

Dr. Abdur Chowdhury

abdur@ir.iit.edu

Dr. Abdur Chowdhury has over 12 years of research and development experience in computer science. He has worked as both a researcher and a developer, which has given him a great perspective on balancing software design with innovation. During this time he has launched over 20 commercial search products, filed more than twenty patent applications and written over 70 publications in magazines, scientific conferences, journals and book chapters covering many computer science topics like AI, networking, operating systems, system scaling, and information retrieval. For the last six years Dr Chowdhury served as the Chief Architect for Search at AOL and held positions at IIT Information Retrieval Lab and Georgetown as and adjunct professor.

PROFESSIONAL EXPERIENCE

AOL (America Online Inc.) – Chief Architect (May 2000 – August 2006)

Leading the AOL Research and Search groups for research, scaling, classification, and evaluation of AOL Systems.

Major accomplishments:

- Researched, designed and led development teams for crawling, classification, indexing and retrieval technology at AOL. Infrastructure was designed to maintain operational requirements of query response time and minimize hardware. These systems power Singing Fish, AOL Audio & Video Search, Real networks, AOL news, AOL Journals, AOL Member Directory, AOL Search and many other smaller systems.
- Lead research team at AOL producing over 20 patents and 25 academic papers.
- Research has lead to improvements in indexing speed, and search time reductions, thus reducing overall hardware needs.
- Designed and led a team to build AOL's query ontology systems. This system is able to handle query loads of 10k qps with existing hardware and can scale linearly as needs increase
- Designed and led a team to build AOL's query analysis system. This system is used by the business to forecast sales of monetized searches that has allowed them to increase revenues by millions of dollars.
- Designed and led a team to build AOL's real-time query trending systems. This system is able to handle all of AOL's query loads and classify and find trends in real-time based on concept or location.

Prior Positions:

- **Georgetown University - Adjunct Professor Computer Science** (Fall 2000 - Spring 2001), teaching: COSC 390 - Networks and Data Communications, COSC 380 - Database Design
- **IITRI** (Illinois Institute of Technology Research Institute) - *Senior Computer Scientist* (February 1999 - May 2000)
- **Group Logic Inc.** - *Senior Developer* (August 1998 - February 1999)
- **University of Maryland College Park, Maryland** - *Lab Manager / Research Engineer* (October 1997 - August 1998)
- **TTC Rockville, Maryland** - *Software Engineer* (January 1994 - October 1997)
- **George Mason University** Fairfax, Virginia - *Research Assistant* (1994 – 1996)

PATENTS

- “Analyzing A Query Log For Use In Managing Category-Specific Electronic Content”, Patent Pending, 2005.
- “Temporal Search Engine Personalization”, Patent Pending, 2005.
- “Real Time Query Trends with Multi-Document Summarization”, Patent Pending, 2005.
- “Search Category Management – Query Classification”, Patent Pending, 2005.
- “Movie Scene Search / Key Scene Detection and Retrieval”, Patent Pending, 2005.
- “Collaborative User Query Refinement”, Patent Pending, 2005.
- “Band/Artist search ranking based on Web Popularity”, Patent Pending, 2005.
- “Domain Expert Search”, Patent Pending, December 2004, (Foreign Filings 2005).
- “Search Fusion”, Patent Pending, December 2004, (Foreign Filings 2005).
- “Query Reformulation”, Patent Pending, December 2004.
- “Filtering Search Results”, Patent Pending, December 2004, (Foreign Filings 2005).

- “Query Routing”, Patent Pending, December 2004, (Foreign Filings 2005).
- “Query Disambiguation”, Patent Pending, December 2004.
- “Simplifying Dictionary Creation in Duplicate Detection Algorithms”, Patent Pending, April 2004.
- “Decreasing the Fragility of Duplicate Document Detecting Algorithms”, Patent Pending, April 2004.
- “Improving the Reliability of Duplicate Document Detecting Algorithms”, Patent Pending, April 2004.
- “Classifier Tuning Based on Data Similarities”, Patent Pending, December 2003.
- “Batch-mode Spam Classification”, Patent Pending, October 2003.
- “Dynamic Suggestion of Information to an Online User”, Patent Pending, August 2003.
- “System for Similar Document Detection”, Patent Pending, March 2000.

PUBLICATIONS

Book Chapters

- A. Chowdhury, "Automatic Evaluation of Web Search Services", ISBN: 0-12-012164-6 Book/Hardback, Advances in Computers, April 2005.

Refereed Journals

- S. Beitzel, E. Jensen, D. Lewis, A. Chowdhury, O. Frieder, “Query Classification via Semi-supervised Learning”, ACM Transactions on Information Systems (TOIS), accepted.
- S. Beitzel, E. Jensen, A. Chowdhury, O. Frieder, D. Grossman, “Temporal Analysis of a Very Large Topically Categorized Web Query Log”, JASIST (to appear).
- S. Beitzel, E. Jensen, A. Chowdhury, O. Frieder, D. Grossman, N. Goharian, “On Fusion of Effective Retrieval Strategies in the Same Information Retrieval System”, Journal of the American Society for Information Science and Technology 2004.
- A. Chowdhury, O. Frieder, D. Grossman, M. McCabe, “Collection Statistics for Fast Duplicate Document Detection,” ACM Transactions on Information Systems (TOIS), 20(2), April 2002.
- A. Chowdhury, O. Frieder, P. Wan, "On the Design, Development, Deployment, and Network Survivability Analysis of the Dynamic Routing System Protocol," Journal of Supercomputing, 21, 243-260, 2002.
- D. Nau, M. Ball, J. Baras, A. Chowdhury, E. Lin, J. Meyer, R. Rajamani, J. Splain, and V. Trichur. "Generating and Evaluating Designs and Plans for Microwave Modules." AI in Engineering Design and Manufacturing, 14(4):273-288, September 2000
- O. Frieder, D. Grossman, A. Chowdhury, G. Frieder, "Efficiency Considerations in Very Large Information Retrieval Servers," Journal of Digital Information, 1(5), April 2000.
- A. Chowdhury, O. Frieder, E. Burger, D. Grossman, K. Makki, “Dynamic Routing System (DRS): Fault tolerance in network routing”, Computer Networks And ISDN Systems (31) 1-2 (1999) pp. 87-97.

Highly Selective Conferences

- A. Kolcz, A. Chowdhury, “Avoidance of Model Re-Induction in SVM-based Feature Selection for Text Categorization”, Twentieth International Joint Conference on Artificial Intelligence 2007, Hyderabad, India, (to appear).
- S. Egelman, L. Cranor, A. Chowdhury, "An Analysis of P3P-Enabled Web Sites among Top-20 Search Results", Eighth International Conference on Electronic Commerce, August 2006.
- M. Jiang, S. Argamon, A. Chowdhury, "ZED: Explaining Temporal Variations in Query Volume", The Second International Conference on Advanced Data Mining and Applications (ADMA 2006), August 2006.
- G. Murray, A. Chowdhury, J. Lin, "Action Modeling: Using Language Models to Predict Query Behavior", SIGIR 2006, Seattle 2006.
- F. Liu, C. Yu, G. Meng, A. Chowdhury, "Effective keyword search in relational databases", SIGMOD 2006, Chicago July 2006.
- E. Jensen, S. Beitzel, A. Chowdhury, O. Frieder, "Query Phrase Suggestion from Topically Tagged Session Logs" In Proceedings of the Seventh International Conference on Flexible Query Answering Systems (FQAS 2006), Milan, Italy, June 2006.
- G. Pass, A. Chowdhury, C. Torgeson, "A Picture of Search" The First International Conference on Scalable Information Systems, Hong Kong, June, 2006.

- S. Beitzel, E. Jensen, D. Lewis, A. Chowdhury, A. Kolcz, O. Frieder, "Improving Automatic Query Classification via Semi-supervised Learning" The Fifth IEEE International Conference on Data Mining, New Orleans, Louisiana, U.S.A., November 27-30, 2005.
- A. Kolcz, A. Chowdhury, "Improved Naive Bayes for Extremely Skewed Misclassification Costs", 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) 2005.
- E. Jensen, S. Beitzel, A. Chowdhury, O. Frieder, "A Framework for Determining Necessary Query Set Sizes to Evaluate Web Search Effectiveness", ACM-WWW 2005.
- S. Beitzel, D. Lewis, A. Chowdhury, A. Kolcz, E. Jensen, O. Frieder, D. Grossman, "Automatic Web Query Classification Using Labeled and Unlabeled Training Data", Proceedings of the 2005 ACM Conference on Research and Development in Information Retrieval (SIGIR-2005), Salvador, Brazil, August 2005.
- S. Beitzel, E. Jensen, A. Chowdhury, G. Pass, O. Frieder, "Surrogate Scoring for Improved Metasearch Precision". Proceedings of the 2005 ACM Conference on Research and Development in Information Retrieval (SIGIR-2005), Salvador, Brazil, August 2005.
- E. Jensen, S. Beitzel, A. Chowdhury, O. Frieder, D. Grossman, "Predicting Query Difficulty on the Web by Learning Visual Clues", Proceedings of the 2005 ACM Conference on Research and Development in Information Retrieval (SIGIR-2005), Salvador, Brazil, August 2005.
- A. Kolcz, A. Chowdhury, J. Alspecter, "Improved robustness of signature-based near-replica detection via lexicon randomization", ACM-SIGKDD 2004.
- S. Beitzel, E. Jensen, A. Chowdhury, D. Grossman, O. Frieder, "Hourly Analysis of a Very Large Topically Categorized Web Query Log", ACM-SIGIR 2004.
- S. Beitzel, E. Jensen, A. Chowdhury, D. Grossman, O. Frieder, "Evaluation of Filtering Current News Search Results", ACM-SIGIR 2004.
- A. Chowdhury, G. Pass, "Operational Requirements for Scalable Search Systems", ACM-CIKM Conference for Information and Knowledge Management, November 2003.
- S. Beitzel, E. Jensen, A. Chowdhury, D. Grossman, "Using Titles and Category Names from Editor-driven Taxonomies for Automatic Evaluation", ACM-CIKM Conference for Information and Knowledge Management, November 2003.
- L. Ma, N. Goharian, A. Chowdhury, "Extracting Unstructured Data From Template Generated Web Documents", ACM-CIKM Conference for Information and Knowledge Management, November 2003.
- S. Beitzel, E. Jensen, A. Chowdhury, O. Frieder, D. Grossman, N. Goharian, "Disproving the Fusion Hypothesis: An Analysis of Data Fusion via Effective Information Retrieval Strategies", ACM Symposium on Applied Computing March 2003.
- A. Chowdhury, I. Soboroff, "Automatic Evaluation of World Wide Web Search Services", ACM-SIGIR, August 2002, ACM-SIGIR, August 2002.
- A. Chowdhury, D. Grossman, O. Frieder, C. McCabe, "Document Normalization Revisited", ACM-SIGIR, August 2002, ACM-SIGIR, August 2002.
- A. Chowdhury, D. Grossman, O. Frieder, C. McCabe, "Analyses of Multiple-Evidence Combinations for Retrieval Strategies", ACM-SIGIR, September 2001.
- K. Goda, T. Tamura, M. Kitsuregawa, A. Chowdhury, O. Frieder, "Query Optimizations for Vector Space Problems", ACM-SIGIR, September 2001.
- C. McCabe, J. Lee, A. Chowdhury, D. Grossman, O. Frieder, "On the Design and Evaluation of a Multidimensional Approach to Information Retrieval", ACM-SIGIR, July 2000.
- C. McCabe, A. Chowdhury, D. Grossman, O. Frieder, "A Unified Environment for Fusion of Information Retrieval Approaches", ACM-CIKM Conference for Information and Knowledge Management, November 1999.
- A. Chowdhury, E. Burger, D. Grossman, O. Frieder, "DRS: A Fault Tolerant Network Routing Protocol over IP", IEEE-IC3N International Conference On Computer Communications and Networks, 1997.
- A. Chowdhury, L. Nicklas, S. Setia, L. White, "Supporting Dynamic Space-sharing on Clusters of Non-dedicated Workstations", IEEE International Conference on Distributed Computing Systems, 1997
- S. Setia, Abdur Chowdhury, Lisa D. Nicklas, Elizabeth White, "Supporting Dynamic Reconfiguration of Parallel Applications on Clusters of Non-dedicated Workstations", Proceedings of the Eighth SIAM Conference on Parallel Processing for Scientific Computing, March 1997

PhD Thesis

- A. Chowdhury, "On the Design of Reliable Efficient Information Systems", PhD Thesis, Department of Computer Science, Illinois Institute of Technology, January 2001.

