Michael Bendersky

University of Massachusetts

Department of Computer Science

140 Governors Drive

Amherst, MA 01003

bemike@cs.umass.edu

http://ciir.cs.umass.edu/~bemike

I am broadly interested in practical applications of information retrieval, natural language processing and statistical machine learning. Specifically, I worked on research problems in a variety of domains, including web search, recommendation systems, focused crawling, passage-based retrieval, query log mining, text reuse detection and sponsored search. As a result of this work, I created open-source and commercial software, and published in leading academic conferences.

My Ph.D. research combines insights from information retrieval, natural language processing and statistical machine learning to improve the effectiveness of search with complex natural language queries. I am excited about this research, since it has a potential to revolutionize the way people search on the web, in the enterprise and on mobile devices, to name just a few possible applications.

I am expected to complete my Ph.D. by the summer of 2012. I am looking for positions that combine interdisciplinary research, dynamic work environment and an ability to have a significant impact on product development and user experience.

EDUCATION

UNIVERSITY OF MASSACHUSETTS AMHERST **2007 – PRESENT**

Ph.D. candidate in Computer Science

Advisor: *Prof. W. Bruce Croft*

Research Topic: *Information Retrieval with Verbose Natural Language Queries*

TECHNION, HAIFA. INDUSTRIAL ENGINEERING AND MANAGEMENT DEPT. **2005 – 2007**

M.Sc. in Information Management Engineering

Advisor: *Dr. Oren Kurland*

Thesis: *Passage language models in ad hoc document retrieval*

TECHNION, HAIFA. INDUSTRIAL ENGINEERING AND MANAGEMENT DEPT. **2001 – 2004**

B.Sc. cum laude in Information Engineering

PROFESSIONAL EXPERIENCE

RESEARCH ASSISTANT. CIIR LAB, UNIVERSITY OF MASSACHUSETTS AMHERST **2007 – PRESENT**

Advisor: *Prof. W. Bruce Croft*

- Improving information retrieval with verbose natural language queries.
- Published papers in leading academic conferences and co-organized a series of Query Representation & Understanding workshops at the ACM SIGIR 2010 – 2011 conferences.

RESEARCH INTERN. YAHOO! RESEARCH, SANTA-CLARA SUMMER 2009

Advisor: *Evgeniy Gabrilovich*

- Development of novel structured indexing & retrieval techniques for sponsored search.
- Published a paper at the WWW 2010 conference and filed a patent application.

RESEARCH INTERN. MICROSOFT RESEARCH, REDMOND SUMMER 2008

Advisor: *Kenneth Church*

- Development of accurate models for query performance prediction and commercial intent detection using large-scale analysis of click data.
- Published a paper at the Learning to Rank for Information Retrieval Workshop at the ACM SIGIR 2009 conference.

SENIOR ENGINEER, QA TEAM. ALCATEL-LUCENT (FORMERLY MOBILITEC ISRAEL LTD.) 2004 -2007

- Development and integration of a unified automated testing framework
- Performance and scalability testing for large-scale web applications.

TEACHING EXPERIENCE

TEACHING ASSISTANT. UNIVERSITY OF MASSACHUSETTS AMHERST SPRING 2011

Instructor: *Prof. W. Bruce Croft*

Course: *"Search Engines"*

- Developed and graded homework assignments based on the course syllabus
- Presented a guest lecture on retrieval models

TEACHING ASSISTANT. TECHNION, HAIFA SPRING 2007

Instructor: *Dr. Oren Kurland*

Course: *"Advanced Topics in Information Systems - Information Retrieval"*

- Developed and presented a series of tutorials on practical aspects of Information Retrieval: indexing and retrieval, evaluation and using web search API's.
- Graded homework assignments.

RESEARCH PROJECTS

INFORMATION RETRIEVAL WITH VERBOSE QUERIES 2007-PRESENT

There is growing empirical evidence that current search engines do not, in general, perform well with verbose natural language queries. As a part of this project, we analyzed a large-scale search log to find under-performing types of queries. In addition, we proposed several effective techniques for improving retrieval performance of verbose queries, including key concepts detection, concept weighting and query segmentation. This project is ongoing, and thus far resulted in numerous conference and workshop publications.

