

# DANIEL ROCCO

*rockdj@cc.gatech.edu*

499 Calhoun St. NW  
Atlanta, GA 30318

404.483.9841

<http://www.cc.gatech.edu/~rockdj>

---

## OBJECTIVE

To strive for quality in all facets of academic life and contribute to the excellence of a prestigious academic institution for the mutual inspiration of my students, my colleagues, and myself.

## EDUCATION

### **Georgia Institute of Technology**

*Fall 2000 to present*

Pursuing Ph.D. in computer science, specializing in database systems. Research interests in the application of database technology to the Internet, automatic Web service discovery and classification, ubiquitous and mobile computing, and applications of XML technology. Minor in management of technology.

Doctoral thesis topic: *On Efficient Discovery, Classification, and Transport of Information on the World Wide Web.*

### **Georgia Institute of Technology**

*Fall 1996 to Fall 1999*

Earned Bachelor of Science in computer science, with high honor. Specialization in database systems and networking.

## PUBLICATIONS

- Daniel Rocco, Ling Liu, and James Caverlee. *XPack: A High-performance Document Encoding System for the Web*. In preparation for conference submission, February 2004.
- David Buttler, Daniel Rocco, and Ling Liu. *Efficient Web Change Monitoring with Page Digest*. To appear, International Conference on the World Wide Web poster symposium, 2004.
- Anne H. H. Ngu, Daniel Rocco, Terence Critchlow, and David Buttler. *Automatic Discovery and Inferencing of Complex Bioinformatics Web Interfaces*. Lawrence Livermore National Laboratory Technical Report UCRL-JRNL-201611, 2003.
- James Caverlee, Ling Liu, and Daniel Rocco. *Discovering and Ranking Data Intensive Web Services: A Source-Biased Approach*. Georgia Institute of Technology CERCS Technical Report GIT-CERCS-03-26, 2003.
- Daniel Rocco and Terence Critchlow. *Automatic discovery and classification of bioinformatics Web sources*. *Bioinformatics*. 2003 Oct 12;19(15):1927-33.
- Daniel Rocco, David Buttler, and Ling Liu. *Page Digest for Large-Scale Web Services*. In Proc. IEEE Conference on E-Commerce, 2003.
- Ling Liu, David Buttler, Terence Critchlow, Wei Han, Henrique Paques, Calton Pu, and Daniel Rocco. *BioZoom: Exploiting Source-Capability Information for Integrated Access to Multiple Bioinformatics Data Sources*. In Proc. of 3rd IEEE Symposium on Bioinformatics and Bioengineering, 2003.

- David Buttler, Matthew Coleman, Terence Critchlow, Renato Fileto, Wei Han, Calton Pu, Daniel Rocco, and Li Xiong. *Querying Multiple Bioinformatics Information Sources: Can Semantic Web Research Help?* SIGMOD Record, Vol 31, No. 4, December 2002.
- Daniel Rocco and Terence Critchlow. *Discovery and Classification of Bioinformatics Web Services*. Lawrence Livermore National Laboratory Technical Report UCRL-JC-149963, 2002.

## TEACHING EXPERIENCE

**Georgia Institute of Technology**, Intel mentor program *Winter 2003 to Spring 2004*

The Intel mentor program is a collaboration between Georgia Tech's College of Computing, the College of Electrical and Computer Engineering, and Intel Corp. that is designed to encourage women and underrepresented minorities to pursue careers in the disciplines of computer science and computer and electrical engineering. As a mentor, I am responsible for directing a student in the completion of a semester-long project designed to teach her about a computing discipline of interest. My duties include supervising my student's progress on the project throughout the semester via weekly meetings, ensuring that she is succeeding academically, assisting with the preparation of an end-of-term poster, and suggesting appropriate resources for academic and social needs.

**Georgia Institute of Technology**, Teaching Assistant *Fall 1996 to Fall 2001*

I served as a teaching assistant in various capacities throughout my academic career for courses ranging from an introductory computer science course through a senior-level database course. My responsibilities included lecturing, project design and implementation, group and one-on-one meetings with students, and administration.

**Georgia Institute of Technology**, Peer Leader *Fall 1997 to Spring 1999*

The Freshman Experience program at Georgia Tech was designed to improve freshman retention rates through proactive involvement in student's lives. My role as a peer leader involved fostering the social and academic development of approximately 15 male students living together in a residence hall. I prepared group activities throughout the year, encouraged attendance at program-wide events, and monitored academic progress while living in the community. Georgia Tech's Freshman Experience program is highly successful and has been a model for similar programs across the country.

