

Jonathan H. Clark

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Newell-Simon Hall 4502
5000 Forbes Avenue
Pittsburgh, PA 15213

Phone: (501) 351-1061
Office: Newell-Simon Hall 3612
jhclark@cs.cmu.edu
<http://www.cs.cmu.edu/~jhclark>

Education

Master of Language Technologies, School of Computer Science August 2007 – May 2009
Carnegie Mellon University, Pittsburgh, PA

Bachelor of Science, Computer Science with minor in Mathematics May 2007
Texas Christian University, Fort Worth, TX
Magna Cum Laude
GPA: 3.85 (4.0 in major)

Publications

- J. Clark**, R. Frederking, L. Levin "Inductive Detection of Language Features via Clustering Minimal Pairs: Toward Feature-Rich Grammars in Machine Translation", *The Second Workshop on Syntax and Structure in Translation (SSST) at the Association for Computational Linguistics (ACL)*, 2008. Columbus, Ohio.
- J. Clark**, R. Frederking, L. Levin "Toward Active Learning in Corpus Creation: Automatic Discovery of Language Features During Elicitation", *The Sixth Language Resources and Evaluation Conference (LREC)*, 2008. Marrakech, Morocco.
- J. Clark**, C. Hannon, "A Classifier System for Author Recognition Using Synonym-Based Features", *Sixth Mexican International Conference on Artificial Intelligence*, November 2007.
- J. Clark**, C. Hannon, "An Algorithm for Identifying Authors Using Synonyms", *ENC 2007*, September 2007.
- M. Bowden, M. Olteanu, P. Suriyentrakorn, **J. Clark**, D. Moldovan, "LCC's PowerAnswer at QA@CLEF 2006," *CLEF 2006 Working Notes*, September 2006.
- C. Hannon, **J. Clark**, "A Cognitive-Based Approach to Learning Integrated Language Components", *The Third International Workshop on Natural Language Understanding and Cognitive Science*, May 2006.

Posters and Presentations

- J. Clark**, C. Hannon, "An Algorithmic Approach to Recognizing Authors Using Synonym Weighting", *TCU Student Research Symposium*, April 2007. (Received Computer Science Best in Department Award)
- J. Clark**, C. Hannon, "A Computational Method for Identifying Authors Using Synonyms", *Presentation at the National Council on Undergraduate Research 2007*, April 2007.
- J. Clark**, C. Hannon, "Author Attribution via Synonyms", *Poster at SIGCSE 2007 Student Research Competition*, March 2007.
- J. Valentino II, **J. Clark**, "The Cerberus System: Using Artificial Vision for 3D Simulation and Robot Navigation", *TCU Student Research Symposium*, April 2005. (Received Computer Science Best in Department Award)

Experience

Carnegie Mellon University, Language Technologies Institute **August 2007 – Present**

- Researched methods for improving machine translation of low-resource languages
- Implemented “navigator” to elicit the most relevant sentence pairs from bilingual informants based on linguistic knowledge base
- Gained hands-on knowledge of current state-of-the art in machine translation

Language Computer Corporation (Richardson, TX) **May - July 2006, 2007**

- Interned with a research-based natural language processing (NLP) company
- Built system to collect parallel corpus for languages with scarce resources
- Participated in 3-person group submitting to NIST MT 06 in both Arabic and Chinese to English machine translation
- Integrated machine translation and question-answering components for CLEF 06 (Cross-Language Question Answering Evaluation) submission

Google Summer of Code **June - August 2005**

- Coded for Gaim, a multi-platform multi-protocol Instant Messaging Tool
- Gained professional experience from the mentoring of Mark Doliner and Gaim developers
- Authored 25-page document that became the basis for implementing OSCAR file transfers in both Kopete in the 2006 Summer of Code and the SHAIM client in fall 2006

Crescent Lab for Intelligent Systems (TCU Dept. of Computer Science) **2005 - 2007**

- Acted as Lead Student Research Assistant, organizing a team of researchers
- Worked one-on-one with Dr. Charles Hannon, assisting in his research
- Selected by department to represent TCU at the NCUR 2007 conference

Ellis Jewelers, Inc. (North Little Rock, AR) **1998 - 2005**

- Created custom automatic daily backup system in Java, tailored to store needs
- Designed user interfaces using Access forms and VBA / SQL aimed at ease of use
- Provided software troubleshooting, hardware maintenance, and network administration

ACM Programming Contests **2004 – 2007**

- Founding member of “/usr/bin/tcu,” a campus organization to learn new algorithms weekly
- Winner of TCU individual programming contest two consecutive years using Java (judged based on solving the most algorithm-based problems in shortest time)
- Selected as part of team for ACM South-Central Regional Programming Competition

Projects

Treegraft SCFG Transduction Parser - <http://code.google.com/p/treegraft/> **2008 - Present**

- Created a transduction parser for synchronous context free grammars in the style of the CMU Avenue XFER3 system
- Reviewed literature on core issues in state-of-the-art synchronous parsing technologies
- Used: Java 5, Eclipse, JUnit

Akerblad Sentence Aligner - <http://akerblad.sourceforge.net/> **July 2007**

- Ported LDC’s Champollion sentence aligner to Java
- Crafted plugin-based architecture to allow for future research-oriented development by others
- Used: Java 5, Eclipse

NumberTrans Java Number Translator - <http://numbertrans.sourceforge.net/> **2006 - Present**

- Created Java library for translating numbers including cardinals, ordinals, and decimals between various human languages and scripts (intended for use in Machine Translation)
- Gained knowledge of over a dozen numeral scripts including Arabic, Chinese, Korean, and Thai
- Used: Java 5, Eclipse, JUnit

- TCU SmartHome Senior Design Project** - <http://www.tcu.dev.com/> **2006 - 2007**

 - Integrated disparate systems in TCU's SmartHome using Agent Oriented techniques as Chief Software Architect
 - Used: PostgreSQL, Eclipse, Java 5, JUnit

- Alchemy / Goal Mind Natural Language Research System** **2005 - 2007**

 - