

連絡先

3563 79th st. Apt. #2D
Jackson Heights, NY, 11372 USA

携帯: +1-917-502-2779
E-mail: hagnosis@gmail.com

職歴

2010年10月 - (現在): シニア・サイエンティスト - 楽天技術研究所 (米国ニューヨーク)

- * 潜在クラスに基づく翻字モデルの研究に従事しています。(NLP2011にて優秀発表賞を受賞)
- * 特にEC(電子商取引)分野における, 日本語形態素解析の分野適応の研究をしています.
- * その他, 集合拡張など Web からの語彙知識獲得等のタスクに取り組んでいます.

2009年4月-2010年9月: RD(エンジニア)- バイドゥ株式会社 (東京 / 上海オフィス勤務)

- * 「不自然言語処理コンテスト」, バイドゥモバイル Web コーパス, 時間軸コーパスなどの企画・プロジェクトを中心となって企画, 開発しました.
- * バイドゥモバイル検索のランキング・アルゴリズムの向上, ページ解析技術, スпам・フィルタの開発を担当しました. また, 絵文字検索および自然言語処理技術を用いた絵文字意味解析を開発しました.
- * 日本語形態素, 同義語マイニング, 固有名詞資源整備, BaiduType (日本語仮名漢字変換) など, その他様々な自然言語処理トピックに取り組み, 検索エンジンの質の向上に貢献しました.

2008年4-7月: リサーチ・インターン - Microsoft Research (米国ワシントン州) (メンター: 鈴木久美氏)

- * 検索クエリの誤り訂正・異表記や翻字の正規化を行う高性能な日本語のクエリ書き換え手法を提案し, 従来手法に対する有意性を確認しました.
- * 数10GBの検索ログを扱うクエリ書き換えシステムを Visual C#, SQL Server, Ruby を用いて実装しました. 本システムは, Microsoft の Live Search (<http://www.live.com/>) に近々組み込まれる予定です.
- * クエリ書き換えペアをセッションログから効率的かつ自動的に生成する手法を開発しました.
- * 本プロジェクトの結果を第3回 NLP 若手の会にて発表し, 奨励賞を受賞しました. さらに, 国際会議 NAACL HLT 2009 に論文が採択されました.

2006年11月 - 2007年8月: 開発エンジニア - 情報処理推進機構 (IPA) 2006年度下期未踏ソフトウェア創造事業 (PM: David J. Farber 教授)

- * 未踏ソフトウェア創造事業に, GPS 付き情報端末を対象とした位置情報適応型メタ SNS "Serendi: A Location-Aware Social Networking Platform" (<http://serendi.org/>) が採択されました.
- * 自然言語処理およびネットワーク解析技術を用いて, ユーザー間の「Compatibility (相性)」を解析するモジュールを開発しました. PHP, JavaScript, Ruby, MySQL, ActiveRecord を使用しました.
- * 同時参加人数 50人以上の大規模なユーザーテストを実施し, 本システムの信頼性を確認しました.

2006年4月 - 2007年3月: リサーチ・アシスタント - 名古屋大学

- * 名古屋大学において採択された21世紀COEプログラム「社会情報基盤のための音声・映像の知的統合」に関連するプロジェクトの研究開発に従事しました.
- * シソーラスをはじめとする言語資源構築の性能改善のため, 語彙類似度計算に用いる文脈の拡張・選択手法をいくつか提案・実装しました.
- * ここでの研究成果は, トップレベルの国際会議および学術雑誌に数多く採録されました (主要論文参照).

2005年9月 - 2006年3月, 2006年9月 - 2007年3月: ティーチング・アシスタント - 名古屋大学

- * 「線形代数および演習」と「線形代数および演習」の講義を担当しました.

2005年8-9月: ソフトウェア・エンジニア・インターン - Google Inc. (米国カリフォルニア州)

(メンター: Dekang Lin 氏・Jun Wu 氏)

- * Google 本社における2ヶ月間のインターンシップに参加しました. 当時インターン制度が始まって1年しか経過しておらず, 数少ない日本人インターンとして選ばれたうちの一人です.
- * 日本語クエリ・サジェスションのプロジェクトに取り組みました. この時の成果は, 現在の Google 検索結果の画面上下部に表示されるクエリ・サジェスションの基礎技術として採用されています.
- * MapReduce など Google 社の保有する並列計算アルゴリズムおよび大規模クラスターなどのインフラを駆使して開発を行いました.

