

Haifeng Liu

Mailing Address:
10 King's College Road
Toronto, ON
Canada, M5S 3G4

Phone: (416)731-1491
Fax: (416)978-4765
hfliu@cs.toronto.edu
<http://www.cs.toronto.edu/~hfliu>

EDUCATION

Ph.D. Computer Science University of Toronto, 2008 (expected)
Thesis: Managing Uncertain, Imprecise, and Semantic Data in Publish/Subscribe Systems
Advisor: Hans-Arno Jacobsen

M.Sc. Computer Science University of Toronto, 2003 (GPA 3.9/4.0)
Thesis: A Publish/Subscribe System Model Supporting Uncertainties
Advisor: Hans-Arno Jacobsen

B.Sc. Computer Science University of Science and Technology of China, 2001 (GPA 88/100)

AWARDS AND SCHOLARSHIPS

Best Paper Award in the Web Intelligence conference (343 submissions, 58 accepted (17%) and 1 best paper)	2007
Monica Ryckman Bursary, Computer Science Department Awards, University of Toronto	2006
Best Paper Finalist in the WWW conference (550 submissions, 77 accepted (14%) and 4 best paper finalists)	2005
Best Paper Runners-up in the WWW conference (504 submissions, 74 accepted (15%), 7 best paper finalist)	2004
University Fellowship, University of Toronto	2001-2005
Meritorious honor (First Prize) in COMAP Mathematical Contest in Modeling, USA (496 participated, 77 meritorious winner (15%))	2000
Outstanding Student Scholarship, University of Science and Technology of China	1996-2001

PUBLICATIONS

Journal Papers

- [1] Haifeng Liu, Milenko Petrovic, Hans-Arno Jacobsen, "Efficient and Scalable Filtering of Graph-based Metadata", *Web Semantics* 3(4): 294-310, December 2005.
- [2] Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma, "Learning important models for web page blocks based on layout and content analysis", *SIGKDD Explorations* 6(2): 14-23, 2004.

Submitted to Peer-Reviewed Journal

- [1] Hyun Chul Lee, Haifeng Liu, Renee Miller, "Geographically-Sensitive Link Analysis", *submitted to peer-reviewed journal: IEEE Trans. Knowledge and Data Engineering*.
- [2] Haifeng Liu and Hans-Arno Jacobsen, "Adaptive Publish/Subscribe System for Modeling Uncertainties", *submitted to peer-reviewed journal: ACM Transactions on Computer Systems*.

Book Chapter

- [1] Haifeng Liu and Hans-Arno Jacobsen "Object-Oriented Publish/Subscribe for Modeling and Processing of Uncertain Information", *Chapter in Advances in Fuzzy Object-Oriented Databases: Modeling and Application*, p. 301-332, IDEA GROUP PUBLISHING, 2005.

Refereed Conference Papers

- [1] Hyun Chul Lee, Haifeng Liu, Renee Miller, “Geographically-Sensitive Link Analysis”, In *Proc. the 2007 IEEE/WIC/ACM Int. Conf. on Web Intelligence (WI 2007)*, (**Best Paper Award!**, 17% acceptance rate)
- [2] Milenko Petrovic, Haifeng Liu, H-Arno Jacobsen, “CMS-ToPSS: Efficient Dissemination of RSS Documents”, In *Proc. IEEE Int. Conf. on Very Large Data Bases (VLDB), Demonstration*, p. 1279-1282, 2005.
- [3] Milenko Petrovic, Haifeng Liu, H-Arno Jacobsen, “G-ToPSS: Fast Filtering of Graph-based Metadata”, in *Proc. 14th Int. World Wide Web Conf. (WWW05)*, p. 539-547, 2005. (**Best Paper Finalist!**, 14% acceptance rate)
- [4] Xiaofei He, Deng Cai, Haifeng Liu and Jiawei Han, “Image Clustering with Tensor Representation”, in *Proc. ACM Multimedia*, p. 132-140, 2005.
- [5] Haifeng Liu, H-Arno Jacobsen, “A-TOPSS – A Publish/Subscribe System Supporting Imperfect Information Processing”, in *Proc. IEEE Int. Conf. on Very Large Data Bases (VLDB), Demonstration*, p. 1281-1284, 2004.
- [6] Haifeng Liu, H-Arno Jacobsen, “Modeling uncertainties in Publish/Subscribe System”, in *Proc. 20th International Conference on Data Engineering (ICDE04)*, p. 539-547, 2004. (14% acceptance rate)
- [7] Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma, “Learning Block Importance Models for Web Pages”, in *Proc. 13th Int. World Wide Web Conf. (WWW04)*, p. 203-211, 2004. (**Best Paper Finalist!**, 15% acceptance rate)
- [8] Xiaofei He, Deng Cai, Haifeng Liu and Wei-Ying Ma, “Locality Preserving Indexing for Document Representation”, in *Proc. 27th Annual International ACM SIGIR Conference (SIGIR’2004)*, p. 96-103, 2004.
- [9] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Haifeng Liu, Hong-Jiang Zhang, “Spectral Analysis for Face Recognition”, in *Proc. Asian Conference on Computer Vision (ACCV04)*, 2004.
- [10] Haifeng Liu, H-Arno Jacobsen, “Approximate Matching in Publish/Subscribe”, in *Proc. 5th IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 2003.
- [11] Haifeng Liu, H-Arno Jacobsen, “A-TOPSS – A Publish/Subscribe System Supporting Approximate Matching”, in *Proc. 28th International Conference on Very Large Data Bases (VLDB2002), Demonstration*, p. 1107-1110, 2002.

