Xitong Liu

Contact

209 Evans Hall
University of Delaware
Newark, DE 19716

Phone : 302-740-0812
E-mail : xtliu@udel.edu
Web : http://xliu.org

Summary

- 5 years of research experience in Information Retrieval.
- Developed and shipped the ethnicity+profession entity carousel on Bing.
- Teamwork with researchers from HP Labs on two projects, which yields three papers and one U.S. patent.
- Developed a system ranked 2nd out of 43 runs in Knowledge Base Acceleration task of the Text REtrieval Conference (TREC), 2012.
- Developed online search evaluation system for course project of Search and Data Mining, which served 60 students so far.
- 3 years of development experience in C++ with 100,000 lines code when in college.
- Led one project team with 6 undergraduate students when in college.

Education

Ph.D. in Computer Engineering **2009.9 ∼ 2015.6 (expected)**
University of Delaware, Newark, DE.

B.E. in Electronic Information Engineering **2005.9 ∼ 2009.6**
Huazhong University of Science and Technology, Wuhan, China

Working Experience

Research Assistant **2009.9 ∼ Present**
Advisor: Dr. Hui Fang, University of Delaware, Newark, DE.

Research Intern **2014.6 ∼ 2014.9**
Mentor: Dr. Hao Hu, Microsoft Research/Bing, Bellevue, WA.

Research Intern **2011.6 ∼ 2011.8**
Mentor: Dr. Fei Chen, HP Labs, China.

Honors and Awards

- University Dissertation Fellows, University of Delaware, 2014.
- Profession Development Award, University of Delaware, 2013.
- CIKM Travel Grants, 2011.

Skills

- Programming Language: C, C++, Perl (proficient), Python, Java, HTML
- Operating System: Mac OS X, GNU/Linux, Windows
- Toolkit/Platform: Lemur/Indri (proficient), Lucene/Solr, Hadoop

Projects

Knowledge Base Acceleration **2012.6 ∼ 2013.8**

- Task: automate the labor-intensive process of selecting citation-worthy documents to keep knowledge base up-to-date.
- Challenges: filter out relevant documents from a massive web document stream.
- Solutions: select related entities from the profile of topic entity, employ an iterative algorithm to assign weight for each related entity, and use the weighted entities to estimate relevance between topic entity and documents.
- Achievements: our initial system reaches 2nd place out of 11 groups in TREC KBA Track 2012 (1.1TB compressed data), the improved system later outperforms state-of-the-art methods in TREC 2012 and reaches 3rd place out of 11 groups in TREC KBA Track 2013 (4.5TB compressed data).

Enterprise Search II **2011.6 ∼ 2012.12**

- Task: improve document search accuracy through leveraging entity information from database.
- Challenges: find related entities from database, integrate related entities to improve document search quality.
- Solutions: train a domain-specific named entity recognizer based on Conditional Random Fields (CRFs); devise a query expansion framework to integrate entities and their relations to update query and improve retrieval performance.
- Achievements: our system significantly outperforms two state-of-the-art methods by 5%∼10% on two enterprise data sets and 5% on one TREC newswire collection.

Enterprise Search I **2010.9 ∼ 2011.5**

- Task: improve enterprise search quality through a case study: finding relevant information of certain types from enterprise data.
- Challenges: identify query requirements; rank information items.
- Solutions: apply semantic similarity based clustering and language modeling to identify content and type requirements; rank information items based on 2-dimensional search model: an relevant information item should satisfy both requirements.
- Achievements: our system significantly outperforms one strong baseline by 20% on one enterprise data set and 7% on one open semantic linked data set.

Entity Search **2009.12 ∼ 2010.8**

- Task: develop a system to retrieve entities instead of documents to fulfill user's information need directly.
- Challenges: related entity extraction; entity ranking.
- Solutions: employ Stanford NER tagger to extract entity candidates which co-occur with query from documents, and apply DBpedia to improve entity candidates quality; extended expert finding probabilistic framework to rank entities based on three requirements from query: input entity, relation and target entity type; improve entity ranking through feedback.
- Achievements: our system can be ranked at the 3rd place in TREC Entity Track 2009 and the 4th place in TREC Entity Track 2010.

Selected Publications

Yue Wang, **Xitong Liu** and Hui Fang. A Study of Concept-based Weighting Regularization for Medical Records Search. In *52nd Annual Meeting of the Association for Computational Linguistics (**ACL'14**)*, pages 603-612, 2014.

Xitong Liu, Fei Chen, Hui Fang and Min Wang. Exploiting Relationship for Query Expansion in Enterprise Search. In *Information Retrieval*, Volume 17, Issue 3, June 2014, pages 265-294.

Xitong Liu, Jerry Darko and Hui Fang. A Related Entity based Approach for Knowledge Base Acceleration. In *online Proceedings of the 22nd Text REtreival Conference (TREC'13)*, 2013.

Xitong Liu and Hui Fang. Leveraging Related Entities for Knowledge Base Acceleration. In *4th International Workshop on Web-scale Knowledge Representation, Retrieval, and Reasoning (Web-KR)*, 2013.

Xitong Liu, Wei Zheng and Hui Fang. An Exploration of Ranking Models and Feedback Method for Related Entity Finding. In *Information Processing & Management*, Volume 49, Issue 5, September 2013, pages 995-1007.

Xitong Liu and Hui Fang. Entity Profile based Approach in Automatic Knowledge Finding. In *online Proceedings of the 21st Text REtreival Conference (TREC'12)*,

2012.

Xitong Liu, Hui Fang, Fei Chen and Min Wang. Entity Centric Query Expansion for Enterprise Search. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (**CIKM'12**)*, pages 1955-1959, 2012.

Xitong Liu, Hui Fang, Cong-Lei Yao and Min Wang. Finding Relevant Information of Certain Types from Enterprise Data. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (**CIKM'11**)*, pages 47-56, 2011.

Patents

- Performing a search based on entity-related criteria, PCT/US2012/061034, WO 2014062192 A1.