Yiqun "Eddie" Cao

PhD Candidate, Computer Science, UC Riverside

T 951-756-8986

3446 Kentucky St
Riverside, CA 92507

cao.yiqun@gmail.com
http://www.cs.ucr.edu/~ycao/

Objective

A computer science and computer engineer career, preferably in interdisciplinary areas of scientific computing, bioinformatics, and cheminformatics; to develop novel computational systems and approaches that better people's health and life.

Experience

Graduate Student Researcher, Bioinformatics Lab, University of California, Riverside — 2006-present

Performed database and machine learning study on chemical compound structure and activity data. Proposed new algorithms to process large scale structural/graph data and built implementations with C, C++, R, Python and Web Application technologies.

- Proposed and implemented a Maximum Common Substructure-based (MCS) algorithm for searching chemical structures and performing ligand-based virtual screening.
- Designed and implemented a general method to greatly accelerate chemical structural similarity search.
- Created a software package for calculating structural descriptors and similarities.
- Co-authored ChemMine, a web-based compound mining environment.
- Co-authored ChemmineR, a chemoinformatics framework using R programming environment.
- Authored the Bioactivity and Phenotype Database (BAPDB), a database for exploring the biological and molecular functions of genes based on available phenotype and screening data.
- Authored the EI website, a gateway for accessing tens of millions of PubChem compounds, which provides ultra-fast highly-accurate similarity search and high-quality clustering results.

Visiting Student, Shoichet's Lab, University of California, San Francisco — Jan. 2007-Apr. 2007

Conducted research on automated receptor-based virtual screening.

- Studied large-scale automated molecular docking and consensus scoring in molecular docking.
- Built a prototype database for automated docking runs using the whole PDB.
- Built systems to simplify the use of Protein Ligand Optimization Program (PLOP) and to facilitate single-mode DOCK run.

Graduate Student Researcher, National University of Singapore — Jan. 2004-Jul. 2005

Conducted research on distributed and parallel computing, with applications in computational biomedical applications.

Undergraduate Student Researcher, Tsinghua University, Beijing, China — Nov. 2001-Jun. 2003

Worked on embedded system and system and network programming.

- Worked on an embedded Bluetooth-based wireless communication system.
- Conducted research on a network-facilitated anti-virus system (similar to today's Microsoft SpyNet).

Freelance Author — Jun. 2001-Dec. 2001

Co-authored *Advanced PHP Programming* (in Chinese) (ISBN 7-302-05344-8).

Freelance Web Developer - Mar. 2005-Aug. 2005

Designed a website and CMS for Party World KTV Pte Ltd Singapore using PHP.

Publications

- **Yiqun Cao**, Tao Jiang, Thomas Girke. *A maximum common substructure-based algorithm for searching and predicting drug-like compounds*. ISMB 2008. Also appeared on Bioinformatics, 24(13). doi:10.1093/bioinformatics/btn186
- **Yiqun Cao**, Anna Charisi, Li-Chang Cheng, Tao Jiang, Thomas Girke. *ChemmineR: a compound mining framework in R*. Bioinformatics 24(15). doi:10.1093/bioinformatics/btn307
- John J. Irwin, Brian K. Shoichet, Michael M. Mysinger, Niu Huang, Francesco Colizzi, Pascal Wassam, **Yiqun Cao**. *Automated Docking Screens: A Feasibility Study*. J Med Chem 52(18). doi: 10.1021/jm9006966
- **Yiqun Cao**, Tao Jiang, Thomas Girke. *Accelerated Similarity Searching and Clustering of Large Compound Sets by Geometric Embedding and Locality Sensitive Hashing*. Bioinformatics 26(7). doi: 10.1093/bioinformatics/btq067

Education

- **Univ. of California, Riverside**, CA — PhD in Computer Science, 2010 (expected)
- **National Univ. of Singapore** — Master of Engineering in Electrical & Computer Engineering, 2005
- **Tsinghua Univ.**, Beijing, P.R.China — B.S in Electrical Engineering, 2003

Skills

Have experience in **parallel and concurrent programming, object-oriented programming and functional programming**. Have extensive skills in **client- and server-side web application development**. Have basic backgrounds in computer architecture, distributed system, operating system, artificial intelligence, data mining, machine learning, database, computer networks, network programming, programming language concepts, theory of computation, algorithm design and analysis, parallel algorithms, sequence analysis, discrete event simulation, and numerical analysis.

- **C/C++** 9+ years of programming experience in general problem solving, UNIX system programming, multithread and parallel programming, and mixed language programming.
- **Java** 2+ years of experience in web application contexts (Servlet, JSP, and Struts).
- **Python** 3+ years of experience, especially in web application development (CherryPy, Kid, Paste, Django, and Google App Engine).
- **Other Languages** R (S PLUS); XML; JavaScript, XHTML/HTML, CSS; MATLAB; SQL; AWK; LaTeX; Bash.
- **OS & Tools** GNU/Linux (6+ years), Mac OS X (3+ years); Vim, Autotools, gcc, gdb, Eclipse; OpenBabel, JOELib; DOCK, PLOP, DockBlaster.

Honors

- Integrative Graduate Education and Research Traineeship (IGERT) Fellowship at Center for for Plant Cell Biology (CEPCEB) of UCR — 2007-present
- Dean's Distinguished Fellowship Award at UCR — 2005-2007
- Graduate Research Scholarship, National University of Singapore — 2003-2005
- Academic Excellence Scholarship, Tsinghua University, Beijing, China — 2000-2002