

Qi Li

CONTACT INFORMATION	School of Information Sciences The University of Pittsburgh 135 N. Bellefield Ave. Pittsburgh, PA 15206 USA	<i>Mobile:</i> +1-650-215-8065 <i>E-mail:</i> qili@sis.pitt.edu <i>WWW:</i> www.sis.pitt.edu/~qili
OBJECTIVE	An academic position doing researches in retrieval, natural language processing, text mining, ontology and semantic web, especially making these techniques benefit clinical practices.	
RESEARCH INTERESTS	Information Retrieval, Entity Extraction, Text Mining, Medical Domain Natural Language Processing, Ontology, Semantic Web, Human-Computer Interaction	
EDUCATION	Ph.D. Candidate, School of Information Sciences, 2006. 9 to present The University of Pittsburgh , Pittsburgh, PA USA <ul style="list-style-type: none">• Thesis Topic: <i>Searching for Entities: When Retrieval Meets Extraction</i>• Adviser: Associate Professor Daqing He M.S., School of Electronic Engineering and Computer Sciences , 2002.9 to 2006.6 The Peking University , Beijing, China <ul style="list-style-type: none">• Thesis Topic: <i>Design and Implement Flaws Detection Tool for Object-Oriented Programs (Java)</i>• Adviser: Professor Weizhong Shao• Area of Study: Software Engineering B.S., Department of Information Management, 1996.9 to 2000.6 The Peking University , Beijing, China <ul style="list-style-type: none">• Information Sciences and Management• Thesis Topic: <i>Design and Implement Yinhan News System</i>• Minor in Economics	
PUBLICATIONS	Li, Q. , Tsui, R., Dowling, J., Christensen, L., Chapman, W., et al. Concept Extraction from Emergency Department Reports for Automatic Case Detection: More than Mapping UMLS Concepts, submitted to <i>AMIA</i> 2011. Li, Q. , He, D., Searching for Entities:When Retrieval Meet Extractions, <i>TREC</i> 2010. Chapman, W., Conway, M., Dowling, J, Tsui, R., Li, Q. , Christensen, L., Harkema, H., Sriburadej, T., Espino, J. Challenges in Adapting an NLP System for Real-time Surveillance, (Abstract), <i>ISDA</i> 2010 Li, Q. , He, D., Towards Classifying Exploratory Search Results with Wikipedia, <i>eCASE & eTech</i> 2010. Li, Q. , He, D., Mao, M., A Study of Relation Annotation in Business Environments Using Web Mining, <i>ICSC</i> 2009. Mao, M., Li, Q. , A Sensemaking-based Information Foraging and Summarization System in Business Environments, <i>EEE</i> 2009. Hsiao, I., Li, Q. and Lin, Y. (2008) Educational Social Linking in Example Authoring, In Proceedings of <i>18th conference on Hypertext and hypermedia, HT '08</i> , Pittsburgh, USA, 19-21 June, 2008, ACM Press.	

Pu, Q. He, D., **Li, Q.**. (2008) University of Pittsburgh at GeoCLEF 2008: Towards Effective Geographic Information Retrieval. In the Proceedings of *Cross-Language Evaluation Forum*.

Ahn, J., Brusilovsky, P., He, D., Grady, J., **Li, Q.** (2008) Personalized Web Exploration with Task Models. In *World Wide Web Conference 2008*.

He, D., Brusilovsky, P., Grady, J., **Li, Q.** (2007) How Up-to-date should it be? The Value of Instant Profiling and Adaptation in Information Filtering. In Proceeding of *International Conference on Web Intelligence*.

WORKING
EXPERIENCE

TREC Entity Retrieval, School of Information Sciences, University of Pittsburgh

Entity Retrieval (Dissertation)

January 2010.1 to Present

- Decomposing the black box of Entity Retrieval into two parts: document retrieval and entity extraction.
- Theoretical demonstrating the decomposition process using probability model.
- Bringing state-of-the-art techniques in document retrieval and entity extraction into entity retrieval.
- Evaluating the entity retrieval system in two layers, which in turns improves the overall system performance.
- A demo system of entity retrieval on museum archive search with the techniques of Indri document indexing, servlet dealing with the front-end queries, and the front-end interface.