## WORK EXPERIENCE

**Lawrence Livermore National Laboratory**, Research Intern *Winter 2003*

Developed a Web data source locator and analyzer that utilizes an abstract description of a class of services to find service instances pertinent to the domain of interest. The first prototype implementation was able to automatically identify 66% of the molecular sequence Web sources in our test set with no false positives. Began construction on the wrapper generation and integration components.

**Lawrence Livermore National Laboratory**, Research Intern

*Summer 2002*

Undertook the construction of an abstract language for defining a class of Web services independent of the actual implementation of those services. Assisted in the development of a strategy for defining Web service classes, locating service instances, and integrating discovered services behind a unified query interface.

**IBM Corp.**, Summer Intern

*Summer 2001*

During my summer internship at IBM T.J. Watson, I participated in the revision process for the Local Positioning Profile, a Bluetooth protocol providing for position data communication between local clients. I assisted the editor in refining the specification by implementing the LPP protocol over a standard Bluetooth stack on an embedded hardware platform running a minimal Linux distribution.

**GeoStats, Inc.**, Research Developer

*Fall 2000*

Implemented several PalmOS applications, including a generic survey administration application used in a one year tourism study by the Alaska Department of Transportation. The survey application exploited the Palm platform's excellent form factor, battery life, and data processing capabilities to collect user input via a dynamic question interface. Several categories of questions could be processed and answers were analyzed in real-time to alter the user's path through the survey based on their previous responses.

**NEO, Inc.**, Application Developer

*Winter 2000 to Summer 2000*

I helped develop and maintain applications for Facility Management that included database integration with AutoCAD. I implemented a barcode inventory system for the Symbol SPT 1500 that tagged over 2000 furniture items in 3 days for a corporate relocation.

## RESEARCH PROJECTS

**Web Data Representation**

*Fall 2001 to present*

The Page Digest is a mechanism for efficient storage and processing of Web documents. The Page Digest design encourages a clean separation of the structural elements of Web documents from their content. Its encoding transformation is invertible without introducing significant additional cost or complexity to normal document parsing. Compared to using standard DOM implementations, our initial experimental results show that Page Digest encoding can provide an order of magnitude speedup when traversing a Web document or comparing two arbitrary Web documents.

We have examined the potential benefits of using Page Digest in other large-scale Web Services such as Web Search Software, Web Data Extraction Services, and Automatic Fragment Detection for Dynamic Content Caching. Our experimental results show that change detection using the Page Digest operates in linear time, offering 75% improvement in execution performance when compared with popular existing change detection and difference systems. In addition, the Page Digest format reduces the tag name redundancy found in Web documents, which provides up to a 50% reduction in the document size without employing data compression techniques.

## **Large-scale Data Access**

*Spring 2002 to present*

As part of Georgia Tech's collaboration with Lawrence Livermore National Laboratories, I am developing a Web crawler for bioinformatics sources. The crawler is part of the joint effort to provide a uniform query interface for online biological data repositories such as BLAST. The crawler's portion of this task is to search the Web for potential new sources, analyze these sources using a domain specific meta language description, and provide integration hints to other portions of the system.

## **Cystic Fibrosis Patient Tracking**

*Spring 2000*

I developed a database application used to track the progress of children with cystic fibrosis. The application tracks changes in lung capacity and nutritional statistics, allowing the medical staff to quickly determine a child's health status and rehabilitation progress, enabling them to adapt treatment to the child's needs.

## **PalmOS Security Survey**

*Fall 2000*

This work explored the security architecture of the Palm Operating System. We cite several security bulletins and show that the PalmOS is insecure at a fundamental level by examining several possible attacks that exploit the vulnerable portions of the platform's security features.

## **COMPUTER SKILLS**

Platforms: x86 (Windows, Linux), PalmOS, embedded Linux  
Languages: Java, Python, C/C++, shell scripting, Visual Basic, assembler  
Web standards: HTML, CSS, XML, XSLT, XML Schema  
Database systems: Microsoft Access, SQL Server, PostgreSQL, Oracle  
Voice interface: Dragon NaturallySpeaking, Python voice scripts

## **HONORS, AWARDS, AND ACTIVITIES**

Distributed Data Intensive Systems Lab Group Coordinator *Fall 2003 to Summer 2004*  
Graduate Student Senate, Computer Science Senator *Fall 2000 to Spring 2001*  
Faculty Honors *Fall 1998*  
Dean's List *Fall 1997 to Spring 1997, Winter 1998 and 1999*  
Matrix Resources Computer Science Scholarship *Spring 1999*  
Architecture Study Abroad in Rome *Summer 1998*  
Georgia Tech Chorale *Fall 1998 to Spring 2001*  
Georgia Tech Westminster Christian Fellowship *Fall 1998 to Spring 2001*