

## OBJECTIVE

Seeking a software development/research position in information technology.

## STRENGTHS

- Consolidated theoretical background and research experience in the fields of **web search, information retrieval, data mining** and **machine learning**.
- Seven years of hands-on experience in developing major **search engine technology**, such as **ranking algorithms, query log analysis**, and **caching algorithms**;
- Rich experience in analyzing massive and high-dimensional scientific and internet data sets.
- Extensive knowledge of **object-oriented** analysis and design.
- Self-motivated and excellent independent problem solving skills.

## COMPUTER SKILLS

- **Programming Languages:** proficient in C++, C, Perl, C#, shell scripting; good knowledge of Java, HTML, XML, Python, and ActiveX.
- **Databases:** familiar with several flavors of relational DBMS, such as Berkeley DB, MySQL, SQL Server.
- **Software and Tools:** MatLab 7.0, Weka 3.4.4, Maple, NUnit, and Visual Studio 2008.
- **Operating Systems:** Solaris 9, Linux 2.6, and Windows 2003/2008 Server.

## EDUCATION

- **Ph.D.** in Computer Information Science, Polytechnic Institute of NYU, Brooklyn, NY, 2008
  - Thesis: "Mining the Web to Improve Search Engine Performance"
- **B. Sc.** in Computer Science and Technology, Beijing Jiao tong University, Beijing, China, 2000

## PROFESSIONAL EXPERIENCE

**Software Development Engineer** (contractor), Microsoft Corp., Redmond, WA (03/2008-present)

- Worked in the CRM&BCM Outlook Client team. Integrated CRM and BCM functionalities into the latest version of Outlook while keeping the backwards compatibility. (**C#, C++, .Net Remoting, MS SQL Server, IIS, NUnit, and Windows 2008 Server**)
  - Remodeled the prototype for Xml-based ribbon customization, resulting in substantially increased productivity and reduced redundancy.
  - Cooperated with web application team and successfully managed the message exchange between CRM server and Outlook client.
  - Applied .Net Remoting to implement a client-server infrastructure which enables NUnit to run unit testing for Outlook Add-Ins, which requires Outlook running in the backend process.

**Research Assistant**, Dept. of Computer Science, Polytechnic Institute of NYU, Brooklyn, NY (09/2001-03/2008)

- Proposed and evaluated improved techniques for **query result caching** in the purpose of optimizing query processing performance in modern web search engines. (**C++, MatLab, Linux**)
  - Described and developed several algorithms for weighted result caching which targets at maximizing cost savings instead of hit ratio. Analyzed and formulated the impact of Zipf-based query distributions in weighted result caching performance.

- Proposed a new set of feature-based cache eviction policies that achieved very significant improvements over previous work on result caching in the literature. The best method on average outperforms common hybrid algorithms by about 4%.
- Investigated **link-based ranking** techniques in order to provide more accurate search results. Derived new I/O efficient algorithms for PageRank based on techniques proposed for out-of-core **graph algorithms**. The experiments were performed based on several web crawls of a few hundred million pages each. The new algorithms resulted in notable improvements over previous approaches in terms of reducing the I/O cost. **(C, Linux)**
- Designed a two-stage approach to improve the performance of **spam detection** techniques. First, implemented a machine learning-based classifier using both C4.5 and SVM. Then by integrating the knowledge about the neighborhood, designed several heuristics to decide if a node should be relabeled based on the pre-classified result. Experimental results showed significant improvements with respect to both precision and recall, with overall accuracy of 90%. **(Perl, C++, Linux)**
- Conducted a detailed survey of **geographic search queries** and proposed a new taxonomy for such queries. Built a C4.5-based geographic classifier that detects queries with geographic intents. Identified opportunities for improving geographical search related technologies, based on an analysis of 36 million real user queries with click through data. **(Perl, Weka, Linux)**

**Software Engineer Intern**, IPValue Management Inc., Bridgewater, NJ (06/2006-08/2006)

- Acted as a subject matter expert to support technical staff and attorneys during assertion opportunities.
- Conducted detailed investigation into specific technologies in support of licensing and technology transfer operations.
- Performed evidence gathering, assisted in claim charting, conducted prior art searches and reverse engineered for promising infringing products.

**Instructor**, Training Center of China National Software & Service Company Ltd., Beijing, China (07/2000-08/2001)

- Taught MCSE and SCJP for diverse customers, including major corporations, small e-business ventures, and government agencies.
- Key team member in development of an e-learning technology that helped the training center as a pioneer generating the highest student satisfaction rate in this area.

**Software Engineer Intern**, Information System Center of China National Software & Service Company Ltd., Beijing, China (02/2000-06/2000)

- Created an internal web site which helps to keep track of employee data, utilizing Java servlet and JDBC.

## SELECT PUBLICATIONS

- **Qingqing Gan**, Torsten Suel “Improved Techniques for Result Caching in Web Search Engines”, to appear in Proceedings of the 18th International World Wide Web Conference (**WWW**), Apr. 2009.
- **Qingqing Gan**, Josh Attenberg, Alexander Markowetz, Torsten Suel “Analysis of Geographic Queries in a Search Engine Log”, in Proceedings of the 1st International Workshop on Location and the Web (in conjunction with **WWW**), Apr. 2008.
- **Qingqing Gan**, Torsten Suel “Improving Web Spam Classifiers Using Link Structure”, in Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web (in conjunction with **WWW**), May 2007.
- Yen-Yu Chen, **Qingqing Gan**, Torsten Suel “Local Methods for Estimating PageRank Values”, in Proceedings of the ACM 13th Conference on Information and Knowledge Management (**CIKM**), Nov. 2004.
- Yen-Yu Chen, **Qingqing Gan**, Torsten Suel “I/O Efficient Techniques for Computing PageRank” in Proceedings of the ACM 11th Conference on Information and Knowledge Management (**CIKM**), Nov. 2002.