

CURRICULUM VITAE

Steven P. Crain

RESEARCH STATEMENT:

- We are interested in helping people use Web search to solve problems. Without doubt, modern Web search is very useful for many activities. However, new developments in Web search will greatly increase the class of problems that it benefits. In this statement, we will describe the research agenda we are pursuing and discuss how it will enable users to better leverage Web search to solve problems.
- Current Web search technology uses feature-based machine learning techniques. Two kinds of features are used: document features reflect the quality of a document independent of the search task; query-document features attempt to capture the relevance of each document to a particular query. Consider the document features (for example, the media type, and Page Rank)---their applicability changes if the user is looking for other people's opinions of a product (which tend to be short and have negligible Page Rank) instead of looking for factual information. Likewise, the significance of query-document features varies for different queries and search tasks. In traditional Web search, the query serves as proxy for the user and the search context (the role the search plays in problem solving). More formally, the relevance of a document is assumed to be independent of the user and the search context when the query is known. We are interested in more complex models that allow the search engine to incorporate its knowledge of the query, the user and the search context when ranking documents.

CURRENT RESEARCH:

- *Query dependent document ranking.* We are attempting to improve the ranking of query results by correctly modeling the impact to the ranking function based on differences in the queries.

RESEARCH PROJECTS:

- *Personality analysis by spectral methods*, 1992-1998.
- *Propagation in epidemic networks*, 2005.
- *Weblog sentiment analysis*, 2007.
- *Autistic behavior characterization*, 2007.
- *Nearly non-negative SVD for Non-Negative Matrix Factorization*, 2007.
- *From libraries to the Web: large scale hierarchical classification of book repositories*, 2007-2008.
- *Multi-view face detection using AdaBoost trees*, 2008.
- *Cache oblivious task queues*, 2008.

PUBLICATIONS:

- Contribution to Knuth, Art of Computer programming, vol I.
- Poster: Crain, S. P.; Huang, J.; and, Zha, H. 2008. A scalable assistant librarian: hierarchical subject classification of books. In *Proceedings of the 31st Annual international ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore, July 20 - 24, 2008). SIGIR '08. ACM, New York, NY, 799-800. DOI=<http://doi.acm.org/10.1145/1390334.1390510>
- Crain, S. P.; Bian, J.; and, Zha, H. 2009. *No query left behind: using the unique ranking needs of queries to improve ranking performance.* Preparing for resubmission.

TALKS:

- *Burning your security at three ends: Security, trust and privacy in the age of disclosure*, 2008.
<http://www.cc.gatech.edu/~scrain6/publications/WebSecurityTalk.pdf>

EDUCATION:

Degrees

- 1992 B.S. Mathematics, Michigan State University, with high honor. GPA 3.9 (4.0 scale)

All other college studies

- 1988-1991, Michigan State University. Extensive course work in Chemical Engineering before switching to Mathematics. Dimensional Analysis, mass/energy balance, control theory.
- 1992-1996, Gordon-Conwell Theological Seminary, S. Hamilton, MA. Extensive coursework in pastoral ministry. Training in library research, teaching, communication.
- 1996, Assemblies of God Theological Seminary. Single course on campus ministry. Study teaching, helping international students.
- 2006-2007, Massachusetts Institute of Technology.
 - Artificial Intelligence; project: *Diffusion wavelet analysis of Partially Observable Markov Decision Processes* (T Lozano-Perez).
 - Computer Vision; project: *Inference using image texture* (W Freeman).
- 2007-present, Georgia Institute of Technology. PhD program in computer science.
 - Web Search; project: *Weblog sentiment analysis* (H Zha).
 - Independent project: *Autistic behavior characterization* (G Abowd and J Rehg)
 - Data Analysis; project: *Optimized implementations of Belief Propagation and Relevance Vector Machine* (A Gray).
 - Military Simulation and War Gaming; project: *WWII simulation of hypothetical German advance west of Paris* (A McLean).
 - Computational Science and Engineering Algorithms; project: *Cache oblivious task queues* (D. Bader).
 - Natural Language Processing; project: *Book summarization: Leveraging ConceptNet to form coherent summaries of long documents* (A Ram).
 - Knowledge Based Artificial Intelligence; project: *On holographic memory* (A Goel).

