

JIANHAN ZHU

Www: <http://kmi.open.ac.uk/people/jianhan/>

RESEARCH STATEMENT

Currently my research is focused on entity search in particular expert search in organizational intranets. I have represented the Open University to achieve the best search results (out of 23 teams) in the TREC2006 Enterprise Track Expert Search task following a two-stage model integrating multiple document features. I have participated in TREC2006 and 2007 enterprise track, and INEX2007 entity ranking track.

I have led the development of the CORDER (Community Relation Discovery by Named Entities) system. We proposed a novel CORDER method which measures associations between named entities in a text corpus by integrating their occurrences, distances in text, and frequencies in documents. We have integrated CORDER with an online social networking tool called BuddyFinder for matchmaking on the web.

I authored the ESpotter tool, which provides effective named entity recognition for a domain by adapting its lexicon and patterns to the domain through a URL tree. ESpotter has a moderate popularity in terms of the number of users.

My PhD research (1999-2003) is concerned with constructing link hierarchies from the link structure of a website for helping website users navigate and search the website for desirable information.

RESEARCH INTEREST

Information retrieval, expert search, enterprise search, multimedia information retrieval, text mining, information extraction, named entity recognition, multimedia information retrieval, semantic web, natural language processing, web link structure mining, web mining

EDUCATION

- Ph.D. Faculty of Informatics, University of Ulster**
May 1999- Oct. 2003 Supervisors: Dr. Jun Hong and Prof. John G. Hughes.
Dissertation title: "Mining Web site link structures for adaptive Web site navigation and search".
Viva date: October 2003
- B.S. (Hons) Tsinghua University, Beijing, China**
Sept. 1993- July 1998 Department of Automation. Emphasis in process control and data communications.

EXPERIENCE

- Research Fellow Knowledge Media Institute, The Open University**
Aug. 2003- Now
- Project manager of JISC-funded DYNIX project investigating and comparing metadata based cross search systems, and developing a dynamic query meta-search interface
 - Key researcher of Expert Search which integrates document features and domain knowledge
 - A prototype expert search application developed for the Open University domain
 - Key researcher of CORDER: software for entity association network discovery from text
 - Co-investigator of "Entity based semantic search and knowledge management" project funded in part by an IBM 2006 UIMA (Unstructured Information Management Framework) Innovation Award
 - Co-investigator of "A novel module integrating domain knowledge in entity-based search for the UIMA framework" project funded in part by an IBM 2007 UIMA (Unstructured Information Management Framework) Innovation Award
 - Research Fellow on EPSRC funded AKT (Advanced Knowledge Technologies), and EU funded Dot.Kom, ELeGI, and Pharos projects
 - Key researcher of BuddyFinder-CORDER: matchmaking and expert search on the web
 - Designer and author of ESpotter: an adaptive named entity recognition framework, ESpotter Internet Explorer plug-in, ESpotter powered KMi semantic website, integrate ESpotter with Magpie. ESpotter has a moderate popularity.
 - Collaborators: Dr. Dawei Song, Prof. Stefan Rueger, Dr. Victoria Uren, Prof. Marc Eisenstadt, Prof. Enrico Motta, Christ Denham, Alexandre Goncalves, Dr. Jun Hong, Prof. John G. Hughes

- Ph.D. Faculty of Informatics, University of Ulster**
- Research Assistant**
May 1999-Aug. 2003
- PhD project: Website adaptation by learning from website structures and user access patterns, in NIKEL (Northern Ireland Knowledge and Engineering Laboratory).
Supervisors: Dr. Jun Hong and Prof. John G. Hughes.
 - Co-author of a website data mining software Easyminer with MineIt Company Ltd. (<http://www.mineit.com>), which won European IST Awards in November 2000
With: Prof. John G. Hughes, Dr. Sarabjot S. Anand, Dr. Alex Büchner, and Maurice Mulvenna

- Research Assistant**
July 1995-Apr. 1999
- Information R&D Centre, Tsinghua University, Beijing, China**
- Work on projects in the context of the Chinese National Ninth 5-Year Program
Supervisors: Prof. Yanda Li and Prof. Yaoxue Zhang.
- Author of a document format conversion software, which transforms academic papers into HTML files
- Software Packages**
- Expert Search, CORDER, ESpotter, EasyMiner (MineIt Company Ltd, 2000)

