

Xiaohua (Davis) Zhou

College of Information Science & Technology, Drexel University

3141 Chestnut Street, Philadelphia, PA 19104

(215) 840-6193

xiaohua.zhou@drexel.edu

<http://www.pages.drexel.edu/~xz37>

Educational Background

- 2003.9~Present College of Information Science & Technology (IST), Drexel University Philadelphia, PA 19104, USA, Ph.D. (expected in June 2008)
- 1999.9~2002.3 Shanghai Jiao Tong University (SJTU), Shanghai, 200030, China
M.S. in Management Science and Engineering, 2002
- 1995.9~1999.7 Shanghai Jiao Tong University (SJTU), Shanghai, 200030, China
B.S. in Automatic Control, B.A. in International Trade

Research Interest

Information Retrieval, Text Clustering and Classification, Question Answering, Biomedical Literature Mining, Information Extraction

Peer Reviewed Publications

Journal Papers:

- **X. Zhou**, X. Hu, and X. Zhang, "Topic Signature Language Model for Ad hoc Retrieval," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 19 no 9, 1276-1287, Sept., 2007
- X. Zhang, X. Hu, X. Shen, **X. Zhou**, and P. Achananuparp, "Utilization of Global Ranking Information in Graph- based Biomedical Literature Clustering," to appear in *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 4 issue 4 (special issue of DaWaK 2007)
- X. Zhang, L. Jing, X. Hu, M. Ng, J. Xia, and **X. Zhou**, "Medical Document Clustering Using Ontology Based Term Similarity Measures," accepted to be published in *International Journal of Data Warehousing and Mining (IJDWM)*
- **X. Zhou**, X. Hu, G. Li, X. Lin, X. Zhang, "Relation-based Document Retrieval for Biomedical IR", *Transactions on Computational Systems Biology (CSB)*, vol. 4, 112-128, 2006

Conference Papers:

- **X. Zhou**, X. Zhang, and X. Hu, "Semantic Smoothing of Document Models for Agglomerative Clustering," in the Twentieth International Joint Conference on Artificial Intelligence (*IJCAI 2007*), Jan. 6-12, 2007, India, 2922-2927 (15.7%)
- **X. Zhou**, X. Hu, X. Zhang, and X. Shen, "A Segment-based Hidden Markov Model for Real-Setting Pinyin-to-Chinese Conversion," In *the Proceedings of ACM 16th Conference on Information and Knowledge Management (CIKM 2007)*, November 6-8, 2007, Lisboa, Portugal, 1027-1030 (26%)
- X. Zhang, L. Jing, X. Hu, M. Ng, and **X. Zhou**, "A Comparative Study of Ontology Based Term Similarity Measures on Document Clustering," *In the 12th International conference on Database Systems for Advanced Applications (DASFAA2007)*, 115-126 (18.7%)

- **X. Zhou**, X. Zhang, and X. Hu, "Dragon Toolkit: Incorporating Auto-learned Semantic Knowledge into Large-Scale Text Retrieval and Mining," In *proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, October 29-31, 2007, Patras, Greece (47%)
- **X. Zhou**, X. Hu, X. Zhang, X. Lin, and I.-Y. Song, "Context-sensitive Semantic Smoothing for Language Modeling Approach to Genomic Information Retrieval", accepted in *the 29th Annual International ACM SIGIR Conference (SIGIR 2006)*, Aug 6-11, 2006, Seattle, WA, USA, 170-177 (18.5%)
- X. Zhang, **X. Zhou**, and Xiaohua Hu, "Semantic Smoothing for Model-based Document Clustering," in *IEEE International Conference on Data Mining (ICDM06)*, Dec. 18-22, 2006, Hong Kong, 1193-1198 (20%)
- X. Hu, X. Zhang, and **X. Zhou**, "Integration of Cluster Ensemble and EM based Text Mining for Microarray Gene Cluster Identification and Annotation", In *the Proceedings of ACM 15th Conference on Information and Knowledge Management (CIKM 2006)*, Arlington, Virginia, USA, 824-825 (25%)
- **X. Zhou**, X. Hu, X. Lin, H. Han, and X. Zhang, "Relation-based Document Retrieval for Biomedical Literature Databases", accepted by the *11th International Conference on Database Systems for Advanced Applications (DASFAA 2006)*, April 12-15, Singapore, 689-701 (25%)
- **X. Zhou**, X. Zhang, and X. Hu, "Using Concept-based Indexing to Improve Language Modeling Approach to Genomic IR", in the *28th European Conference on Information Retrieval (ECIR2006)*, April 10-12, 2006, London, UK, 444-455 (21%)
- **X. Zhou**, X. Zhang, and X. Hu, "MaxMatcher: Biological Concept Extraction Using Approximate Dictionary Lookup", In the *9th biennial The Pacific Rim International Conference on Artificial Intelligence (PRICAI 2006)*, Aug 9-11, 2006, Guilin, China, 1145-1149 (30.9%)
- X. Zhang, D. Wu, **X. Zhou**, and X. Hu, "A Language Modeling Text Mining Approach to the Annotation of Protein Community", In *the 6th IEEE Symposium on Bioinformatics and Bioengineering (BIBE 2006)*, 12-19 (38.8%)
- **X. Zhou**, H. Han, I. Chankai, A. Prestrud, and A. Brooks, "Approaches to Text Mining for Clinical Medical Records ", accepted by the *21st Annual ACM Symposium on Applied Computing 2006, Technical tracks on Computer Applications in Health Care (SAC 2006)*, April 23 -27, 2006, Dijon, France (32%)
- **X. Zhou** and H. Han, "Survey of Word Sense Disambiguation Approaches", Accepted by *the 18th International FLAIRS Conference*, May 15-17, 2005, Florida, FL, USA, 307-313. (53.4%)