Submitted to Peer-Reviewed Conferences

- [1] Haifeng Liu, Milenko Petrovic, H-Arno Jacobsen, “Efficient Filtering of RSS Documents on Computing Cluster”, *submitted to peer-reviewed conference: World Wide Web Conf (WWW08)*, 2008.

Technical Report

- [1] Hyun Chul Lee, Haifeng Liu, Renee Miller, “Geographically-Sensitive Link Analysis”, *Technical Report CSRG-544, Department of Computer Science, University of Toronto*, November 2006.

PATENT

Patent Method and System for Calculating Importance of a Block within a Display Page. Filed, May 2004.

INVITED TALKS

1. Research Directions in Publish/Subscribe, IBM China Research Lab, Beijing, China, 2003.
2. Modeling Uncertainties in Publish/Subscribe System, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 2003.

RESEARCH EXPERIENCE

Doctoral Research

Sep. 2003 – present

Middleware System Research Group, University of Toronto

- 1) Developed a Graph-based model for the Toronto Publish/Subscribe System (G-ToPSS) for dissemination of RSS documents that allows the use of ontologies as a way to provide additional information about the data.
- 2) Developed a cluster-based publish/subscribe system, propose two indexing algorithms that leverage the relationship among subscriptions and matching statistics to distribute subscriptions, and develop a pipelined filtering algorithm for graph-based data in a computing cluster environment.
- 3) Study the issue of ranking web pages using geographically-sensitive link analysis algorithms to improve search performance. Propose several alternative graph models capturing the semantics of the geographic-entity-to-page and the page-to-page relations and propose geographically-oriented link analysis algorithms for each graph model.

Master Research

Sep. 2001 – June 2003

Middleware System Research Group, University of Toronto

- 1) Research on approximate matching in publish/subscribe system: proposed a language model using fuzzy set and possibility theory to represent vague query language terms and approximate matching, designed and implemented matching algorithms.
- 2) Conducted the implementations of a real-world publish/subscribe system: the deployment of A-TOPSS as an information dissemination service collocated to an apartment rental market web server.

Summer Intern

Summer 2003

Microsoft Research Asia (MSRA), Beijing, China

Mentor: Jirong Wen, Wei-Ying Ma

- 1) Web Search: Learning block importance models based on block spatial features for web pages to improve search performance (**shown to Bill Gates!**)
- 2) Face Recognition: Spectral analysis for face recognition
- 3) Information Retrieval: Locality Preserving Indexing (LPI) for document indexing that discovers the local structure and obtains a compact document representation subspace that best detects the essential semantic structure.

Research Assistant

Summer 1999, Summer 2000

National High Performance Computing Center, Univ. of Sci. and Tech. of China

- 1) Developed a MIS for merchandize management, responsible for database management and interface programming (PowerBuilder)
- 2) Developed a distant education system which integrated teaching and learning on the web.
- 3) Research on inexact string alignment algorithm for DNA analysis and implemented TSP algorithm which is applied to solve the DNA mapping and sequencing problem.

PROFESSIONAL ACTIVITIES

Technical Reviewer

Sep. 2001 – present

IEEE Transactions on Knowledge and Data Engineering

Very Large Data Bases Conferences
IEEE International Conference on Distributed Computing Systems
International Middleware Conference
International Workshop on Software Engineering and Middleware

TEACHING EXPERIENCE

Teaching Assistant

APS105, Computer Fundamental	2002, 2003, 2005, 2006
Supervised laboratory sessions, marked assignments and exam papers.	
CSC309, Programming on the Web	2004, 2005, 2006
Conducted tutorials, marked assignments and exam papers.	
ECE344, Operating Systems	2002, 2003, 2004, 2005
Supervised laboratory sessions, marked assignments and exam papers.	
CSC148, Introduction to Computer Science	2002, 2003, 2004, 2005
Conducted weekly tutorials, marked assignments and exam papers.	
ECE369, Operating Systems	Summer 2004
Marked assignments and exam papers.	
CSCA48, Introduction to Computer Programming	Winter 2003
Conducted weekly tutorials, supervised laboratory sessions, marked assignments and exam papers.	
CSC326, Programming Language	Fall 2001
Supervised laboratory sessions, marked assignments and exam papers.	

Undergraduate Projects Supervision

1. CMS-ToPSS: Efficient Dissemination of RSS Documents	Summer 2005
Jerry Shen: NSERC Summer project	
2. Self-tuning Publish/Subscribe System for Modeling Uncertainties	Summer 2004
Zoe Chow: Computer Engineering Summer Project	
3. Quick Update on Data Gathering for Pub/Sub Workload Approximation	Summer 2003
Jason Jianye Lin and Francisca Chrysostom, Computer Engineering Summer Project	
4. A real-world Publish/Subscribe System Supporting Approximate Matching	Summer 2002
Enping Tu and Leo Zhanhong Luo: NSERC Summer project	

SKILLS

Languages: C/C++, Java, XML, HTML, Lisp and Meta HTML

Database System: database programming with SQL, MS SQL server, MySQL, JDBC and ODBC

Web/Network: ASP, JavaScript and socket programming, network technologies such as TCP/IP, Client/Server, XML and XML document filtering, content-based routing in the distributed systems using advertise/publish/subscribe.

Operating Systems: Unix (Solaris), Linux, GCC, GDB, Makefile, CVS, Windows (9x/NT/XP), Matlab,

Others: Research skills: modeling, analytical and quantitative skills Work skills: both cooperation and independent Communication: fluent in English speaking and writing.