Workshops / Conferences / TreC / Technical Reports / Other

- S. Beitzel, E. Jensen, A. Chowdhury, O. Frieder, D. Grossman, "Using Manually-built Web Directories for Automatic Evaluation of Known-Item Retrieval", ACM-SIGIR Workshop on Evaluations, August 2003.
- L. Ma, N. Goharian, A. Chowdhury, "Automatic Data Extraction From Template Generated Web Pages" International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 03), Las Vegas, Nevada, 2003.
- W. Xi, K. Sidhu, A. Chowdhury, "Latent Query Stability", International Conference on Information and Knowledge Engineering IKE, Las Vegas, Nevada, 2003.
- S. Beitzel, E. Jensen, A. Chowdhury, Herman Vandermolten, "Statistical Combinations of Document Structure Information for Named and Entry Page Search", NIST Text Retrieval Conference, November 2003.
- A. Kolcz, A. Chowdhury, J. Alspector, "On the Effects of Data Duplication on Classifier Accuracy", The Twentieth International Conference on Machine Learning (ICML-2003) Workshop on Learning from Imbalanced Data Sets II.
- S. Beitzel, E. Jensen, A. Chowdhury. "Semi-Automatic Evaluation via Editor-driven Taxonomies". ACM SIGIR Workshop on Defining Evaluation Methodologies for Terabyte-Scale Collections, Toronto Canada, 2003.
- S. Beitzel, E. Jensen, A. Chowdhury, D. Grossman, N. Goharian, O. Frieder. "Recent Results on Fusion of Effective Retrieval Strategies in the Same Information Retrieval System". ACM SIGIR Workshop on Distributed Information Retrieval, New Orleans USA, 2003.
- O. Frieder, D. Grossman, A. Chowdhury, "On Scalable Information Systems", IEEE International Symposium on Network Computing and Applications, April 2003.
- A. Chowdhury, S. Beitzel, E. Jensen, "Analysis of Combining Multiple Query Representations with Varying Lengths in a Single Engine", IEEE International Conference on Information Technology: Coding and Computing, April 2002.
- A. Chowdhury, S. Beitzel, E. Jensen, D. Grossman, O. Frieder, "IIT in TREC 11", NIST Text Retrieval Conference, November 2002.
- A. Chowdhury, "Adaptive Phrase Weighting", International Symposium on Information Systems and Engineering, June 2001.
- A. Chowdhury, D Holmes, D. Grossman, O. Frieder, "IIT in TREC 10", NIST Text Retrieval Conference, November 2001.
- C. McCabe, A. Chowdhury, D. Grossman, O. Frieder, "Fusion of Information Retrieval Approaches", IEEE International Conference on Information Technology: Coding and Computing (ITCC 2001), April 2001
- O. Frieder, A. Chowdhury, D. Grossman, M. C. McCabe, "On the Integration of Structured Data and Text: A Review of the SIRE Architecture," DELOS Workshop on Information Seeking, Searching, and Querying in Digital Libraries, Zurich, Switzerland, December 2000.
- A. Chowdhury, "Memory Mapped Files and 3rd Party Libraries", December 2000, Inside Solaris Journal.
- N. Goharian, A. Chowdhury, D. Grossman, T. Ghazawi, "Efficiency Enhancements for Information Retrieval using Sparse Matrix Approach" Parallel and Distributed Processing Techniques and Applications, June 2000.
- A. Chowdhury, P. Luse, O. Frieder, P. Wan, "Network Survivability Simulation of the Commercially Deployed Dynamic Routing System Protocol", IEEE Workshop on Fault-Tolerant Parallel and Distributed Systems, May 2000.
- A. Chowdhury, D Holmes, C. McCabe, D. Grossman, O. Frieder, "Improved Query Precision using a Unified Fusion Model", NIST Text Retrieval Conference, November 2000.
- D. Nau, J. Meyer, M. Ball, J. Baras, A. Chowdhury, E. Lin, R. Rajamani, and V. Trichur. "Integrated Product and Process Design of Microwave Modules Using AI Planning and Integer Programming." In Fourth Workshop on "Knowledge Intensive CAD (KIC-4), pp. 186-196. Parma, Italy, 2000. IFIP Working Group 5.2.
- D. Nau, J. Meyer, M. Ball, J. Baras, A. Chowdhury, E. Lin, R. Rajamani, and V. Trichur. "Integrating AI Planning and Integer Programming for Use in Integrated Product and Process Design." In AAAI-2000 Workshop on Integration of AI and OR Techniques for Combinatorial Optimization. Austin, TX, 2000. AAAI.
- A. Chowdhury, A. Spitzer, "Virtual Memory Priority Paging with Solaris 2.6 and 2.7", January 2000, Inside Solaris Journal.