FINDING TEXT REUSE ON THE WEB 2008-2009

The concept "text reuse" captures a wide scope of text transformations, including exact recapitulations, loose restatements, and reports on similar topics. In this project, we addressed the problem of finding instances of text reuse on the web. To this end, we proposed a general architecture for a text reuse web search engine, and proposed novel algorithms for several of its components, including ranking, timeline detection and link graph construction. This project resulted in a conference publication.

PASSAGE LANGUAGE MODELS IN AD HOC DOCUMENT RETRIEVAL 2006-2007

Passage-based document-retrieval uses the information from some of the document parts (or passages) to rank a document in response to a query. The proposed model generalizes some of the previously proposed methods for passage-based document retrieval and introduces a notion of document homogeneity, which controls the amount of document information used for retrieval. This project resulted in several conference and journal publications, and a thesis.

CRAWL-BY-EXAMPLE (GOOGLE SUMMER OF CODE) SUMMER 2006

Available at: http://sourceforge.net/projects/crawlbyexample/

Crawl-By-Example is a supervised classification module for Internet Archive's *Heritrix* crawler. *Crawl-By-Example* addresses the problem of finding relevant information in large collections of crawled web-pages. It classifies the processed pages by subjects and finds the best pages in each category, according to the examples provided by the operator. *Crawl-By-Example* is one of the first supervised focused crawler implementations, freely available as a part of an open-source crawler.

R-ECOMMAGENT (UNDERGRADUATE RESEARCH PROJECT) SUMMER 2004

Advisor: Dr. Onn Shehory, IBM Haifa Research Lab.

R-eCommAgent is an agent-based distributed platform, designed to provide shopping recommendations based on user system activity and profile. *R-eCommAgent* employs collaborative-filtering algorithms that produce improved recommendations as user system activity increases.

AWARDS AND HONORS

- Best Paper Runner-up: Search Category
  The 5th ACM International Conference on Web Search and Data Mining (WSDM 2012)
- Honorable Mention Award
  The 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011)
- Outstanding Synthesis Award (sponsored by Yahoo!)
  Department of Computer Science, University of Massachusetts Amherst (2011)

- Ph.D. Qualifier with Distinction
  Department of Computer Science, University of Massachusetts Amherst (2010)
- Yahoo! Award for Accomplishments in Search and Mining *(2009)*

INVITED TALKS

- "Parameterized Concept Weighting for Information Retrieval"
  (RALI Lab, Université de Montréal, September 2011)
- "Learning Concept Importance Using a Weighted Dependence Model"
  (Machine Learning & Friends Lunch, UMass Amherst, February 2010)
- "Discovering Key Concepts in Verbose Queries" (Technion, Haifa, Israel, January 2009)
- "Long Queries and Short Documents" (Microsoft Live Search, Redmond, August 2008)

CONFERENCE PUBLICATIONS

1. *Michael Bendersky, Donald Metzler and W. Bruce Croft*
   "Effective Query Formulation with Multiple Information Sources"
   In the Proc. of the 5th ACM International Conference on Web Search and Data Mining (WSDM 2012) (Selected for oral presentation) **(Best Paper Runner-up: Search Category)**
2. *Hang Li, Gu Xu, W. Bruce Croft, Michael Bendersky, Ziqi Wang, Evelyne Viegas*
   "QRU-1: A Public Dataset for Promoting Query Representation and Understanding Research"
   In the Proc. of the 2nd Workshop on Web Search Click Data at WSDM 2012
3. *Michael Bendersky, Donald Metzler and W. Bruce Croft*
   "Parameterized Concept Weighting in Verbose Queries"
   In the Proc. of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011) **(Honorable Mention Award)**
4. *Michael Bendersky, W. Bruce Croft and David A. Smith*
   "Joint Annotation of Search Queries"
   In the Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011)
5. *Michael Bendersky, W. Bruce Croft and Yanlei Diao*
   "Quality-Biased Ranking of Web Documents"
   In the Proc. of the 4th ACM International Conference on Web Search and Data Mining (WSDM 2011) (Selected for oral presentation)
6. *Michael Bendersky, David Fisher and W. Bruce Croft*
   "UMass at TREC 2010 Web Track"
   In the Proc. of the Text REtrieval Conference (TREC 2010)
7. *Michael Bendersky, W. Bruce Croft and David A. Smith*
   "Structural Annotation of Search Queries Using Pseudo-Relevance Feedback"
   In the Proc. of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010) (Short Paper)
8. *Van Dang, Michael Bendersky and W. Bruce Croft*
   "Learning to Rank Query Reformulations"

