Jinru He
#1466 67th Street
Brooklyn, NY, 11219

513-907-7905
mikehe1117@gmail.com

Education

- **New York University, Polytechnic Institute**
  PhD Candidate, Computer Science, *Sep. 2007 - Present*
  - Relevant courses: Web Search Engine, Advanced Database System, Machine Learning, Information Theory, Distributed System, Advanced Algorithms
- **Huazhong University of Science and Technology**
  Bachelor of Science in Software Engineering *Sep. 2003- Jun. 2007*
  - Graduated with Honors.
  - Relevant courses: Computer Architecture, Data Structure, Introduction to Computer Compiler, Computer Networks, Operating System

Work Experience

- **Core Search, Bing, Microsoft**
  Research Intern *Jun. 2011 - Aug.2011*
  - Worked as Research intern with Dr. Trishul Chilimbi and Maguro team on the next generation of index structure inside Bing search.
- **New York University, Polytechnic Institute**
  Research Assistant *Sep. 2007 - Present*
  - Working with Professor Torsten Suel on the inverted index structures in Information Retrieval systems, and high-performance query processing.
- **Huazhong University of Science and Technology**
  Research Assitant *Sep. 2006 - Jun. 2007*
  - Worked with Professor Changqing Chen on OLAP applications in bank queueing systems. The final project was runner-up at the Second Citigroup Software Competition.

Publications

- Text vs. Space: Efficient Geo-Search Query Processing, with Maria Christoforaki, Constantinos Dimopoulos, Torsten Suel and Alex Markowetz, 20th ACM Conference on Information and Knowledge Management (CIKM 2011) (15% accepted).
- Faster Temporal Range Queries over Versioned Text, with T. Suel, 34th Annual International ACM SIGIR Conference (SIGIR 2011) (19.8% accepted)
- Improved Index Compression Techniques for Versioned Document Collections, with J. Zeng, T. Suel, 19th ACM Conference on Information and Knowledge Management (CIKM 2010) (13.4% accepted).

- Compact Full-Text Indexing of Versioned Document Collections, with H. Yan, T. Suel, 18th ACM Conference on Information and Knowledge Management (CIKM 2009) (14.5% accepted)
- Using Graphics Processors for High Performance IR Query Processing, with S. Ding, H. Yan, T. Suel, 18th International Conference on World Wide Web (WWW 2009) (11.8% accepted)

Selected Research Projects

- **Text vs. Space: Efficient Geo-Search Query Processing** Nov. 2010 - Jun. 2011

  An number of search services allow users to constrain text queries (e.g., photography classes) to a geographic location (e.g., Santa Monica). This includes local search engines such as Google Local, mobile search services accessible from smart phones. This motivates the problem of how to efficiently execute search queries that contain a mix of textual and spatial constraints

  In this project, we take a new look at this problem. Executing such queries requires a combination of textual (e.g., inverted lists) and spatial (e.g., R-trees, space-filling curves) data structures. We describe several existing and new algorithms that make different choices on this trade-off between text and space, and evaluate them on large data sets. Our results indicate that an efficient query processor needs to first get the textual aspects of the problem right. In fact, even a naive approach that applies spatial filtering at the end appears to outperform many previous schemes, while even better results are obtained by integrating some light-weight spatial structure into the inverted index design. (C++)

- **Faster Temporal Range Queries over Versioned Text** Jun. 2010 - Nov. 2010

  Versioned textual collections are collections that retain multi- ple versions of a document as it evolves over time. Important large-scale examples are Wikipedia and the web collection of the Internet Archive. Search queries over such collections of- ten use keywords as well as temporal constraints, most com- monly a time range of interest. In this paper, we study how to support such temporal range queries over versioned text. Our goal is to process these queries faster than the corre- sponding keyword-only queries, by exploiting the additional constraint. A simple approach might partition the index into different time ranges, and then access only the relevant parts. However, specialized inverted index compression techniques are crucial for large versioned collections, and a naive partitioning can negatively affect index size and query through- put. We show how to achieve high query throughput by using smart index partitioning techniques that take index compres- sion into account. Experiments on over 85 million versions of Wikipedia articles show that queries can be executed in a few milliseconds on memory-based index structures, and only slightly more time on disk-based structures. We also show how to efficiently support the recently proposed stable top-k search primitive on top of our schemes.(C++)

- **Compact Full-Text Indexing of Versioned Document Collections** Jan.2009 - Jun.2010

  Current Information Retrieval systems use inverted index structures for efficient query processing. Due to the extremely large size of many data sets, these index structures are usually kept in compressed form, and many techniques for optimizing compressed size and speed have been proposed. In this work, we focus on the case of versioned document collections, that is, collections where each document is modified over time, resulting in multiple versions of the document. Consecutive versions of the same document are often very similar, and several researchers have explored ideas for exploiting this similarity to decrease the overall index size.

  We propose new index compression techniques for versioned document collections that achieve significant reductions in index size over previous methods. In particular, we first propose several bitwise compression techniques that achieve very compact index structure but that are too slow for most applications. Based on the lessons learned, we then propose additional techniques that come close to the sizes of the bitwise technique while also approximately matching the speed of the best previous methods. (C++)

- **Using Graphics Processors for IR Query Processing** Sep.2008 - Oct.2009

  Web search engines are facing formidable performance challenges due to data sizes and query loads. The major engines have to process tens of thousands of queries per second over tens of billions of documents. To deal with this heavy workload, such engines employ massively parallel systems consisting of thousands of machines. The significant cost of operating these systems has motivated a lot of recent research into more efficient query processing mechanisms. We investigate a new way to build such high performance IR systems using graphical processing units (GPUs). GPUs were originally designed to accelerate computer graphics applications through massive on-chip parallelism. Recently a number of researchers have studied how to use GPUs for other problem domains such as databases and scientific computing. Our contribution here is to design a basic system architecture for GPU-based high-performance IR, to develop suitable algorithms for subtasks such as inverted list compression, list intersection, and top-k scoring, and to show how to achieve highly e?cient query processing on GPUbased systems. Our experimental results for a prototype GPU-based system on 25.2 million web pages shows promising gains in query throughput. (CUDA, C, Linux, GTS8800)

- **Hustonline FTP Search Engine** Sep.2006 - Jun.2007

  There are thousands of FTP sites full of useful files and data in CERNET(Chinese Education and Research NETworks). FTP Search Engine is one of the solutions to seek specific files on FTP sites in CERNET. I designed the architecture of the FTP search engine. I also implemented crawler, indexer and data storage. Up to now, the number of FTP sites collected by the FTP Search Engine has reached 6000, and nearly 25 million data have been stored. The website can be reached by http://fdigg.net. (C++, Windows)

Awards

- Second Citigroup Software Competition, Runner-up, 2006
- Microsoft Imagine Cup Software Competition, Honorable Mention, 2006
- Class Three of People's Scholarship, 2004

Skills

- Proficient in GPU (graphic processing unit) programming with CUDA.
- Extensive experience with object-oriented programming languages (C++, Java, C#).
- Familiar with several flavors of DBMS, such as Berkeley DB, MySQL, SQL Server.
- Familiar with Java, ASP.net, Javascript, PHP.
- Experienced in using Solaris, Linux, and Windows.