

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

**WORK EXPERIENCE** *Software Engineer* **Google Inc., Kirkland, WA**  
Kirkland, WA **May. 2012 – Present**  
I work in the demographics team to infer user demographics based on user's behavior on the Web and the mobile space and the goal is to improve ad quality in Google's display network. I also worked in the doubleclick search team to improve reporting experience. Selected projects include: integrating Adwords and Google Analytics data in reporting, adding landing page testing support on new conversion datasource, created and managed the demo version of the product used in sales pitch.

*Software Development Engineer* **Bing, Microsoft Corp, Redmond, WA**  
Redmond, WA **Feb. 2009 – May. 2012**  
I work in the ranking and intent team in Bing on projects that improve the relevance of Bing search result. Selected projects include: finding conepts and entities in queries; identifying visually important phrases from html pages; matching query intent and document category; top N result reranking by CTR and authority signal; identifying important words in query.

*Software Engineering Intern* **Google Inc, Mountain View, CA**  
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, Los Angeles, CA**  
Department of Computer Science, UCLA **Jul. 2004 – Jun. 2008**  
During my PhD study, I've worked on research problems centered around the 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, Cupertino, CA**  
Cupertino, CA **Jun. 2007 - Sep. 2007**  
I focus on a research problem that speed up personalized aggregate queries over blog data, the proposed method reduces query time by 70% with 70% approximation accuracy on average. Results was published in an academic conference. I've also added English support to an opinion extraction system built by NEC researchers in Japan, and involved in the implementation of a blog visualization tool.  
Manager: Belle L. Tseng (belle@yahoo-inc.com)

*Research Intern* **NEC Labs, 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 to collect 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 using struts framework for managing internal research reports.

Manager: Kailash Swarna (kswarna@amgen.com)

**COMPUTING SKILLS** Proficient in C++, intermediate in C#, Java, and Python  
Familiar with mining data in large-scale distributed system using map-reduce paradigm  
Basic Perl, PHP, and Javascript  
Proficient in Matlab and MySQL administration and tuning  
linux system administration and shell scripting

**EDUCATION** University of California, Los Angeles  
PhD in Computer Science, 2009  
Dissertation: Challenges and Opportunities in Building Personalized Online Content Aggregators

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.

**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", to appear 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 and **Ka-Cheung Sia**, “Peer Clustering and Firework Query Model”, in Poster Proceedings of the 11th World Wide Web Conference (WWW), 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

### PATENTS

Search Results By Mapping Associated With Disparate Taxonomies, US 2013/0238608 A1, filed 2012.

Using Part-of-Speech Tagging to Improve Query Alteration and Ranking, filed 2013.

### HONORS AND AWARDS

Microsoft Research Technology Transfer Award, 2011

Microsoft Gold Star Award, 2010

Placed top two in three consecutive UCSD Datamining competitions from 2006 - 2008

### PROFESSIONAL ACTIVITIES

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

### REFERENCES

Dr. Steven Yao, *qiyao@microsoft.com*  
Partner Dev. Manager, Bing, Microsoft

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

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