DISTINCTIONS

- Grants**
- JISC DYNIQX project (with Prof. Marc Eisenstadt), 2007, £50,000
- Awards**
- Represent the Open University to get No. 1 position in TREC2006 Enterprise Track Expert Search task
 - IBM UIMA 2007 Innovation Award (with Dr. Dawei Song), \$24,000USD, 2007
 - IBM UIMA 2006 Innovation Award (with Dr. Dawei Song), \$19,000USD, 2006
 - Vice Chancellor Research Scholarship (VCRS), University of Ulster, 1999-2002
 - Excellent Student Scholarship, Tsinghua University, Beijing, China, 1993-1998
- Professional Activities**
- Program committee member in Future Challenges in Expertise Retrieval Workshop proposal for ECIR (European Conference on Information Retrieval) 2008
 - Reviewer: JASIST, CIKM2007, IEEE TKDE, ACM Transactions on Asian Language Information Processing (TALIP), WAIM2006, Information Systems Journal, Int. J. of Information Technology & Decision Making, ACM CHI 2006 Notes, WWW 2006 posters, ESWC 2006, ASWC 2006, ACM CHI 2003, IFIP AI-2006, IEEE CEC 2006/EEE, AMCIS 2006, CCSCE 2003, SOAIC2006
 - Member: ACM, IEEE, IEEE-Computer
- Supervision**
- Visiting PhD student: Alexandre Goncalves (2004-2005), Intern student: Camilo Thorne (2005)
- Invited talk**
- “Mining Conceptual Link Hierarchies for Adaptive Web Navigation”, Microsoft Research, July 2003, Cambridge, UK
 - “Expert Search, Entity Retrieval, and Beyond”, IBM T.J. Watson Research Center, Nov 6th 2007, NY, USA

PUBLICATIONS (Numbers in square brackets are citation counts from Google scholar)

- Refereed Journal Papers** Jianhan Zhu, Alexandre L. Gonçalves, Victoria S. Uren, Enrico Motta, Roberto Pacheco, Marc Eisenstadt and Dawei Song (2007a) Relation Discovery from Web Data for Competency Management. *Web Intelligence and Agent Systems: An International Journal*, Vol.5, 4, IOS Press.
- Jianhan Zhu, Marc Eisenstadt, Alexandre L. Gonçalves, Chris Denham, Victoria S. Uren and Dawei Song (2007b) Social Search With Missing Data: Which Ranking Algorithm? *Journal of Digital Information Management Special Issue on Web Information Retrieval*, Pit.Pichappan, Keith van Rijsbergen and Iadh Ounis (eds), Vol.5, 5, pp. 249-261.
- Jianhan Zhu, Jun Hong, and John G. Hughes (2004) PageCluster: Mining Conceptual Link Hierarchies from Web Log Files for Adaptive Web Site Navigation. *ACM Transactions on Internet Technology (TOIT) special issue on "Machine Learning for the Internet"*, Gary William Flake, Paolo Frasconi, C. Lee Giles, Marco Maggini (eds.) Vol. 4, No. 2, pp. 185-208. [6]
- Joseph Fong and Jianhan Zhu (2002) Online Web Mining Transactions Association Rules using Frame Metadata Model. *Journal of Applied Systems Studies (JASS) special issue on "Web Information Systems Applications"*, Vol. 3, No. 2, pp. 313-328, Cambridge International Science Publishing. [3]
- Not Refereed Journal Papers** Victoria Uren, Jianhan Zhu, Alexandre Goncalves (2005) Applying Mined Social Networks to Knowledge Management Scenarios. *AIS SIGSEMIS Bulletin*, November 2nd.
- Refereed Conference Papers** Jianhan Zhu, Alexandre L. Gonçalves, Victoria S. Uren, Enrico Motta, Roberto Pacheco, Dawei Song, and Stefan Rueger (2007c) CORDER: COMMUNITY Relation Discovery by named Entity Recognition. In Proc. of International Conference on Machine Learning and Cybernetics (ICMLC) 2007, Hong Kong, China, IEEE.
- Alexandre Goncalves, Jianhan Zhu, Dawei Song, Victoria Uren, Roberto Pacheco (2006) LRD: Latent Relation Discovery for Vector Space Expansion and Information Retrieval. In Proc. of The Seventh International Conference on Web-Age Information Management (WAIM 2006), June, Hong Kong, China (Best student paper award). [3]
- Jianhan Zhu, Marc Eisenstadt, Dawei Song, Chris Denham (2006) Exploiting Semantic Association to Answer Vague Queries. In Proc. of Fourth International Conference on Active Media Technology (AMT 2006), June, Brisbane, Australia, pp. 73-78, IOS Press. [1]
- Yuangui Lei, Marta Sabou, Vanessa Lopez, Jianhan Zhu, Victoria Uren, Enrico Motta (2006) An Infrastructure for Acquiring High Quality Semantic Metadata. In Proc. of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro. [6]
- Jianhan Zhu, Marc Eisenstadt, Alexandre L. Gonçalves, and Chris Denham (2005a) BuddyFinder-CORDER: Leveraging Social Networks for Matchmaking by Opportunistic Discovery. In Proc. of ISWC2005 Workshop on Semantic Network Analysis, November 7, Galway, Ireland. [1]
- Jianhan Zhu, Alexandre L. Gonçalves, Victoria S. Uren, Enrico Motta, and Roberto Pacheco (2005b) Mining Web Data for Competency Management. In Proc. of Web Intelligence 2005 (WI'2005), France, September 19-22, 2005, pp. 94-100.
- Jianhan Zhu, Alexandre L. Gonçalves, Victoria S. Uren, Enrico Motta, and Roberto Pacheco (2005c) CORDER: COMMUNITY Relation Discovery by named Entity Recognition. In Proc. of Third Int. Conf. on Knowledge Capture (K-Cap'2005), October, Banff, Canada, pp. 219-220 [2]
- Jianhan Zhu, Victoria Uren, and Enrico Motta (2005d) ESpotter: Adaptive Named Entity Recognition for Web Browsing. In Proc. of Professional Knowledge Management Conference (WM), Springer, 518-529. [22]
- Jianhan Zhu, Victoria Uren, and Enrico Motta (2005e) ESpotter: Adaptive Named Entity