- A. Chowdhury, O. Frider, Spyro Papademetriou, Manish Karir, "Unnoticeable Jitter in ATM Workstation Configurations", May 1999, ISHPC - International Symposium on High Performance Computing, 1999. pp. 235-241.
- D Holmes, C. McCabe, A. Chowdhury, K. Alford, D. Grossman, O. Frieder, "Improved Manual Query Processing and Using Stemming Equivalence Classes as a Basis for Relevance Feedback", NIST Text Retrieval Conference, November 1999.
- A. Chowdhury, A. Spitzer, "Improving Server Apps with Scheduling under Solaris", January 1999, Inside Solaris Journal.
- A. Chowdhury, "IPC with Doors on Solaris 2.6", October 1998, Inside Solaris Journal
- A. Chowdhury, M. Karir, S. Papademetriou, "Evaluation of Jitter Introduced by ATM Switches and Workstations a Performance Evaluation", May 1998, TR-98-23, Institute of Systems Engineering UMD
- J. Meyer, M. Ball, J. Baras, A. Chowdhury, E. Lin, D. Nau, R. Rajamani and V. Trichur, "Process Planning in Microwave Module Production." In 1998 Artificial Intelligence and Manufacturing: State of the Art and State of Practice, September 1998.
- D Holmes, C. McCabe, D. Grossman, A. Chowdhury, O. Frieder, "Use of Query Concepts and Information Extraction to Improve Information Retrieval Effectiveness", NIST Text Retrieval Conference, November 1998.
- A. Chowdhury, C. McCabe, "Performance Improvements to Vector Space Information Retrieval Systems with POS", June 1998, TR-98-48, Institute of Systems Engineering UMD
- O. Frieder, D. Grossman, A. Chowdhury, G. Frieder, "Phrase Processing with the Relational Model and IDF Update Effects on Precision/Recall", IBM Technical Report 05.500, April 1997, (IBM Confidential until May 2002)
- A. Chowdhury, "Dynamic Reconfiguration: Checkpointing Code Generation", IEEE 5th International Symposium on Assessment of Software Tools and Technologies (SAST97), 1997
- A. Chowdhury, L. Nicklas, S. Setia, and E. White. "Supporting Dynamic Space-sharing on Clusters of Non-dedicated Workstations". Technical Report CS-97-01, George Mason University, March 1997.
- D. Grossman, C Lundquist, J Reichart, D Holmes, O. Frieder, A. Chowdhury, "Overview of the Fifth Text REtrieval Conference", NIST Text Retrieval Conference, December 1996.
- A. Chowdhury, "DRS: A Fault Tolerant Network Routing System For Mission Critical Distributed Applications", Masters Thesis, Department of Computer Science, George Mason University, December 1996.

Invited Talks / Papers

- G. Pass, A. Chowdhury, T. C., "A picture of Search", First International Conference on Scalable Information Systems (INFOSCALE 2006).
- "Duplicate Data Detection", Illinois Institute of Technology. Department of Computer Science, November 2004.
- "Operational Requirements for Scalable Search Systems", Virginia Tech. Department of Computer Science, November 2004.
- "Evaluation of World Wide Web Search Services", Virginia Tech Department of Computer Science, November 2004.
- "Automatic Evaluation of World Wide Web Search Services", University of Maryland Human Computer Interaction Lab, November 2003.
- "Operational Requirements for Scalable Search Systems", Georgetown University Department of Computer Science, October 2003.
- "Scaling Search Services Theory", Illinois Institute of Technology Department of Computer Science, November 2002.

EDUCATION

1999 - 2001, Illinois Institute of Technology, Chicago Illinois

Received: Graduate Outstanding Academic Achievement Award

Ph.D. in Computer Science

1992 - 1996, George Mason University Fairfax, Virginia

Bachelors of Science in Computer Science & Masters of Science in Computer Science

PROFESSIONAL ACTIVITIES

Editorial Boards

- Editorial Board for the Journal of the American Society for Information Science and Technology 2004-Present

Conference Committees

- Conference Committee, World Wide Web Conference 2007.
- Organizing Committee, First International Conference on Scalable Information Systems (INFOSCALE 2006) Co-sponsored by IEEE CS Society 2006.
- Organizing Committee, Track Chair Knowledge Management, for the 14th Conference on Information and Knowledge Management ACM-CIKM 2005.
- Peer reviewer for the IEEE Transactions on Information Systems, 2005-2006.
- Conference Committee, IEEE International Conference on Information Technology: Computing and Coding, 2004-2005.
- Conference Committee, International Conference on Information & Knowledge Management ACM-CIKM 1997-2003
- Vice Chair for the 2003 International Conference on Information and Knowledge Engineering
- Conference Committee Conference for ACM Special Interest Group for Information Retrieval SIGIR 2003
- National Science Foundation Panelist Reviewer - 2001
- Peer reviewer for the IEEE Transactions on Distributed Systems, 2000.
- Peer reviewer for the International Symposium on High Performance Computing, 2000.
- Peer reviewer for the IEEE International Conference On Computer Communications and Networks, 1996-2000.
- Peer reviewer for the International Symposium on High Performance Computing, 1999.