

# Chad Cumby

Accenture Technology Labs  
161 N Clark St Chicago, IL 60601  
(312) 693-2615  
chad.m.cumby at accenture dot com

RESEARCH   ◇ **Researcher**, Accenture Technology Labs,  
Accenture, Inc. (Aug 2003 - Present)  
<http://www.accenture.com/techlabs/>

At ATL I've worked on:

- **Social Media Analysis Platform** - Here we built a large-scale platform for training and deploying learned categorizers and taggers to analyze a stream of aggregated online media (via Spinn3r). The system does keyword indexing (Lucene) and uses off-the-shelf NLP annotators (UIMA and custom) to prepare datasets and do feature extraction. Then within a web-based interface a user can bootstrap an extractor and do labeling with an active learning loop. Pre-built domain specific classifiers for sentiment detection were created for 200 product categories. Potential pilot applications include automated warranty issue tracking for products, and adverse event detection for pharmaceuticals.
- **Enterprise Knowledge Management** - This set of projects focused on using machine learning and IR methods to augment enterprise applications (staffing, proposal writing, etc) with data extracted from large company knowledge-bases. I developed a new entity retrieval and ranking system to allow people to better find domain experts. Also I've been working on a data anonymization system for text and unstructured data that would preserve the data's utility for other tasks. (classification, adversarial classification, search algorithms, optimization)
- **Information Extraction** - In my group we have implemented several prototype information extraction systems for Accenture clients. In one case we used a semi-supervised learning algorithm (Co-EM), with a novel active learning component, to extract new unknown product attributes and their values for an online retailer. In another case we used several strategies to extract clinical trial design attributes and outcomes from research papers for a major pharma company. (semi-sup learning, structured learning, active learning, entity recognition)
- **Individual Consumer Modeling** - This project used grocery point of sale purchase data to predict customers' shopping lists as they enter the store on an individual basis, in addition to modeling several other aspects of shopping behavior. I created an *Promotion Planning* tool that uses these behavior models to plan promotions for a store and project their outcomes. The first iteration of this tool ran on large hires collaborative touch-sensitive display. This work was featured in *The Numerati* by Stephen Baker. (Large scale learning, classification, feature extraction, visualization, optimization, simulation using data mining results)
- **Human Performance Enhancement via UWB Locationing** - This project leads an industrial worker through a standard maintenance scenario, where the environment can sense the location of the worker's hand and provide context specific assistance through a mobile device. The procedure is recorded and annotated using a pan-tilt-zoom camera and event information extrapolated from the locationing.

◇ **Research Assistant**, Department of Computer Science,  
University of Illinois at Urbana-Champaign (Sept 2001 – May 2003)

<http://l2r.cs.uiuc.edu/~cogcomp/>

Advisor: Dan Roth.

Assistantship with the Cognitive Computations lab. My research has focused on Machine Learning, as it applies to applications in natural language processing and other domains with structured or semi-structured data. Topics I have focused on include: relational learning, feature construction, propositionalization, kernels for structured data. Research projects:

- Developed a generalized object model and description language for the FEX feature extraction system. FEX can be used to generate features for propositional learners such as the SNoW system from text and other structured data. Developed a library of classes in C++ to use feature extraction methods and the SNoW learning system together, in preparation for developing a so-called Learning Based Programming platform. This platform would allow researchers to create programs with learned functions and variables.
- Developed a kernel-learning method, parameterizable by the above description language, for structured and relational data. Conducted classification experiments in Bioinformatics and Named Entity recognition.
- Assisted our group in the TREC-11 Question Answering competition. Worked on heuristic Answer Selection component for detecting relations in candidate passages to match relations present in questions.
- Worked on learning method for adapting parameters in the Extended Boolean information retrieval model based on features of queries, in order to optimize retrieval relevance.

- ◇ **Software Development Intern**, Microsoft Corp.,  
Natural Language Group (June 2001 – August 2001)

Summer internship with Microsoft's Natural Language Group. Designed and implemented a demo command and control application in C++ utilizing aspects of the semantics engine being developed by the group.

- ◇ **Undergraduate Research Programmer**, Dept. of Computer Science,  
University of Illinois at Urbana-Champaign, (January 1999 – May 2001)

EDUCATION ◇ **University of Illinois at Urbana-Champaign**, Urbana, IL.

M.S. in Computer Science, May 2003.

Master's project: *Feature Extraction Languages for Propositionalization in Relational Learning*

B.S. in Mathematics and Computer Science with High Distinction, May 2001.

