

Jonathan Koren

131 Pala Avenue, San Jose, CA 95127
mobile: +1 408 859 0984

jonathan@jonathankoren.com
<http://www.jonathankoren.com>

EXPERIENCE

2017 – Present **Senior Staff Software Engineer** **Mozilla; Mountain View, CA**
Lead the machine learning team for Pocket.

2015 – 2017 **Senior Software Engineer** **Ozlo; Palo Alto, CA**
Principal architect responsible for information extraction, knowledge base, and lexicon. Together, these allow for the discovery of attributes of entities from unstructured natural language, and a more robust understanding of natural language.

Using structured, semistructured, and unstructured data sources, along with actual query logs, worked with the product team to define and scope additional attributes for various entities (places of interest, film and television, music, recipes, etc.) would be supported. These attributes were novel to Ozlo, and not merely passed through from external data providers.

Advocated for, designed and implemented a knowledge representation layer (knowledge base) to allow inference over attributes to deduce second-order facts about entities and to allow intelligent backoff and understanding of user queries. The knowledge base became one of the core technologies around all natural language efforts inside Ozlo.

Advocated and led the construction of a lexicon separate from the knowledge base. The lexicon contained English words that mapped to concepts inside the knowledge base. By explicitly breaking the lexicon and knowledge base apart, recall of concepts improved, and the the query understanding logic became more robust to variation. Additionally, the separating the lexicon out from the knowledge base allowed for multilingual retrieval.

Aligned multiple external corpora to enhance the coverage of the lexicon and to add support for sentiment, word relationships (e.g. synonym, antonym, hypernym, etc.), verb valance, lemmatization / conjugation, and singularization / pluralization. The lexicon became a core component in the natural language understanding and natural language synthesis components of the Ozlo chatbot product.

Created visualization tools for the knowledge base and lexicon.

Championed and developed data driven methods to measure and improve knowledge base and lexicon quality. This was important not only to understand current quality but also to aid in bringing additional data sources online.

Managed interns. Led architecture discussions.

2014 – 2015 **Staff Software Engineer** **Facebook; Menlo Park, CA**
Search NLP group. Worked on key-voice detection in trending news articles. Used n-grams, noun phrases, and named entities along with Wikipedia and other external data sources to identify important entities related to trending news topics, and then identify relevant news feed posts from these entities.

Search Discovery. Trending news relevance optimization. Migrated product from feed ranking infrastructure to a more search oriented infrastructure with negligible relevance impact. Championed and implemented explore-exploit framework to increase relevance and adaptability of the ranking system to changing and rapidly developing trends. This resulted in an increase in CTR of the trending module. Leveraged interest-based features to increase personal relevance of the trending module.

2010 – 2014 **Staff Applied Research Engineer** **LinkedIn; Mountain View, CA**

Launch team for LinkedIn Today (now LinkedIn Pulse) – LinkedIn's news aggregation and recommendation product. Designed and developed real-time indexing and near-real-time retrieval and recommendation platform for news content. Developed data collection strategies, relevance performance metrics, and statistical models that combined both content and social activity features to optimize personalized relevance. Lead the modeling efforts to transition LinkedIn Today, from using a licensed Twitter data feed, to a natively produced data feed with no noticeable impact in result quality or user experience.

Sponsored content optimization. Developed statistical models to optimize CTR and eCPM of sponsored content (i.e. sponsored status updates and sponsored recommendations) appearing in users' update streams, while maintaining content relevance. Engineered model features and maintained real-time inference engine.

Machine learning infrastructure group. Developed large-scale machine learning infrastructure for near-realtime inference, rapid model training, along with automatic model retraining. Utilized a combination of Kafka, Hadoop, and a custom logistic regression inference engine.

Designed, implemented, and managed the transition from a Spring RPC-based to a REST-based API. Scoped, managed, and tested the ads relevance infrastructure rollout to additional data centers, thereby enabling active-active redundancy.

2006 – 2010 **Graduate Student Researcher** **University of California, Santa Cruz**

Advisor: Dr. Yi Zhang

Researched automatic generation of personalized adaptive faceted search interfaces.

Researched petabyte-scale distributed metadata storage systems, with emphasis on enterprise search, including ranking, among heterogeneous documents, context-aware, personalized, and collaborative filtering.

2009 **Intern** **Lawrence Livermore Nat. Lab; Livermore, CA**

Conducted research on ontology automatic construction from unstructured text collections. Ontologies were then used to build faceted search interfaces.

