Japan. Computer Science GPA: 3.9

RESEARCH INTERESTS	1. **Japanese and Chinese word segmentation, PoS tagging, morphological analysis**
	2. **Machine transliteration based on latent classes, machine translation**
	3. **"Un" natural Language Processing**
	4. **Lexical Knowledge Acquisition using Machine Learning and Graph-Theoretic Approaches**

COMPUTER SKILLS	Languages (in the order of fluency): Clojure, Python, JavaScript. Java, C/C++
	Applications: Solr, MongoDB, MySQL, NLTK, Node.js
	5+ years of Web application development experience, including LAMP architecture

NATURAL LANGUAGE SKILLS	Japanese: Native
	English: Fluent - TOEIC score 990 (full score, 2012)
	Chinese (Mandarin) : Advanced – New HSK (汉语水平考试) Grade 6 (Dec. 2010), HSK Speaking (汉语水平考试口试) Advanced (高级) (Mar. 2013)

PUBLICATIONS (SELECTED)

Books and Articles

D. Conway, J. M. White (author), M. Hagiwara, Y. Okuno, T. Mizuno, T. Kinoshita (translation). 入門 機械学習 (Machine Learning for Hackers). O'Reilly Japan, 2012

S. Bird, E. Klein, E. Loper (author) M. Hagiwara, T. Nakayama, T. Mizuno (translation). 入門 自然言語処理 (Natural Language Processing with Python). O'Reilly Japan, 2010

M. Hagiwara, K. Murakami, G. Neubig, Y. Matsubayashi: Robust NLP for Real-world Data : 7. ANPI_NLP - Mining Safety Information after Disasters Using Natural Language Processing -. *Information Processing Society of Japan Magazine*. Vol. 53, No. 3, pp. 241-248, 2012.

M. Hagiwara: Recommendation for Overseas Internship. JSAI Journal, Vol. 29, No. 2, pp. 209-211, 2014

Journal Papers

M. Hagiwara, Y. Ogawa, K. Toyama. Bootstrapping Lexical Knowledge from Unsegmented Text using Graph Kernels. *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 26, No. 3, pp. 440-450. 2011

M. Hagiwara, Y. Ogawa, K. Toyama. Supervised Synonym Acquisition Using Distributional Features and Syntactic Patterns. *Journal of Natural Language Processing*, Vol. 16, Num. 2, pp. 59-83, 2009.

M. Hagiwara, Y. Ogawa, K. Toyama. A Comparative Study on Effective Context Selection for Distributional Similarity. *Journal of Natural Language Processing*, Vol. 5, Num. 5, pp. 119-150. 2008.

M. Hagiwara, Y. Ogawa, K. Toyama. Effective Use of Indirect Dependency for Distributional Similarity. *Journal of Natural Language Processing*, Vol. 15, Num. 4, pp. 19-42, 2008.

M. Hagiwara, Y. Ogawa, K. 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

M. Hagiwara, S. Sekine. Lightweight Client-Side Chinese/Japanese Morphological Analyzer Based on Online Learning. COLING 2014 system demonstration, pp. 39-43, 2014.

H. Li, M. Hagiwara, Q. Li, H. Ji. Comparison of the Impact of Word Segmentation on Name Tagging for Chinese and Japanese, LREC 2014, pp.2532-2536, 2014.

M. Hagiwara, S. Sekine. Accurate Word Segmentation using Transliteration and Language Model Projection, ACL 2013, pp. 183-189.

M. Hagiwara, S. Masuko. KooSHO: Japanese Text Input Environment based on Aerial Hand Writing. *The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL/HLT 2013),* demo session, pp. 24-27, 2013.

Y. Hayashibe, M. Hagiwara, S. Sekine. phloat: Integrated Writing Environment for ESL learners, Second Workshop on Advances in Text Input Methods (WTIM 2012), pp.57-72, 2012.

M. Hagiwara, S. Sekine. Latent Semantic Transliteration using Dirichlet Mixture. *NEWS 2012 (the 4th Named Entities Workshop),* pp. 30-37, 2012.

G. Neubig, Y. Matsubayashi, M. Hagiwara, K. Murakami. Safety Information Mining - What can NLP do in a disaster -, *IJCNLP 2011*.

M. Hagiwara and S. Sekine. Latent Class Transliteration based on Source Language Origins. *ACL-HLT 2011*.

M. Hagiwara and H. Suzuki. Japanese Query Alteration Based on Lexical Semantic Similarity. *NAACL HLT 2009*, pp. 191-199, 2009.

N. Shimizu, M. Hagiwara, Y. Ogawa, K. Toyama, H. Nakagawa. Metric Learning for Synonym Acquisition, *Proc. of COLING 2008*, pp. 793-800, 2008.

M. Hagiwara. A Supervised Learning Approach to Automatic Synonym Identification based on Distributional Features. *Proc. of ACL 2008 Student Research Workshop*, pp. 1-6, 2008.

M. Hagiwara, Y. Ogawa, K. Toyama. Bootstrapping-based Extraction of Dictionary Terms from Unsegmented Legal Text. *Proc. of JURISIN 2008,* pp. 63-72, 2008.

M. Hagiwara, Y. Ogawa, K. Toyama. Context Feature Selection for Distributional Similarity. *IJCNLP 2008*, pp. 553-560, 2008.

M. Hagiwara, Y. Ogawa, K. Toyama. Effective Proximity Distance for Word-Based Context. *Proc. of SNLP 2007*, pp. 105-110, 2007.

M. Hagiwara, Y. Ogawa, K. Toyama. Effectiveness of Indirect Dependency for Automatic Synonym Acquisition. *Proc. of CoSMo 2007*, pp. 1 - 8, 2007.

M. Hagiwara, Y. Ogawa, K. Toyama. Selection of Effective Contextual Information for Automatic Synonym Acquisition. *Proc. of COLING/ACL 2006*, pp. 353 - 360, 2006.

M. Hagiwara, Y. Ogawa, K. Toyama. PLSI Utilization for Automatic Thesaurus Construction. *IJCNLP 2005*, pp. 334 - 345, 2005.

AWARDS & PROFESSIONAL ACTIVITIES

* 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"

* Invited presentation at IPSJ 2012 "Real-world Natural Language Processing"

* Leading editorial member of a special issue on "UnNatural Language Processing, " J. of Nat. Lang. Proc., 2011.

* Panelist at the joint workshop "Relationship between industrial, students, universities, and students in the NLP field" at the 17th Annual Meeting of the Association for Natural Language Processing

* Best Paper Award at NLP2009 "Semantic Category Extraction from Unsegmented Text using Graph Kernels"

* Paper Award at the 3rd NLP Symposium for Young Researchers. Presentation: "A Unified Approach to Japanese Query Alteration based on Semantic Similarity"

* Paper Award at the 22nd IMI Seminar of the 21st Century COE Program. Presentation: "Utilization of Probabilistic Latent Semantics for Automatic Thesaurus Construction"

* Program Committee at SANCL 2012, the Student Research Workshop (SRW) at ACL-IJCNLP 2009 and ACL 2012.

* Program Committee at ACL 2014 (morphology) and COLING 2014 (machine translation).