Center of Disease Control and Prevention, Department of Biomedical Informatics, University of Pittsburgh

Research Assistant

January 2010.1 to June 2011

- Advisors: Dr. Wendy W Chapman, Rich Tsui, Henk Harkema
- Working on the project of medical domain entity extraction (Topaz) and disease detection, especially on the disease of Shigellosis and Influenza.
- Designing and Implementing on a medical entity extraction result evaluation tool, called ARR, to facilitate the extraction result monitoring.
- Maintaining and debugging the GATE-version Topaz system; updating the rules using in the system; conducting the experiments and evaluating the performance.
- Maintaining and debugging the UIMA-version Topaz system.
- Studies on the different features, such as POS, for the medical entity extraction.

Collaborative Filtering Search System (NSF), School of Information Sciences, University of Pittsburgh

Research Assistant

September 2009 to December 2009

- Advisors: Dr. Daqing He
- Studies on the user collaborative search behavior on the Web environment, especially in Citeulike website.
- Mining the Citeulike dataset, including the users information (user id), the document information (title, abstract, author, keywords, url, etc.), group information (which users belong to the special group), and library information (which books are in the someone's library).
- Updating a Firefox plug-in for the user behavior tracking, including user login/logout, user browsing, user search, user bookmarking, etc.
- Collecting the front-end user data, and storing the data in the back-end (MySQL) database for the user behavior study.

Continuous Information SenseMaking SAP Research, Palo Alto, California USA

Internship

May 2008 to August 2009

- Advisors: Dr. Ming Mao
- Working in Topic Track Continuous Sensemaking which focuses on the free-text document analysis.
- Studying and implementing the entity and relation extraction from Knowledge Base (Freebase) and the free texts.
- Classifying and disambiguating search results according to the context.
- Estimating the importance of topics at each stage the according to the detected entities and relations on the ontology network; detecting the topics continuously by summarizing the topics at each stage.

Geography Query Parsing, School of Information Science, University of Pittsburgh USA

CLEF Geographical Information Retrieval Track **January 2008 to May 2008**

- Geography name entity identification. Mining geography name entities with their corresponding coordinates from Wikipedia; geography name entity extractors building on the Wikipedia-mined geography thesaurus.
- Geography name entity disambiguation. Designing components to disambiguate locations by estimating the distance between two locations and geography context information.
- Query type identification. Classifying the query types by identifying the nearest-neighbor nodes in semantic knowledge networks; generating different queries according to the different query types.

Comprehensive Exam, School of Information Science, University of Pittsburgh
December 2009

Topic: *Information Retrieval and Semantic Retrieval* with Prof. Daqing He.

Topic: *Text Mining and Ontology* with Prof. Micheal Spring

Topic: *Adaptive Web and Personalization* with Prof. Peter Brusilovsky

Prelime Exam, School of Information Science, University of Pittsburgh USA

Toward Classify Exploratory Search Results with Wikipedia **September 2007 to December 2008**

- Classifying exploratory search results to support people for Request For Information (RFI) search task.
- Identifying the potential concepts in queries using Wikipedia entries, and evaluate the effectiveness of the algorithm.
- Demonstrating the possibility of using Wikipedia articles as training data sets to help classifying search results task and concluding the rules for collecting the training datasets.

GAIL (DARPA), School of Information Science, University of Pittsburgh USA

Research Assistant

January 2007 to December 2007

- Passage-level Document Relevancy Evaluation
- Continuing and completing the passage-level precision and recall evaluation tool component.
- This tool is Java-based with inputs of snippets and outputs of passage-level precision.

TECHNICAL SKILLS

Programming: Java, Servlet, JavaScript, PHP, PYTHON,XML, HTML, SQL, etc.

Indexing and Retrieval Tools: **Indri/Lemur**, Google/Yahoo API, MySQL, Protege,

NLP Tools: **GATE**, UIMA, Stanford NLP tools

Machine Learning and Statistic Tools: Weka, Matlab, SPSS

Operating Systems: Microsoft Windows family, Linux,