Ka Cheung "Richard" SIA *kcsia@cs.ucla.edu*

11140 Rose Ave. APT 103, Los Angeles, CA 90034
(310)948-2935, http://oak.cs.ucla.edu/∼sia

Objective

I am looking for a a full-time position involved in research and development in the area of Web data management, analysis and datamining.

Education

University of California, Los Angeles
PhD Candidate in Computer Science, 2003 - Present, expect to graduate in Fall 2008
Tentative Thesis Title: Efficient Monitoring, Management and Analysis of Dynamic Web Content
Advisor: Prof. Junghoo Cho (cho@cs.ucla.edu)

The Chinese University of Hong Kong
M.Phil. in Computer Science, 2003
Thesis: Content-Based Image Retrieval: Reading One's Mind and Helping People Share

The Chinese University of Hong Kong
B.Eng. in Information Engineering (first class honor), 2001.

Research and Industry Experience

Software Engineering Intern **Google Inc.**
Mountain View, CA **Jun. 2008 – Sep. 2008**
I work in the index metrics team to build a tool that monitor and analyze competitor's performance based on crawl volume and search engine referrer counts statistics.
Manager: Bensong Chen (bensong@google.com)

Graduate Student Researcher **UCLA**
Department of Computer Science, UCLA **Jul. 2004 – Jun. 2008**
During my PhD study, I've worked on research problems centered around a NSF funded project "Blogocenter". It includes: Weblog parser; efficient scheduling policy for Weblog data collection; opinion extraction and sentiment analysis; efficient personal aggregate queries on blogs; analysis of word evolution in social tagging system. Some other projects that I've worked on includes: efficient query answering in sensors network; analysis of collaborative document authoring; leveraging bittorrent network for DDoS attack. I've also guided several students to complete their graduation projects for Master degree.
Advisor: Professor Junghoo Cho (cho@cs.ucla.edu)

Research Intern **NEC Labs America, Cupertino, CA**
Cupertino, CA **Jun. 2007 - Sep. 2007**
I worked on research problems in speeding up personalized aggregate queries over blog data, the proposed method reduces query time by 70% with 70% approximation accuracy on average. Results of this study was published in an academic conference. I've also added English support to an opinion extraction system built by NEC researchers in Japan, and participated in the implementation of a blog visualization tool.
Manager: Belle L. Tseng (belle@yahoo-inc.com)

Research Intern **NEC Labs America, Cupertino, CA**
Cupertino, CA **Jun. 2006 - Sep. 2006**
I implemented a Firefox browser plugin as a personal information manager prototype. I worked on research problems in user-based RSS feeds monitoring and user interest identification and adaptation which resulted in two publications in academic conference. I've also built an annotation tool for team members in preparing spam blog training data.
Manager: Belle L. Tseng (belle@yahoo-inc.com)

Summer Intern **Amgen Inc., Thousand Oaks, CA**
Research Informatics Department **Jun. 2005 - Sep. 2005**
I have worked on three projects including: applying datamining techniques in patient enrollment statistics forecast; bioinformatics studies of protein molecules in bone-remodeling process; development of a J2EE web application in struts framework for managing internal research reports.
Manager: Kailash Swarna (kswarna@amgen.com)

Graduate Student Researcher **CUHK, Hong Kong**
Multimedia Information Processing Lab **Sep. 2001 – Aug. 2003**
Research, design, and implementation of the project DISCOVIR (DIStributed COntent-based Visual Information Retrieval). It built on an open source software, LimeWire, and added the capability of content-based image retrieval over the Gnutella P2P network. The project resulted in several publications in academic conference and journal.
http://www.cse.cuhk.edu.hk/~miplab/discovir.
Advisor: Professor Irwin King (king@cse.cuhk.edu.hk)

Selected Publications

Journal articles

Ka Cheung Sia, Junghoo Cho, and Hyun-Kyu Cho, "Efficient Monitoring Algorithm for Fast News Alert", in IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 19, no.7, pp. 950-961, July 2007.

Irwin King, Cheuk Hang Ng, and **Ka Cheung Sia**, "Distributed Content-Based Visual Information Retrieval on Peer-to-Peer Networks", in ACM Transactions on Information Systems (TOIS), Vol. 22, Issue 3, July 2004, pages 477-501.

Conference papers

Ka Cheung Sia, Junghoo Cho, Yun Chi, and Belle L. Tseng, "Computing Personal Aggregate Queries Efficiently on Blogs", in the 14th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2008.

Ka Cheung Sia, Junghoo Cho, Koji Hino, Yun Chi, Shenghuo Zhu, and Belle L. Tseng, "RSS Feeds Monitoring Driven by User Browsing Pattern", in Proceedings of the First International Conference on Weblog and Social Media (ICWSM), 2007.

Ka Cheung Sia, Shenghuo Zhu, Yun Chi, Koji Hino, and Belle L. Tseng, "Capturing User Interests by Both Exploitation and Exploration", in Proceedings of the 11th International Conference on User Modelling (UM), 2007.

Yu-Ru Lin, Wen-Yen Chen, Xiaolin Shi, **Richard Sia**, Xiaodan Song, Yun Chi, Koji Hino, Hari Sundaram, Jun Tatemuran, and Belle L. Tseng, "The Splog Detection Task and A Solution Based on Temporal and Link Properties", in Proceedings of the 15th Text Retrieval Conference (TREC), 2006.

Zhenyu Liu, **Ka Cheung Sia**, and Junghoo Cho, "Cost-Efficient Processing of Min/Max Queries over Distributed Sensors with Uncertainty", in Proceedings of the 20th ACM Symposium on Applied Computing (SAC), 2005.

Cheuk Hang Ng, **Ka Cheung Sia**, and Chi Hang Chan, "Advanced Peer Clustering and Firework Query Model in the Peer-to-Peer Network", in Poster Proceedings of the 12th World Wide Web Conference (WWW), 2003.

Cheuk-Hang Ng and **Ka-Cheung Sia**, "Peer Clustering and Firework Query Model", in Poster

Proceedings of the 11th World Wide Web Conference (WWW), 2002.

K.C.Sia and Irwin King, “Relevance Feedback Based on Parameter Estimation of Target Distribution”, in Proceedings of International Joint Conference on Neural Networks (IJCNN) ,2002.

Technical Reports

Ka Cheung Sia, Susan Chebotariov, and Junghoo Cho, “Theory of Mergeability: Safely Merging Two Versions of a Document”, June 2007

Ka Cheung Sia, “DDoS Vulnerability Analysis of the Bittorrent Protocol”, June 2006

Computing Skills

Proficient in Java, intermediate in C/C++ and Python
Basic Perl, PHP, and Javascript
Proficient in Matlab and MySQL administration and tunning
Linux system administration and shell scripting

Honors and Awards

Placed top two in three consecutive UCSD Datamining competitions in 2006, 2007, and 2008 (http://mill.ucsd.edu)

Student travel award for ICWSM, UM, and SIGKDD conference.

Qualifications

MCP, MCP+I, MCSE, Sun Certified Java Programmer (SCJP)

Professional Activities

I have been the external reviewer of several academic conferences including SIGMOD, VLDB, WWW, CIKM, ER and the WebDB workshop.

References

Prof. Junghoo Cho, *cho@cs.ucla.edu*
Department of Computer Science, UCLA

Prof. Carlo Zaniolo, *zaniolo@cs.ucla.edu*
Department of Computer Science, UCLA

Dr. Belle L. Tseng, *belle@yahoo-inc.com*
International Search Ranking Team, Yahoo! Research

Dr. Bensong Chen, *bensong@google.com*
Index Metrics Team, Google Inc.