Workshop Papers:

- **X. Zhou**, X. Hu, X. Zhang, and D. Wu, "A Mixture Language Model for Class-Attribute Mining from Biomedical Literature Digital Library," *International Workshop on Knowledge Discovery and Management in Health Informatics*, Nov. 2-4, 2007, Fremont, California
- **X. Zhou**, H. Han, I. Chankai, A. Prestrud, and A. Brooks, "Converting Semi-structured Clinical Medical Records into Information and Knowledge", In the *International Workshop on Biomedical Data Engineering*, April 3-4, 2005, Tokyo, Japan.
- **X. Zhou** and I.-Y. Song, "Conceptual Modeling of Genetic Studies and Pharmacogenetics", *the International Workshop on Data Mining and Bioinformatics*, May 09-12, 2005, Singapore, 402-415.

Book Chapters:

- X. Hu, X. Zhang, and **X. Zhou**, "A Comprehensive Comparison Study of 7 Methods of Mining Hidden Links from Biomedical Literatures," *Knowledge Discovery in Bioinformatics: Techniques, Methods and Applications*, X. Hu & Y. Pan (Eds). Wiley & Son, 2006

Papers under Review:

- **X. Zhou**, X. Zhang, and X. Hu, "Semantic Smoothing for Bayesian Text Classification and Clustering," *ACM Transactions on Information Systems (TOIS)*
- **X. Zhou**, X. Zhang, and X. Hu, "Semantic Smoothing for Bayesian Text Classification with Small Training Data," *2008 SIAM Conference on Data Mining (SIAM2008)*

Teaching Experiences

- Instructor of Database Management Systems, College of IST, Drexel University, Fall 2007
- Teaching Assistant of Database Management Systems, College of IST, Drexel University, Summer 2007
- Instructor of Microsoft Certified Solution Developer (MCSD) Program
Nan Yang Computer Training Center, Shanghai, China, Feb. 2001– Aug. 2001
Wisenet Education Center, Shanghai, China, July 2001– Oct. 2001

Conference Presentations

- International Joint Conference on Artificial Intelligence (IJCAI), January, 2007, Hyderabad, India
- ACM SIGIR, August, 2006, Seattle, WA, USA
- The 28th European Conference on Information Retrieval (ECIR2006), April, 2006, London, UK

Academic Services

- Program Committee Member:
 - 2007 IEEE International Symposium on Data Mining and Information Retrieval (DMIR)
 - International Workshop on Data Mining in Bioinformatics with IEEE ICDM 2006
- Ad hoc Journal Reviewer: TKDE, KAIS, Bioinformatics, IJSOI
- External Referee: ER2004, eCOMO2004, DGOV2004, CAiSE2005, NSF ITR Program 2005

Working Experiences

- CTO of HiSun Software, Shanghai, China, 2002-2003.
Developed a prototype of Intelligent Tutoring System (ITS) for Grade 7 Math consisting of three components: profiling and personalizing, diagnostics, and problem solving wizard.
- System Analyst of ecSolutions, a Hong Kong based Software Company in Shanghai, China, 2002
 - Developed Document Preparation Systems for Citic Ka Wah Bank, Hong Kong, which allowed clients of banks to prepare documents needed for issuing or cashing Letters of Credit online.
 - Developed Internal Settlement System for Shanghai Power Group, which exchanged accounting information with legacy systems distributed in 8 commercial banks and 49 subsidiary companies.

Software and Demos

- **Dragon Toolkit**: A Java-based Toolkit for Language Modeling Approach to Information Retrieval and Text Mining (<http://www.dragontoolkit.org>, more than 500 downloads since the first release in April, 2007)
- **Web-based Factoid Question Answering** (online demo, <http://www.dragontoolkit.org/qa.asp>)
- **Web-based Entity Classifier** (online demo, <http://www.dragontoolkit.org/cptc.asp>)
- **Dragon Pinyin**: A Segment-based Hidden Markov Model for Real-Setting Pinyin-to-Chinese Conversion (online demo, <http://www.dragontoolkit.org/pinyin.asp>)

Honors and Awards

- Research Day Winner, Drexel University, April 2007
- Research Day Winner, Drexel University, April 2006
- Dean's Award (Research Day), IST, Drexel University, April 2006
- SIGIR Student Travel Award, August 2006
- Research Assistantship, Drexel University, 2003-2007
- Excellence Scholarship, Shanghai Jiao Tong University, 1995-1998
- Gold Medal, National High School Olympic Mathematical Contest, 1995.

Technical Skills

- Programming experiences: C, C++(6 years), Visual Basic(5 years), and Java(6 years)
- Databases: MS SQL Server, Oracle, and DB2
- Distributed Computing: DCOM, CORBA, and EJB
- Web Application Development: ASP, PHP, and JSP/Servlet
- Mathematical modeling software: SAS 8.2, SPSS 11, and Matlab 6.1

Professional Memberships

- IEEE Student Member
- ACM Student Member
- Certified System Analyst - China, April 2001

References

- Xiaohua Hu, Associate Professor
College of Information Science & Technology, Drexel University
3141 Chestnut Street, Philadelphia, PA 19104-2875, (215) 895-0551, thu@ischool.drexel.edu
- Il-Yeol Song, Professor
College of Information Science & Technology, Drexel University
3141 Chestnut Street, Philadelphia, PA 19104-2875, (215) 895-2489, song@drexel.edu
- Xia Lin, Associate Professor
College of Information Science & Technology, Drexel University
3141 Chestnut Street, Philadelphia, PA 19104-2875, (215) 895-2482, xlin@ischool.drexel.edu