

RICHARD WANG

rcwang@gmail.com

+1-412-448-5959

<http://www.rcwang.com>

SUMMARY

My expertise is organizing world's information, using big data, machine learning, natural language processing, information retrieval and extraction. I have many years of hands-on technical lead experience as well as strong research background.

EDUCATION

Carnegie Mellon University 2005 - 2009

Ph.D. , Computer Science, Language Technologies

Carnegie Mellon University 2003 - 2005

M.S. , Computer Science, Language Technologies

Carnegie Mellon University 1999 - 2003

B.S. , Computer Science

EXPERIENCE

Baidu USA 03 / 2014 - Present

Senior Software Architect

I lead a knowledge search team of 6-7 people onsite and 3-4 people remotely. By building a state-of-the-art medical knowledge graph with 200+ million facts, combined with deep learning and NLP, our system is able to understand 90% of medical-related user-search queries with 95% accuracy. When applied to online search ads, we increased CTR of medical-related ads by 0.5% and RPM by 1.5%, which equates to an annual revenue of \$12M USD for the company.

EnFind, Inc. 10 / 2012 - 03 / 2014

Founder / CTO

I founded a company to deep-learn and deep-understand online articles by automatically identifying and distinguishing all people, places, and things mentioned, and to provide most relevant information to the publishers, advertisers, and users. We launched on many publishers' websites in 2013 and reached 1M+ daily users. By using our proprietary technology in NLP, AI, and knowledge graph, we achieved a super high mouse-hover rate of 20% (every 1 out of 5 users would mouse over our annotations) and an astonishing re-hover rate of 80% (every 4 out of 5 users who hovered once would hover another one). Please visit www.enfind.com for examples and demos.

Google 11 / 2009 - 10 / 2012

Software Engineer

Google's Knowledge Graph, used widely in search and ads, contains more than 44 million topics and 2.4 billion facts, and is considered the largest knowledge-base with 99% accuracy. I lead a project that further increased its coverage and completeness by more than 5%, using unsupervised clustering and semi-supervised learning on semi-structured data from the web.

I also worked on the Google Squared project, an automatic instance extraction system for categorical user queries (eg. car makers, US presidents, etc.). I implemented and integrated my thesis system SEAL into Google Squared and improved its overall precision and recall by more than 30%.

Carnegie Mellon University

07 / 2003 - 11 / 2009

Research Assistant on Machine Learning and Information Extraction

I conducted research and implemented systems related to machine learning, natural language processing, and information extraction for the RADAR personal assistant project and my thesis project at CMU.

Carnegie Mellon University

01 / 2009 - 05 / 2009

Teaching Assistant on Natural Language Processing

I was the teaching assistant for the course "Advanced Natural Language Processing Seminar" at CMU.

Carnegie Mellon University

01 / 2007 - 05 / 2007

Teaching Assistant on Information Retrieval

I graded student's assignments and exams, made exam questions, and held office hours for the course "Information Retrieval" at CMU.

Google

05 / 2006 - 09 / 2006

Summer Engineering Intern

I developed a generic learning and extraction system that automatically produces complex and well-formulated regular expressions for extracting information learned from annotated training data.

Carnegie Mellon University

01 / 2003 - 07 / 2003

Machine Translation Research Programmer

I developed a system for the RADD machine translation project at CMU that crawls the web and automatically identifies and extracts parallel corpora from online web pages for training machine translation models.

Carnegie Mellon University

08 / 2001 - 07 / 2003

Writing Assistant for Advanced Chinese Courses

I held office hours and tutored hundreds of students on advanced Chinese composition for the courses "Chinese Language and Culture" and "Topics in Contemporary Culture in China" at CMU.

Carnegie Mellon University

08 / 2002 - 05 / 2003

Carnegie Mellon University Web Programmer

I designed and programmed part of the official School of Computer Science website at <http://www.cs.cmu.edu>.

Carnegie Mellon University

05 / 2002 - 09 / 2002

Merck Computational Biology Summer Scholar

I developed a system that automatically extracts cell/protein names from image captions in biological journal articles, to use as features for training a cell/protein image classifier.

Edimax USA

05 / 2000 - 09 / 2000

System Administrator and Technical Writer Summer Intern

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

SKILLS

Machine Learning	Algorithms	Computer Science
Natural Language Processing	Artificial Intelligence	Information Extraction
Data Mining	Pattern Recognition	Information Retrieval
MapReduce	C++	Java
Scalability	Text Mining	Software Engineering
Computational Linguistics	Big Data	Product Management
Programming	Software Development	Hadoop
PHP	Git	MySQL
SQL	Perl	Databases
JavaScript	Web Development	

LANGUAGES

English (Native or bilingual proficiency) **Chinese** (Native or bilingual proficiency) **Japanese** (Limited working proficiency)

PATENTS

Refining Search Results

Issued , 04 / 2015

US Patent: 9015143 · Inventors: Engin Cinar Sahin, Daniel A. Rocha, Richard Wang · <http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=9015143>

Methods, systems, and apparatus, including computer programs encoded on a computer storage medium, for refining search results. In one aspect, a method includes presenting search results that are responsive to a search query on a search engine results page and each search result includes a snippet. The method also include receiving a refinement to the search query through the search engine results page and generating a request to update one or more of the snippets based on the refinement. The method also includes receiving one or more updated snippets in response to the request and updating one or more of the search results presented on the search engine results page using the one or more updated snippets.

