

CONTACT INFORMATION Hasegawa Corpo 202
1-30-16 Higashiyama, Meguro-ku,
Tokyo, 153-0043 JAPAN

Phone: +81-90-6806-3457
E-mail: hagisan@gmail.com

WORK EXPERIENCE 4/2009 – Present: **Research and Development Engineer - Baidu Japan, Inc.** (in Shanghai / Tokyo)

- * Worked on the ranking and some analytical algorithms for Baidu mobile search (<http://m.baidu.com>). Also implemented machine-learning based algorithms to the spam detection module.
- * Completed the mobile emoticon search using various NLP semantic analysis techniques. This result is presented at the NLP2010 conference.
- * Technical supervision of BaiduType (Japanese input method - <http://type.baidu.jp/>)
- * Also worked on various NLP topics including – word / sentence analysis technologies, synonym mining and dictionary construction, proper noun detection, etc.

4/2008 – 7/2008: **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 the 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.

11/2006 – 8/2007: **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.

4/2006 – 3/2007: **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.
- * Proposed and implemented some extension and selection methods of context for lexical similarity computation, to increase the performance of linguistic resources construction such as thesauri.
- * Published several papers at the top-tier international conferences as well as in journals. (see the "Publications" section)

9/2005 – 3/2006, 9/2006 – 3/2007: **Teaching Assistant -- Nagoya University**

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

8/2005 – 9/2005: **Intern (Software Engineer) -- Google Inc., CA, USA.**

(Mentors: Dekang Lin and Jun Wu)

- * 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.
- * 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.

**OTHER
PROJECTS**

frippa (<http://www.frippa.com/>)

- * Developed the entire system of this community-based classified ads service, one of the most active peer-to-peer trading communities in Japan with more than 2,000 users.
- * Runs on an original MVC framework based on Linux, MySQL, ActiveRecord, Ruby, etc.
- * Implemented a functionality to provide users with related items using natural language processing.
- * Provided the item database in the joint project with the Reuse Market for furniture and appliances at Nagoya University in 2007, as a social contribution activity.

Also worked on user interface utilizing Ajax and Flash, as a temporary developer at a few IT start-up companies including RINEN.inc (<http://rinen.cc/>) and Anchor (<http://anchor.vc/>)

EDUCATION

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
* Entered using the grade-skipping system.
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
EXPERIENCE**

1. Lexical Knowledge Acquisition using Machine Learning and Graph-Theoretic Approaches

Acquiring lexical knowledge such as synonyms, hypernyms, and hyponyms from large corpora is an important issue in NLP. Many of the acquisition methods are based on the *distributional hypothesis*, where the word relatedness is measured by the commonality of the contexts of words. I've been working on the extraction, selection, and extension of useful contextual information for distributional similarity, and proposed several extension and selection methods to enhance the acquisition performance. I also applied the probabilistic latent semantic models (PLSA) in order to solve the sparseness problem and to improve the performance.

I have also been working on the lexical knowledge acquisition using syntactic patterns and machine learning techniques. The recent results show that it achieves more than 50% performance improvement compared to the conventional methods using the vector space model and the distributional similarity. Besides, we can acquire semantic categories (sets of proper nouns belonging to the same class) from unsegmented Japanese corpora using graph-based semantic kernels.

2. Japanese Query Alteration

Query correction is an important problem for the robust Web information retrieval. I proposed and implemented a unified approach for Japanese query alteration at Microsoft Research. The points are (1) it can handle re-writing of candidates that are semantically similar but distinct in spelling, and (2) it uses graph kernel-based semantic similarity to avoid the data sparseness problem.

I also proposed an efficient method for generating query-candidate re-writing pairs from search session log. It showed a better re-writing performance compared to conventional English query correction methods.

**COMPUTER
SKILLS**

Languages: C, C++, C#, Python, Ruby, (D)HTML, JavaScript
Platforms: Windows, Linux
5+ years of Web application development experience, including LAMP architecture

**NATURAL
LANGUAGE
SKILLS**

Japanese: Native
English: Advanced - TOEIC score 960 (2007)
Chinese (Mandarin): Chinese Proficiency Test Grade 3

**PUBLICATIONS
(SELECTED)**

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 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, 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. *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. *Proc. of 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.

**AWARDS &
PROFESSIONAL
ACTIVITIES**

* Outstanding Presentation Award at the Annual Meeting of the Association for Natural Language Processing. Presentation: "Semantic Category Extraction from Unsegmented Text using Graph Kernels"

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

* Outstanding Presentation 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 the Student Research Workshop (SRW) at ACL-IJCNLP 2009 (Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing).

**OTHER
ACHIEVEMENTS**

* Skipped a year in undergraduate and admitted to the graduate school using the grade-skipping system due to the excellent academic performance. Became the very first case of the grade-skipping in the current department. Worked on graduate project in the third grade (title: "Automatic Construction of Multilingual Thesaurus for Cross Lingual Information Retrieval")

* Planned and lead an English study group, which was held weekly to practice presentation. (one year)

* Worked as a technical translator (three years) between English and Japanese in the biomedical, dentistry, and patent translation field.

INTERESTS

Composing playing pieces of music on computers and playing musical instruments in bands.
Reading and learning (natural/computer) languages.

