

**CONTACT** 3563 79<sup>th</sup> st, Apt. #2D  
**INFORMATION** Jackson Heights, NY, 11372 USA

*Phone:* +1-917-502-2779  
*E-mail:* hagsan@gmail.com

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

- EXPERIENCE**
- \* Working on machine transliteration based on latent semantic classes (NLP2011 paper award)
  - \* Working on domain adaptation for Japanese morphological analysis in electronic commerce domain
  - \* Lexical knowledge acquisition from Web, such as set expansion

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.

<b>OTHER PROJECTS, ACTIVITIES</b>	<p><b>NLTK Japanese Corpora</b></p> <ul style="list-style-type: none"> <li>* introductions and corpus readers for freely available Japanese corpora for NLTK</li> </ul> <p><b>frippa (<a href="http://www.frippa.com/">http://www.frippa.com/</a>)</b></p> <ul style="list-style-type: none"> <li>* Developed the entire system of community-based classified ads service, one of the most active peer-to-peer trading communities in Japan with more than 2,000 users.</li> <li>* Runs on an original MVC framework based on Linux, MySQL, ActiveRecord, Ruby, etc.</li> </ul> <p>Also worked on user interface utilizing Ajax and Flash, as a temporary developer at a few IT start-up companies including RINEN.inc (<a href="http://rinen.cc/">http://rinen.cc/</a>) and Anchor (<a href="http://anchor.vc/">http://anchor.vc/</a>)</p> <ul style="list-style-type: none"> <li>* Worked as a technical translator (three years) between English and Japanese and Chinese-Japanese.</li> <li>* Lead language study meet-up groups in NewYork: “Chinese, Japanese, Korean in Practice” <a href="http://bit.ly/JHVOrd">http://bit.ly/JHVOrd</a> and “Business Chinese” <a href="http://bit.ly/rvWRbF">http://bit.ly/rvWRbF</a></li> </ul>
<b>EDUCATION</b>	<p>Apr. 2006 – Mar. 2009: <b>Ph.D., Department of Information Engineering,</b> Graduate School of Information Science, Nagoya University Doctoral Thesis: “Modeling and Selection of Context for Better Synonym Acquisition”</p> <p>Apr. 2004 - Mar. 2006: <b>Master's Program in Department of Information Engineering,</b> Graduate School of Information Science, Nagoya University Overall GPA: 3.8</p> <ul style="list-style-type: none"> <li>* Skipped a year in undergraduate and admitted to the graduate school based on the grade-skipping system due to the excellent academic performance (the very first case in the current department). Graduate project: “Automatic Construction of Multilingual Thesaurus for Cross Lingual Information Retrieval”</li> </ul> <p style="padding-left: 40px;">Master’s Thesis: “Utilization of Probabilistic Latent Semantics for Automatic Thesaurus Construction”</p> <p>Apr. 2001 - Mar. 2004: <b>Information Engineering Course, School of Engineering,</b> Nagoya University, Japan. Computer Science GPA: 3.9</p>
<b>RESEARCH EXPERIENCE</b>	<p><b>1. “Un” natural Language Processing</b></p> <p>UnNatural Language Processing (UNLP) is one research field of NLP, which deals with “real” and “noisy” language data which cannot be captured by conventional “text-book” NLP techniques. Targets of UNLP include, but not limited to: twitter, emoticons, misspellings, irregular NERs, unknown words, informal languages, and so on. The projects I’ve worked on so far are:</p> <ul style="list-style-type: none"> <li>– Emoticon processing for mobile search engines</li> <li>– The First Unnatural Language Processing Contest hosted by Baidu Japan</li> <li>– The second Unnatural Language Processing Thematic Session at NLP2011</li> <li>– ANPI_NLP (Safety Information Mining Project for 2011 Tohoku Region Pacific Coast Earthquake in Japan)</li> </ul> <p><b>2. Transliteration and Japanese Query Alteration</b></p> <ul style="list-style-type: none"> <li>– focusing on multi-lingual latent semantic transliteration models and query alteration</li> </ul> <p><b>3. Lexical Knowledge Acquisition using Machine Learning and Graph-Theoretic Approaches</b></p> <ul style="list-style-type: none"> <li>– worked on the use of latent semantic models in acquiring lexical knowledge from large corpora. Recently focusing on the use of graph-kernels for knowledge extraction from unsegmented Japanese text</li> </ul>
<b>COMPUTER SKILLS</b>	<p>Languages (in the order of fluency): Clojure, Python, Java, C, C++, JavaScript. Ruby, C#</p> <p>Applications: Solr, MongoDB, MySQL, NLTK, Platforms: Windows, Linux</p> <p>5+ years of Web application development experience, including LAMP architecture</p>
<b>NATURAL LANGUAGE SKILLS</b>	<p>Japanese: Native</p> <p>English: Fluent - TOEIC score 960 (2007)</p> <p>Chinese (Mandarin) : Advanced – New HSK (汉语水平考试) Grade 6 (Dec. 2010)</p>

## PUBLICATIONS (SELECTED)

### Books and Articles

Steven Bird, Ewan Klein, Edward Loper (author) Masato Hagiwara, Takahiro Nakayama, Takaaki Mizuno (translation). 入門 自然言語処理 (Natural Language Processing with Python). O'Reilly Japan, 2010

Masato Hagiwara, Koji Murakami, Graham Neubig, Yuichiroh 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.

### Journal Papers

Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko 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

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, Satoshi Sekine. Latent Semantic Transliteration using Dirichlet Mixture. *NEWS 2012 (the 4<sup>th</sup> Named Entities Workshop)*, 2012 (to appear).

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. *ACL-HLT 2011*.

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

Nobuyuki Shimizu, Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama, 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. Bootstrapping-based Extraction of Dictionary Terms from Unsegmented Legal Text. *Proc. of JURISIN 2008*, pp. 63-72, 2008.

Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Context Feature Selection for Distributional Similarity. *IJCNLP 2008*, pp. 553-560, 2008.

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

Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. PLSI Utilization for Automatic Thesaurus Construction. *IJCNLP 2005*, pp. 334 - 345, 2005.

### AWARDS & PROFESSIONAL ACTIVITIES

- \* 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,” *Journal of Natural Language Processing*, 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 of SANCL 2012, the Student Research Workshop (SRW) at ACL-IJCNLP 2009 and ACL 2012.