その他
開発経験
特筆事項

NLKT 日本語コーパス

* NLTK (Natural Language Toolkit)のためのフリー日本語コーパス用ライブラリの開発をしました。

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

* 地域型個人間取引仲介・案内・3行広告 Web サービス frippa のシステム全体の開発を担当しました。現在約 2,000 人のユーザーが参加しており、活発な個人間取引サイトへと成長しました。

* Linux, MySQL, ActiveRecord, Ruby 等をベースにした独自 MVC フレームワーク上で動作しています。

その他に、RINEN.inc (<http://rinen.cc/>), Anchor (<http://anchor.vc/>) など IT 系ベンチャー企業においてアルバイトとして開発経験があり、Ajax や Flash 等を駆使したユーザーインターフェースを担当しました。

* 日本語・英語間、および、日本語・中国語間の技術翻訳実務に従事しました (3年間)

* ニューヨークにて、語学学習グループを主催しています：“Chinese, Japanese, Korean in Practice”
<http://bit.ly/JHVOrd> “Business Chinese” <http://bit.ly/rvWRbF>

学歴

2006 年 4 月 – 2009 年 3 月: 名古屋大学 大学院 情報科学研究科 博士課程後期課程

情報システム学専攻ソフトウェア論講座

博士論文: 「Modeling and Selection of Context for Better Synonym Acquisition」

(日本語訳: 類義語獲得の性能向上のための文脈のモデル化と選択)

2004 年 4 月 – 2006 年 3 月: 名古屋大学 大学院 情報科学研究科 博士課程前期課程

情報システム学専攻 ソフトウェア論講座 GPA: 3.8

* 成績優秀のため飛び級入学し、当該研究科において初の飛び級事例となりました。学部 3 年次に卒業論文相当の研究に取り組みました (題目: 「クロスリンガル情報検索のための多言語シソーラス自動構築に関する研究」)

修士 (情報科学) 修士論文: 「Utilization of Probabilistic Latent Semantics for Automatic Thesaurus Construction」 (日本語訳: シソーラス自動構築のための確率的潜在意味の利用)

2001 年 4 月 – 2004 年 3 月: 名古屋大学 工学部 電気電子・情報工学科

専門課程 GPA: 3.9

2001 年 3 月: 愛知県立国府高等学校卒業

研究経験

1. 「不自然言語処理」

「不自然言語処理」とは、従来の「教科書的な」自然言語処理ではとらえることのできない、実世界のリアルな言語現象に対する技術の総称を指します。具体的には、twitter などのソーシャルメディア、絵文字、スペルミス、変則的な固有名詞、くだけた表現などが対象として挙げられます。この枠組の中で、私は以下のような活動に取り組んできました:

- モバイル検索エンジンのための絵文字処理
- バイドゥ株式会社主催の第一回不自然言語処理コンテスト
- NLP2011 におけるテーマセッション「不自然言語処理～枠に収まらない言語表現の処理」
- ANPL_NLP-2011 年東日本大震災のための被災者安否情報マイニングプロジェクト

2. 翻字・日本語クエリ書き換え

- 原言語の起源と潜在クラスを用いた翻字モデル、および、日本語のクエリ書き換えに取り組みました。

3. 機械学習・グラフ理論に基づく語彙知識獲得

- 潜在意味モデルに基づく大規模コーパスからの語彙知識の獲得、および、グラフカーネルに基づく日本語非分かち書き文からの意味カテゴリ獲得に取り組みました。

技術スキル

言語 (使用頻度順): Clojure, Python, Java, C, C++, JavaScript, Ruby, (D)HTML

アプリケーション: Solr, MongoDB, MySQL, NLTK, プラットフォーム: Windows, Linux

LAMP アーキテクチャに基づいた Web アプリケーション経験 (5年以上)

語学スキル

日本語: ネイティブ

英語: 会話・読み書き上級 TOEIC 960 点 (2007 年取得)

中国語 (標準語): 会話・読み書き上級 (2010 年 新 HSK 6 級取得)

主要論文・発表

著書・訳書・記事

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

萩原正人, 村上浩司, Graham Neubig, 松林優一郎: 不自然言語処理 - 枠に収まらない「リアルな」言語処理 - : 7.ANPI_NLP-NLP 技術を応用した震災時安否情報確認支援-. 情報処理. Vol. 53, No. 3, pp. 241-248, 2011.

論文誌 (査読あり)

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.

国際会議・ワークショップ (査読あり)

Masato Hagiwara, Satoshi Sekine. Latent Semantic Transliteration using Dirichlet Mixture. *Proc. of NEWS 2012 (the 4th 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. *Proc. of ACL-HLT 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, 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. 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.

受賞・研究貢献

* 言語処理学会第 17 回年次大会(NLP2011) 優秀発表賞「原言語の起源に基づく潜在クラス翻字モデル」

* IPSJ 2012 「実世界自然言語処理への挑戦」パネリストおよび講演

* 論文誌「自然言語処理」特集号「不自然言語処理」編集委員長

* NLP2011 併設ワークショップ「自然言語処理における企業と大学と学生の関係」パネリスト

* 言語処理学会第 15 回年次大会(NLP2009) 最優秀発表賞「グラフカーネルに基づく非分かち書き文からの意味的語彙カテゴリの抽出」

* NLP 若手の会 第 3 回シンポジウム 奨励賞受賞

発表: 「意味的類似度を利用した日本語クエリ書き換えのための統一的アプローチ」

* 21 世紀 COE プログラム「社会情報基盤のための音声・映像の知的統合」第 22 回 IMI セミナー 優秀賞受賞 発表: 「シソーラス自動構築における確率的潜在意味の利用」

* ACL-IJCNLP 2009 および ACL 2012 Student Research Workshop, SANCL 2012 プログラム委員