

Robert Krovetz

2804 Kent Place
Hillsborough, NJ 08844
908-566-7409
rkrovetz@lexicalresearch.com
<http://lexicalresearch.com>

RESEARCH INTERESTS

Computational linguistics; information retrieval; artificial intelligence and medicine.

EDUCATION

Ph.D., Computer Science, University of Massachusetts
M.S., Computer Science, University of Maryland
B.S., Computer Science, SUNY at Stony Brook

RESEARCH AND DEVELOPMENT EXPERIENCE

NLP Engineer

NYC Health and Hospitals, NYC, NY
2018 to 2019

Developed software to extract information from the provider's notes. Used this information to show improved compliance with metrics to evaluate hospital performance. Improved one metric from 17% to 45%. Also corrected over one million address mistakes in the database. Extracted provider's names from the notes in order to get better recognition of doctor/patient encounters. This work involved heuristics for resolving ambiguity of the names.

Senior Research Scientist

Right Answers, Edison NJ
2016 to 2018

Design and implement software for natural language processing and information retrieval. This includes work on multi-word expressions, morphology, and lexical semantics in order to improve search performance.

President

Lexical Research, Hillsborough, NJ
2006 to Present

Consultant on design and implementation of software for natural language processing. Conduct applied research on the lexicon, particularly word sense disambiguation. From 2010 to 2016 I worked on projects for the Educational Testing Service (ETS). This work involved 1) developing better methods to assess breadth and depth of vocabulary, 2) creating an inventory and ranking of multi-word expressions, 3) characterizing workplace vocabulary, 4) automatic item generation, 5) development of features involving

morphology for determining proficiency in English, and 6) improved metrics for assessing readability, My work led to a patent with ETS and a book on vocabulary teaching and assessment. I also worked for the University of Michigan on a system to assess student understanding of lexical ambiguity, and for other clients on a thesaurus that incorporates connotation, and a search engine that takes better account of semantics.

Principal NLP Engineer

CodeRyte, Bethesda, MD
2005 to 2006

Designed and implemented software for natural language processing of clinical patient records. Developed preprocessors to normalize the language and structure of the records. Developed software to assign a semantic class to a word based on medical morphology, and improved the handling of variant forms in the lexicon. Customized the coding engine to handle customer-specific requirements.

Senior Research Scientist and Manager of natural language group

Teoma (Ask.com), Piscataway, NJ
2003 to 2005

Responsible for projects to improve the performance of the search engine using natural language processing. These projects included: improved stop word processing, spelling correction, query paraphrase, classification of web documents, and evaluation of routines for stemming and smart answers.

Scientist

NEC Research Institute, Princeton, NJ
1996 to 2003

Performed research on word-sense disambiguation and information retrieval. Created large-scale inventory of word senses crucial for natural language applications. Manager and developer of 'Phrasebank' technology transfer system. Automatically extracted a set of multi-word terms from a large corpus to support a speech-to-speech machine translation system. The system significantly exceeded the goals of the project.

PATENTS

Krovetz, Robert and Deane, Paul. Computer-Implemented Systems and Methods for Non-Monotonic Recognition of Phrasal Terms. U.S. Patent 9,208,145. Issued Dec. 2015.

TEACHING EXPERIENCE

Invited Lecturer

European Summer School for Logic, Language, and Information, Birmingham, UK
2000.

Taught a one-week course on word-sense disambiguation. The course covered the central issues in this area and reviewed the state of the art.

Adjunct Associate Professor

University of Pennsylvania, Philadelphia, PA

Columbia University, New York, NY

1999-2001

Co-taught a graduate seminar entitled "Information Retrieval, Digital Libraries, and the Web".

SELECTED REFEREED PUBLICATIONS

Krovetz R, Deane P, and N Madnani, ``The Web is not a PERSON, Berners-Lee is not an ORGANIZATION, and African-Americans are not LOCATIONS: An Analysis of the Performance of Named-Entity Recognition", *Proceedings of the ACL Workshop on Multi-Word Expressions: From Parsing and Generation to the Real World (MWE 2011)*.

Krovetz R, "Word Sense Disambiguation, Lexical Semantics, and NLP Applications", *Proceedings of the NSF-Sponsored Symposium on Semantic Knowledge, Discovery, Organization, and Use*, 2008

Chen Y, Wang J, and R Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning", *IEEE Transactions on Image Processing*, Vol. 14(8), pp. 1187-1201, 2005

Park S, Pennock D, Giles C L, and R Krovetz, "Analysis of Lexical Signatures for Improving Information Persistence on the World Wide Web", *ACM Transactions on Information Systems*, Vol. 22(4), pp. 540-572, 2004

Chen Y, Wang J, and R Krovetz, "An Unsupervised Learning Approach to Content Based Image Retrieval", *Proceedings of the IEEE International Symposium on Signal Processing and its Applications*, pp. 197-200, 2003 (Invited paper)

Glover E, Pennock D, Lawrence S, and R Krovetz, "Inferring Hierarchical Descriptions", *Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM)*, pp. 507-514, 2002

Ugurel S, Krovetz R, Giles C L, Pennock D, Glover E, and H Zha, "What's the Code? Automatic Classification of Source Code Archives", *Proceedings of the Eighth ACM SIGKDD Conference on Knowledge and Data Discovery*, pp. 632-638, 2002

Lawrence S, Pennock D, Flake G, Krovetz R, Coetzee F, Glover E, Nielsen F, Kruger A, and C Lee Giles. "Persistence of Web References in Scientific Research", *IEEE Computer*, Vol. 34(2), pp. 26-31, 2001

Krovetz R, "More than One Sense per Discourse", *Proceedings of the ACL-SIGLEX Workshop*, 1998

Krovetz R, "Homonymy and Polysemy in Information Retrieval", *Proceedings of the 35th Meeting of the Association for Computational Linguistics*, pp. 72-79, 1997. Also appears as NEC Technical Report 97-141

Krovetz R, "Viewing Morphology as an Inference Process", *Proceedings of the Sixteenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 191-202, 1993. Also appears as UMASS-COINS Technical Report TR-93-36

Krovetz R, "Panel on Corpus Linguistics and Information Retrieval", *Proceedings of the Fifteenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 348-351, 1992 [Organizer and Chair of panel]

Krovetz R and Croft W B, "Lexical Ambiguity and Information Retrieval", *ACM Transactions on Information Systems*, Vol. 10(2), pp. 115-141, 1992

Invited Talks:

"More than One Sense per Discourse", USC-ISI, 2003, Sarnoff Research Lab, 2000, University of Massachusetts Amherst, 1998

"Homonymy and Polysemy in Information Retrieval", Princeton University, 1999

"The Use of Lexical Semantics for Information Retrieval", invited member of ACL-SIGLEX panel on "Lexical Semantics and Natural Language Applications", 1998

"Word Sense Disambiguation for Large Text Databases", NEC Research Institute, 1995, University of Maryland, 1998

PROFESSIONAL SERVICE

Reviewer for:

Machine Learning
Journal of Information Retrieval
Journal of Natural Language Engineering
IEEE Transactions on Knowledge and Data Engineering
Journal of the American Society for Information Science
Computational Linguistics
IEEE Expert
Information Processing and Management
ACM Transactions on Information Systems
Communications of the ACM

International Conference on Computational Linguistics (COLING)
International World Wide Web Conference
ACM Conference on Research and Development in Information Retrieval

Program Committee for SIGIR
Chair of Distinguished Lectureship committee for NJ-ASIS

HONORS AND AWARDS

Graduate Fellowship (General Electric), B.S. with Honors.