SPECIALIZATIONS:

- Machine Learning, especially Belief Propagation, SVM, Naïve Bayes, language modeling, AdaBoost, hierarchical classification, reinforcement learning (3 years)
- Information Security (15 years)
- Optimization (20 years)
- Efficiently coding complex algorithms (10 years)
- Software architecture (10 years)
- Database schema design (9 years)
- C (20 years)
- Java (8 years)
- C++ (1 year)
- Programming Languages: C, C++, Java, Perl, Sh, Awk, Assembly, Sybase Transact-SQL, LOGO, Pascal
- Architectures: System V and BSD Unix, Linux, Windows
- Sybase database administration (9 years)
- Project design, planning, coordination, implementation, release, documentation, training, support, maintenance

EMPLOYMENT:

- | | |
|--|----------------|
| Graduate Research Assistant | 2008 – present |
| Georgia Institute of Technology | Atlanta, GA |
| Computational Science and Engineering | |
| ➤ Research assistant to Hongyuan Zha | |
| ➤ <i>From libraries to the Web: large scale hierarchical classification of book repositories</i> (Zha) | |
| ➤ <i>Query-dependent document ranking</i> (Zha) | |
| Research Intern | 2008 |
| Mitsubishi Electric Research Lab | Cambridge, MA |
| ➤ <i>Multi-view face detection using AdaBoost trees</i> (M Jones) | |
| Graduate Research Assistant | 2007 |
| Georgia Institute of Technology | Atlanta, GA |
| Computational Science and Engineering | |
| ➤ Research assistant to Hongyuan Zha and Haesun Park | |
| ➤ <i>Weblog sentiment analysis</i> (Zha) | |
| ➤ <i>From libraries to the Web: large scale hierarchical classification of book repositories</i> (Zha) | |
| ➤ <i>Nearly non-negative SVD for Non-Negative Matrix Factorization</i> (Park) | |

EMPLOYMENT (continued):

- Senior Software and Data Architect 2000 – 2007
Primus Telecommunications Group, Inc. Lynn, MA
- Provided vision for development projects
 - Oversight of database architecture
 - Served on Information Security team
 - Designed and implemented database schema and API for online purchase of Internet services. The API was implemented in Sybase stored procedures, was object oriented, and included session management, role-based security, and attribute validation. Role-based security included some extensions from second-order logic to facilitate business requirements.
 - Rewrote invoice generation code using Sybase stored procedures and strictly batch SQL. Close attention was paid to complex business rules, e.g. taxation, rounding, aging, proration, credits, reassigning services to different accounts
 - Designed and implemented XML feed of invoices to a printing and mailing vendor using Perl, Cocoon, XSLT. Integration with vendor systems using HttpClient over SSL.
 - Worked on migration from and integration with other databases
 - Performed security audits on all software development projects, some databases
 - Also educated other developers on security issues
 - Provided assistance to other developers on security, threading, Java, C, Perl, XML, XSLT, regular expressions, algorithms, project design
 - Set up and supported LDAP and RADIUS servers
 - Administered Sybase database
- Senior Developer and Unix System Administrator 1996 – 2000
Eco Software, Inc. Lynn, MA
- Designed database schema for billing and provisioning system
 - Oversaw design and security of customer self-service website
 - Wrote specifications for RPC-based UNIX provisioning system. Later, I supported and extended it
 - Designed and implemented usage (e.g. internet online time, website traffic) collection and billing infrastructure; worked with IETF RADIUS working group to revise Radius Accounting protocol to serve our business needs
 - Designed and implemented a tool for customers to view their usage
 - Installed, modified, and supported RADIUS server, including tools in Perl to manage the complex business rules in the users database and tools in C to generate reports from the accounting logs SunOS and Solaris Unix system administration
- Library Automation Assistant 1992 – 1996
Gordon-Conwell Theological Seminary S. Hamilton, MA
- Designed circulation system and periodicals catalog database
 - Set up and administered Linux system for Dial-up and Internet access to catalog
 - Recommended hardware purchases and repaired computers
 - Wrote many programs including data conversion, device drivers, email parsers

EMPLOYMENT (continued):

Teaching Assistant 1989 – 1991
MSU Math dept. E. Lansing, MI
➤ TA for Calculus I-IV, present new material, review, drill, proctor, grade exams

Research Assistant 1988 – 1989
MSU Chem E department E. Lansing, MI
➤ Assisted Dr. Jayaraman with determining how to effectively use FIDAP, a software application for finite mesh analysis of fluid and heat dynamics. Skills required include experimentation, problem solving, documentation.

VOLUNTEER:

- Chi Alpha treasurer, small group leader, board member
- Gordon-Conwell library committee
- Church of the Redeemer, board member
- Serve in Salem Mission soup kitchen
- Volunteer staff at Rescue Atlanta
- Royal Rangers leader (similar to Boy Scouts)

LANGUAGES:

- Native: English
- 2 years college: German
- Reading: Hebrew, Greek
- Single course in each of: Aramaic, Akkadian, Old Norse, French, Russian, Ugaritic

SCHOLARSHIPS:

- 1988-1989 Gerstacker Award
- 1988-1992 National Merit Scholarship
- 2007-present President's Fellowship

AFFILIATIONS:

- Chi Omega Epsilon, Chemical engineering honor society
- Pi Mu Epsilon, Mathematics honor society
- Phi Beta Kappa, honor society
- IEEE, IEEE Computer Society
- ACM

OTHER INTERESTS:

- Learning
- Linguistics
- Writing fiction
- Organic gardening
- Wood working, historical crafts, living history
- Camping, hiking
- Baking