PUBLICATIONS

Distant IE by Bootstrapping Using Lists and Document Structure

02 / 2016

Artificial Intelligence (AAAI 2016) · Authors: Lidong Bing, Mingyang Ling, Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/aaai-2016.pdf>

Improving Distant Supervision for Information Extraction Using Label Propagation Through Lists

09 / 2015

Empirical Methods in Natural Language Processing (EMNLP 2015) · Authors: Lidong Bing, Sneha Chaudhari, Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/emnlp-2015.pdf>

Never-Ending Learning

01 / 2015

Conference on Artificial Intelligence (AAAI 2015) · Authors: Tom M. Mitchell, William Cohen, Estevam Hruschka, Partha Pratim Talukdar, Justin Betteridge, Andrew Carlson, Richard Wang, et al. · <http://www.cs.cmu.edu/~rcwang/papers/aaai-2015.pdf>

A Tale of Two Entity Linking and Discovery Systems

11 / 2014

Knowledge Base Population Text Analysis Conference (KBP-TAC 2014) · Authors: Kathryn Mazaitis, Richard Wang, Frank Lin, Bhavana Dalvi, Jakob Bauer, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/kbp-2014.pdf>

Coupled Semi-Supervised Learning for Information Extraction

2010

Third ACM International Conference on Web Search and Data Mining (WSDM 2010) · Authors: Andrew Carlson, Justin Betteridge, Richard Wang, Estevam Hruschka, Tom M. Mitchell · <http://www.cs.cmu.edu/~rcwang/papers/wsdm-2010.pdf>

Populating the Semantic Web by Macro-Reading Internet Text

2009

8th International Semantic Web Conference (ISWC 2009) · Authors: Tom M. Mitchell, Justin Betteridge, Andrew Carlson, Estevam Hruschka, Richard Wang · <http://www.cs.cmu.edu/~rcwang/papers/iswc-2009.pdf>

Character-Level Analysis of Semi-Structured Documents for Set Expansion

2009

Conference on Empirical Methods in Natural Language Processing (EMNLP 2009) · Authors: Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/emnlp-2009.pdf>

Automatic Set Instance Extraction using the Web

2009

Joint Conference of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009) · Authors: Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/acl-ijcnlp-2009.pdf>

Iterative Set Expansion of Named Entities using the Web

2008

IEEE International Conference on Data Mining (ICDM 2008) · Authors: Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/icdm-2008.pdf>

Automatic Set Expansion for List Question Answering

2008

Conference on Empirical Methods in Natural Language Processing (EMNLP 2008) · Authors: Richard Wang, Nico Schlaefer, William Cohen, Eric Nyberg · <http://www.cs.cmu.edu/~rcwang/papers/emnlp-2008.pdf>

Integrating a Natural Language Message Pre-Processor with UIMA

2008

LREC Workshop - Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP · Authors: Eric Nyberg, Eric Riebling, Richard Wang, Robert Frederking · <http://www.cs.cmu.edu/~rcwang/papers/lrec-2008.pdf>

Learning to Extract Gene-Protein Names from Weakly-Labeled Text

2008

CMU SCS Technical Report Series (CMU-LTI-08-004) · Authors: Richard Wang, Anthony Tomasic, Robert Frederking, Isaac Simmons, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/cmu-lti-08-004.pdf>

Language-Independent Set Expansion of Named Entities using the Web

2007

IEEE International Conference on Data Mining (ICDM 2007) · Authors: Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/icdm-2007.pdf>

NER Systems that Suit Users Preferences: Adjusting the Recall-Precision Trade-off for Entity Extraction

2006

Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL 2006) ·

Authors: Einat Minkov, Richard Wang, Anthony Tomasic, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/naacl-2006.pdf>

Extracting Personal Names from Emails: Applying Named Entity Recognition to Informal Text 2005

Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005) ·

Authors: Einat Minkov, Richard Wang, William Cohen · <http://www.cs.cmu.edu/~rcwang/papers/emnlp-2005.pdf>

Understanding Captions in Biomedical Publications

2003

Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining · Authors: William Cohen, Richard Wang, Robert

Murphy · <http://www.cs.cmu.edu/~rcwang/papers/ismb-2003.pdf>