- ◇ **Awards & Honors**

- Recipient of AFCEA educational fund Ralph Schrader Scholarship - 2003
- Nominated for CRA Outstanding Undergraduate Award - 2000, 2001
- Recipient of NSF REU (Research for Undergraduates) Grant,
- UIUC College of Liberal Arts and Sciences James' Scholars,
- Dean's List (Fall 1997 – Spring 1999),
- Robert C. Byrd Scholarship (Fall 1997 – May 2001)

SKILLS ◇ Programming in C/C++ under UNIX, Win32

- ◇ Programming in C# (linq, gui, web-services, ASP.net)

- ◇ Programming in Java

- ◇ Numerical programming with Matlab, C++ (uBLAS)
- ◇ Perl, PHP, bash shell scripting
- ◇ Apache Thrift
- ◇ MySQL/MS-SQLServer db design, ad-hoc query, data load
- ◇ MongoDB
- ◇ Data/Text mining packages (SPSS Clementine, UIMA, WEKA)
- ◇ Interactive visualization tools (Advizor, Tableau)
- ◇ Search engine administration (FAST, Lucene/Solr)
- ◇ Building with make, ant, ivy, cvs, svn
- ◇ Creating documents with L<sup>A</sup>T<sub>E</sub>X

- PUBLICATION HISTORY ◇ A Machine Learning Based System for Semi-Automatically Redacting Documents (with R. Ghani). In Proceedings of *IAAI*. 2011. *to appear*
- ◇ A search engine for finding highly relevant applications (with M. Grechanik, C. Fu, Q. Xie et al). In Proceedings of *ICSE*. 2010.
  - ◇ Exemplar: EXEcutable examPLes ARchive. (with M. Grechanik, C. Fu, Q. Xie et al). In Proceedings of *ICSE*. 2010.
  - ◇ Protecting Sensitive Topics in Text Documents with PROTEXTOR. In Proceedings of *ECML/PKDD*. 2009.
  - ◇ Guide: A GUI differentiator (with Q. Xie, M. Grechanik, C. Fu). In Proceedings of *ICSM*. 2009.
  - ◇ Retrieval and Ranking of Entities for Enterprise Knowledge Management Tasks (with K. Probst, R. Ghani). Semantic Search workshop at *WWW 2009*. 2009.
  - ◇ Graph Structure Learning for Task Ordering (with Y. Yang et al). In Proceedings of *ICEIS 2009*. 2009.
  - ◇ Data Mining for Individual Consumer Models and Personalized Retail Promotions (with R. Ghani, A. Fano, and M. Krema). Book Chapter - Data Mining Methods and Applications. 2007.
  - ◇ OpenFactory: Enabling Situated Task Support in Industrial Applications (with S. Kurth, and A. Fano). In Proceedings of Pervasive 2006 Workshop "Pervasive Technology Applied". 2006.
  - ◇ Learning Individual Customer Models for Personalized Promotions: A Data Mining Case Study (with A. Fano, R. Ghani, and M. Krema). Workshop on Data Mining for Business at ECML 2005. 2005
  - ◇ Building Intelligent Shopping Assistants Using Individual Consumer Models (with A. Fano, R. Ghani, and M. Krema). In Proceedings of *IUI 2005*. 2005.
  - ◇ Predicting Customer Shopping Lists from Point-of-Sale Purchase Data (with A. Fano, R. Ghani, and M. Krema). In Proceedings of *KDD 2004*. 2004.
  - ◇ Feature Extraction Languages for Propositionalized Relational Learning (with Dan Roth). In Proceedings of *IJCAI 2003 Workshop on Learning Statistical Models from Relational Data*. 2003.
  - ◇ On Kernel Methods for Relational Learning (with Dan Roth). In Proceedings of *ICML'03*. 2003.
  - ◇ Question-Answering via Enhanced Understanding of Questions (with D. Roth, X. Li, P. Morie, R. Nagarajan, V. Punyakanok, N. Rizzolo, K. Small, W. Yih). In *Proceedings of TREC'02*. 2002
  - ◇ Learning with Feature Description Logics (with Dan Roth). In *Proceedings of ILP'02*. 2002.

◇ Relational Representations that Facilitate Learning (with Dan Roth). In *KR 2000: Principles of Knowledge Representations and Reasoning*. 2000.

◇ SNoW User's Guide (with A. Carlson, J. Rosen, and Dan Roth). *UIUC Tech Report UIUC-DCS-R-99-210*. 1999.  
(Available at <http://12r.cs.uiuc.edu/~danr/Papers/userguide.ps.gz>)

PROF.

ACTIVITIES

◇ **R**ewriter:

- WWW, International World Wide Web conference
- SIGIR, ACM Sig on Information Retrieval
- ICML, International Conference on Machine Learning
- KDD, ACM Sig on Knowledge Discovery in Databases

REFERENCES

Available on request.