Raphael Hoffmann, PhD
Mountain View, CA 94043
425-246-5932 | raphael.hoffmann@gmail.com | linkedin.com/in/raphaelhoffmann

ENGINEER & SCIENTIST

Excellent engineer – Created patented technology for HighSpot & leading NLP technology for DARPA Memex. As lead engineer, built distributed systems, ML/NLP algorithms, and web apps.

Innovative – Created state-of-the art NLP models. Published novel research at top-tier conferences, and received more than 1000 citations.

Leadership skills – Co-founded Lattice Data, Inc. (20+ employees, backed by top VCs, Fortune-500 customers).

Technical tools
scala, java, python, javascript/node/react/angular, spark/hadoop/sql, probabilistic graphical models

Experience

LATTICE DATA, INC. – Menlo Park, CA
Co-Founder & Lead Engineer, 01/2015 – 11/2016
Developed the Lattice platform that uses large-scale statistical inference for extraction & linking.

- Developed models for entity linking using large-scale joint inference (factor graphs).
- Developed models for relation extraction using large-scale joint inference (factor graphs).
- Developed framework for expressing, prioritizing, and inspecting distant supervision rules and features, enabling an enormous reduction in model development time for all engineers and scientists at the company (python).
- Developed, tested, evaluated, refined several dozen extractors for various customers, including leading extractors for DARPA Memex (python).
- Developed customer-facing widgets for data annotation (javascript, angular).
- Developed customer-facing application for searching and browsing unstructured and extracted content (coffeescript, angular, node, elasticsearch).
- Developed and maintained real-time data ingestion pipelines (kafka, airflow, redis).
- Developed system that enables customers to subscribe to notifications when their custom queries match extractions in the data processed in real-time (coffeescript, node).
- Designed and developed framework for distributed ingestion pipelines (python, greenplum).
- Developed containers for sandbox development, production systems, and demos (docker).
- Managed Lattice resources on AWS, including all production systems.
- Interviewed almost 100 candidates for engineering positions, and brought several top candidates on board.

HIGHSPOT, INC. – Seattle, WA
Sr. Data Scientist, 07/2014 – 04/2015
Developed patented technology to search for semantically and visually related sales slides.

- Developed and refined features over parsed PDF/PPTX documents and their embedded images (clojure).
- Developed and evaluated LSA, VSM, LSH models for similarity (clojure, spark, solr).
- Optimized code for scalability to productize as part of HighSpot Content Genomics.

ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE – Seattle, WA
Visiting Scientist, 12/2013 – 12/2014
Developed interactive tool for building relation extractors.

- Developed NLP visualization and annotation widgets (javascript, angular).
- Developed interactive bootstrapping tool for relation extraction (scala, RDBMS indexes).
- Developed language for distant supervision based on first-order logic (scala, SQL).

MICROSOFT RESEARCH – Redmond, WA
Research internship, 2008
Developed patented query tagging technique with large-scale lexicon mining from the Web.

- Developed discriminative PCFG model for query tagging (C#).
- Using Cosmos, mined structured lists from Web crawl (Scope, C#).

UNIVERSITY OF WASHINGTON – Seattle, WA
Research assistant, 09/2005 – 12/2012
Pushed state-of-the-art in extraction with distant supervision and mixed-initiative systems, eg.:

- Developed factor graph model for distant supervision of relation extractors which can handle overlapping relations (java).
- Developed system for learning self-supervised extractors from Wikipedia (java).
- Developed system that synergistically combines community content creation and extraction (java, javascript).
- Developed search engine for source code (java, lucene, javascript).
- Conducted three user studies to evaluate different user-facing systems I have built.
- Published research at top-tier conferences, and received more than 1000 citations.

Education

UNIVERSITY OF WASHINGTON – Seattle, WA
PhD in Computer Science, 2012
MS in Computer Science, 2007

UNIVERSITÄT PASSAU – Germany
Diplom in Informatik, 2005

Awards

Best Paper Nomination at CHI 2009
Yahoo Key Scientific Challenges Award
University of Washington College of Engineering Award
Fulbright Graduate School Scholarship
Graduation with Distinction (Diplom)

Patents

Systems and methods for identifying semantically and visually related content, United States US20160217343 A1

Acquisition of semantic class lexicons for query tagging, United States US20100268725 B1

Publications

Raphael Hoffmann, Luke S. Zettlemoyer, Daniel S. Weld. *Extreme Extraction: Only One Hour per Relation.* CoRR abs/1506.06418 (2015)

Wei Xu, **Raphael Hoffmann**, Le Zhao, Ralph Grishman. *Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction*. ACL 2013

Congle Zhang, **Raphael Hoffmann**, Daniel S. Weld. *Ontological Smoothing for Relation Extraction with Minimal Supervision*. AAAI 2012

Raphael Hoffmann. *Interactive Learning of Relation Extractors with Weak Supervision*. PhD Dissertation 2012

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, Daniel S. Weld. *Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations*. ACL 2011

Raphael Hoffmann, Congle Zhang, Daniel S. Weld. *Learning 5000 Relational Extractors*. ACL 2010

Ye-Yi Wang, **Raphael Hoffmann**, Xiao Li, Daniel S. Weld. *Semi-Supervised Learning of Semantic Classes for Query Understanding --- From the Web and For the Web*. CIKM 2009

Daniel S. Weld, **Raphael Hoffmann**, Fei Wu. *Using Wikipedia to Bootstrap Open Information Extraction.* ACM SIGMOD Record 2009

Raphael Hoffmann, Saleema Amershi, Kayur Patel, Fei Wu, James Fogarty, Daniel S. Weld. *Amplifying Community Content Creation with Mixed-Initiative Information Extraction*. CHI 2009

Fei Wu, **Raphael Hoffmann**, Daniel S. Weld. *Information Extraction from Wikipeda: Moving Down the Long Tail*. KDD 2008

Daniel S. Weld, Fei Wu, **Raphael Hoffmann**, Eytan Adar. *Intelligence in Wikipedia*. AAAI 2008

Raphael Hoffmann, Patrick Baudisch, and Daniel S. Weld. *Evaluating Visual Cues for Switching Windows on Large Screens*. CHI 2008

Raphael Hoffmann, James Fogarty, and Daniel S. Weld. *Finding and Leveraging Implicit References in a Web Search Engine for Programmers*. UIST 2007

Krzysztof Gajos, David B. Christianson, **Raphael Hoffmann**, Tal Shaked, Kiera Henning, Jing Jing Long, and Daniel S. Weld. *Fast and Robust Interface Generation for Ubiquitous Applications*. Ubicomp 2005

Raphael Hoffmann. *Model Selection for Support Vector Machines*. Diplom Thesis 2005

Krzysztof Gajos, **Raphael Hoffmann**, and Daniel S. Weld. *Improving User Interface Personalization*. UIST 2004

Markus Ramsauer, Michael Coduro, and **Raphael Hoffmann**. *Tool Supported Development of Energy-aware Real-time Applications for Embedded Systems*. ECRTS 2004

Other

Created one of Europe's largest online education sites with 6M+ users, antolin.de.

References available on request.