

Richard C. Wang

ADDRESS: 4625 5th Ave. Apt. 405
Pittsburgh, PA 15213

CELL: (412) 448-5959
E-MAIL: rcwang@cmu.edu

OBJECTIVE: Seeking a full-time job in information extraction/retrieval and natural language processing.

EDUCATION: **Carnegie Mellon University**, Language Technologies Institute, Pittsburgh, PA
Candidate for Ph.D. in Language Technologies, Aug. 2009.

Carnegie Mellon University, Language Technologies Institute, Pittsburgh, PA
Masters in Language Technologies, May 2005.
Cumulative QPA: 3.85

Carnegie Mellon University, Computer Science Department, Pittsburgh, PA
Bachelor of Science, May 2003.
Major: **Computer Science**. Minor: **Business Administration**

WORK & RESEARCH EXPERIENCE: **Research / Teaching Assistant** July 2003 ~ Present

Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA

- Conducted research related to information extraction and the CMU RADAR project.
- Responsible for grading assignments & exams, making exam questions, and holding office hours for the courses: 1) Information Retrieval and 2) Advanced Natural Language Processing Seminar.

Google Summer Engineering Intern May 2006 ~ Aug. 2006

Google Inc., Mountain View, CA, USA

- Developed a system that learns from annotated text and automatically produces regular expressions for extracting information.

CMU School of Computer Science Web Programmer Aug. 2002 ~ May 2003

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA

- Designed and programmed the official School of Computer Science website.

Merck Computational Biology Summer Scholar June 2002 ~ Aug. 2002

Biological Sciences Department, Carnegie Mellon University, Pittsburgh, PA, USA

- Developed a system that extracts information from image captions in biological journal articles.

System Administrator and Technical Writer May 2000 ~ Aug. 2000

Edimax Technology Co., Ltd., Taipei County, Taiwan

- Installed and maintained intranet systems of Novell/Windows servers, hubs, and switches on LAN.
- Translated real-time conversations, emails, web pages, user manuals between English and Chinese.

RESEARCH PROJECTS: **www.BooWa.com** – A list extractor for many languages

- Automatically finds all hyponyms of a given noun (e.g. “car makers” → Ford, Nissan, Toyota, etc.)
- Automatically expands a given set of entities into a more complete set (like Google Sets).

Personal Names Extractor for E-mails

- Recognizes personal names in e-mails using machine-trained sequential learners.

Information Extractor for Space-Request E-mails

- Extracts and processes information from e-mails regarding conference and office space requests.

Automatic Named Entity Translator

- Automatically translates named entities in multiple languages using snippets returned by search engines.

Personalized Meta-Search Engine

- Extracts, filters, and merges search results from multiple search engines.
- Personalizes results by re-ranking snippets based on keywords that occurred in user’s visited snippets.

FIRST AUTHOR PUBLICATIONS: • Character-Level Analysis of Semi-Structured Documents for Set Expansion. EMNLP 2009.

• Automatic Set Instance Extraction using the Web. ACL-IJCNLP 2009

• Iterative Set Expansion of Named Entities using the Web. ICDM 2008

• Automatic Set Expansion for List Question Answering. EMNLP 2008

• Language-Independent Set Expansion of Named Entities using the Web. ICDM 2007

AWARDS: Awarded **Google Research Grant Award** in 2008.

LANGUAGES: Java, JavaScript, HTML, PHP, JSP, C, C++, Perl, SQL, Matlab, Lisp, and SML.
Proficient in English and Chinese (both Mandarin and Taiwanese), Intermediate in Japanese.