In the Proc. of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010) (Poster Paper)

9. *Michael Bendersky, Evgeniy Gabrilovich, Vanja Josifovski and Donald Metzler*
   "The anatomy of an ad: Structured Indexing and Retrieval for Sponsored Search"
   In the Proc. of the 19th International World Wide Web Conference (WWW 2010)
10. *Michael Bendersky, Donald Metzler and W. Bruce Croft*
    "Learning Concept Importance Using a Weighted Dependence Model"
    In the Proc. of the 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010)
11. *Michael Bendersky, W. Bruce Croft and David Smith*
    "Two-Stage Query Segmentation for Information Retrieval"
    In the Proc. of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2009) (Poster Paper)
12. *Michael Bendersky and Kenneth Church*
    "Priors in Web Search"
    Learning to Rank for Information Retrieval Workshop at SIGIR 2009
13. *Eyal Krikon, Oren Kurland and Michael Bendersky*
    "Utilizing Inter-Passage and Inter-Document Similarities for Re-ranking Search Results"
    In the Proc. of the 18th ACM International Conference on Information and Knowledge Management (CIKM 2009) (Short Paper)
14. *Michael Bendersky and W. Bruce Croft*
    "Analysis of Long Queries in a Large Scale Search Log"
    Workshop on Web Search Click Data at WSDM 2009 **(Selected as a Top Search Marketing and Social Media Project by the Internet Marketing Blog)**
15. *Michael Bendersky and W. Bruce Croft*
    "Finding Text Reuse on the Web"
    In the Proc. of the 2nd ACM International Conference on Web Search and Data Mining (WSDM 2009)
16. *Michael Bendersky and W. Bruce Croft*
    "Discovering Key Concepts in Verbose Queries"
    In the Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008)
17. *Michael Bendersky and Oren Kurland*
    "Re-ranking Search Results Using Document-Passage Graphs"
    In the Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008) (Poster Paper)
18. *Michael Bendersky and Oren Kurland*
    "Utilizing Passage-Based Language Models for Document Retrieval"
    In the Proc. of the 30th European Conference on Information Retrieval (ECIR 2008)

JOURNAL PUBLICATIONS

1. *Eyal Krikon, Oren Kurland and Michael Bendersky*
   Utilizing inter-passage and inter-document similarities for re-ranking search results
   (ACM Transactions on Information Systems, Volume 29, Number 1)
2. *Michael Bendersky and Oren Kurland*
   Utilizing passage-based language models for ad hoc document retrieval
   (Journal of Information Retrieval, Volume 13, Number 2, 157-187)

THESES

1. *Michael Bendersky*
   Passage Language Models in Ad Hoc Document Retrieval (Master Thesis) [Also available as a Tech. Report IE/IS-2007-04]

TRAVEL GRANTS

- Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR): 2008 – 2011
- ACM International Conference on Web Search and Data Mining (WSDM): 2009 – 2012
- Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT): 2011

PATENTS FILED

1. United States Patent Application 20110258034
   "Hierarchically-Structured Indexing and Retrieval"
   Inventors: Donald Metzler, Evgeniy Gabrilovich, Vanja Josifovski, Michael Bendersky

COMMITTEES & REVIEWING

- Co-organizer and co-chair
  - Workshop on Query Representation and Understanding (SIGIR 2010-2011)
- Program Committee Member
  - Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR): 2010 – 2012
  - International World Wide Web Conference (WWW): 2011, 2012
  - ACM International Conference on Web Search and Data Mining (WSDM): 2011
  - ACM International Conference on Information and Knowledge Management (CIKM): 2011
  - Annual Meeting of the Association for Computational Linguistics (ACL): 2012
  - Workshop on Web Search Click Data at WSDM 2012
- Member of a Poster Program Committee
  - Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR): 2009
- Member of a Demo Program Committee

    - International World Wide Web Conference (WWW): 2010
- Reviewer:
    - Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT): 2012
- Journal Reviewer
    - ACM Transactions on Information Systems
    - Knowledge and Information Systems
    - Machine Learning