2008 **Intern** **Microsoft Research; Redmond, WA**

Conducted research into static ranking of web pages based on language models of URLs actually browsed to.

2006 **Intern** **Yahoo! Search; Santa Clara, CA**

Conducted research on determining intent from query logs. This research was used to diversify search results.

2003 – 2005 **Research Assistant** **Southern Illinois University, Carbondale**

Advisor: Dr. Norman F. Carver III

Conducted research on multiagent distributed sensor interpretation systems using distributed Bayesian networks and decentralized Markov decision processes. Research focused on problem complexity and domain monotonicity.

2005 – 2006 **Developer** **Th(i)ng Ed; Carbondale, IL
(formerly SchoolCenter)**

Used PHP and SQL in a Linux-Apache (LAMP) environment, to develop a market leading K-12 content management system.

1999 – 2001**Sr. Software Engineer****Motorola; Arlington Heights, IL**

Designed and implemented support for EDGE (a.k.a. EGPRS) hardware in the Configuration Management and Network Management System processes. This involved porting the process from proprietary operating system on a 68x000 based architecture to VxWorks on a PowerPC based architecture.

Developed and maintained real-time (RTOS) software for the Man-Machine Interface and the Configuration Management processes for the GSM Base Station Subsystem using C and C++.

EDUCATION

PhD 2006 – 2010 ON LEAVE**Computer Science****University of California, Santa Cruz**

Advisor: Dr. Yi Zhang

Research Concentration: Faceted Search, Personalization/Recommendation, Adaptive Interfaces, Machine Learning

MS 2005**Computer Science****Southern Illinois University, Carbondale**

Advisor: Dr. Norman F. Carver III

Thesis: Minimizing Communication Cost in an N-Agent Distributed Bayesian Network by Using a Decentralized MDP

BS 1998**Computer Science****Southern Illinois University, Carbondale**

PUBLICATIONS

Blog posts

J. Koren. "Ozlo's Knowledge Base" <https://medium.com/teamozlo/ozlos-knowledge-base-eb74204d7e6c>

Open Source Projects

GOWN. A Golang implementation of WordNet <https://github.com/ozlo/gown>

Conference Proceedings

D. Tunkelang, **J. Koren**, P. Ogilvie, J. Wang. Social Navigation: A Position Paper. Proceedings of the 5th Workshop on Human-Computer Interaction and Information Retrieval (HCIR '11). Poster Session. Mountain View, CA, USA.

J. Koren, Y. Zhang, and X. Liu. Personalized Faceted Search. Proceedings of the 17th International Conference on the World Wide Web (WWW '08). Beijing, China.

J. Koren, Y. Zhang, S. Ames, A. Leung, C. Maltzahn, and E. Miller. Searching and Navigating Petabyte Scale File Systems Based on Facets. Proceedings of the 2nd ACM Petascale Data Storage Workshop (PDSW '07). International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '07). Reno, NV, USA.

Y. Zhang and **J. Koren**. Efficient Bayesian Hierarchical User Modeling for Recommendation Systems. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '07). Amsterdam, The Netherlands.

Book Chapters

M. Stefaner, S. Ferré, S Perugini, **J. Koren**, Y. Zhang "User Interface Design," in Dynamic Taxonomies and Faceted Search G. Maria and Y. Tzitzikas (Eds.). Springer 2009.

INVITED TALKS

June 2009 **Fuji Xerox Palo Alto Lab** **Palo Alto, CA**
Personalizing Faceted Search

April 2009 **Lawrence Livermore National Lab** **Livermore, CA**
Faceted Metadata Search for File Systems

PROFESSIONAL ACTIVITY

Local Arrangements Committee 17th International Conference on Information and Knowledge Management (CIKM 2008)

Program Committee Second International Workshop on Dynamic Taxonomies and Faceted Search (FIND '08).

Reviewer 31st - 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08 - '10).

TECHNICAL SKILLS

Programming languages: Go, Python, Java, Perl, Lisp, C/C++, Scala

Data Technologies: Hadoop / MapReduce, Pig, Kafka, SQL / Hive, Lucene

Web Technologies: HTML, PHP, XML, XSL, CSS, Javascript, D3

Experience: big data, data analytics, distributed systems, machine learning, personalization recommendation systems, search, natural language processing and understanding, data visualization