Recognition for Web Browsing. In Proc. of Workshop on IT Tools for Knowledge Management Systems at WM2005 Conference, Kaiserslautern, Germany, April 11-13. [22]

Jianhan Zhu, Jun Hong, and John G. Hughes (2002a) Using Markov Models for Web Site Link Prediction. In Proc. of the Thirteenth ACM conference on Hypertext and Hypermedia (Hypertext'02), pp. 169-170, ACM Press, College Park, MD, USA, June 11-15. [17]

Jianhan Zhu, Jun Hong, and John G. Hughes (2002b) Using Markov Chains for Link Prediction in Adaptive Web Sites. In Proc. of Soft-Ware 2002: First International Conference on Computing in an Imperfect World, pp. 60-73, Lecture Notes in Computer Science, Springer, Belfast, April. [26]

Jianhan Zhu, Jun Hong, and John G. Hughes (2001) PageRate: Counting Web users' votes. In Proc. of the Twelfth ACM conference on Hypertext and Hypermedia (Hypertext'01), pp. 131-132, ACM Press, Århus, Denmark, August. [10]

Jianhan Zhu (2001) Using Markov Chains for Structural Link Prediction in Adaptive Web Sites. In Proc. of 8th Int. Conf. on User Modeling. LNCS, pp. 298-300, Sonthofen, Germany, July 13-17. [3]

**Not Refereed
Conference
Papers**

Jianhan Zhu, Dawei Song, Stefan Rueger, Marc Eisenstadt, Enrico Motta (2007d) Open University at TREC 2006 Enterprise Track Expert Search Task. In Proc. of the Fifteenth Text REtrieval Conference (TREC 2006), Gaithersburg, USA. (Invited for presentation in TREC2006) [2]

Jianhan Zhu, Dawei Song, Stefan Rueger (2007e) Open University at TREC 2007 Enterprise Track Document Search and Expert Search Tasks. To appear in Proc. of the Sixteenth Text REtrieval Conference (TREC 2007), Gaithersburg, USA. (Invited for presentation in TREC2007)

Jianhan Zhu, Dawei Song, Stefan Rueger (2007f) Open University at INEX 2007 Entity Ranking Track. To appear in Proc. of the INitiative for the Evaluation of XML Retrieval (INEX) 2007 Workshop, Dagstuhl, Germany.

Jianhan Zhu, Victoria Uren, and Enrico Motta. ESpotter: A Prototype System for Adaptive Named Entity Recognition Supporting Web Browsing. In Proc. of the Fifteenth ACM Conference on Hypertext and Hypermedia (Hypertext'04) Demo, Santa Cruz, USA, August 9-13, 2004.

Jianhan Zhu, Victoria Uren, and Enrico Motta. ESpotter: A Prototype System for Adaptive Named Entity Recognition Supporting Web Browsing. In Proc. of the Fourteenth International Conference on Knowledge Engineering and Knowledge Management (EKAW'2004) Demo, Whittlebury Hall, Northamptonshire, UK, October 5-8, 2004.

Other Papers

PhD dissertation: Mining Web site link structures for adaptive Web site navigation and search [6] Supervisors: Dr. Jun Hong and Prof. John G. Hughes

**Submitted
papers**

Jianhan Zhu, Dawei Song, Stefan Rueger (2007) Integrating Document Features for Expert Search.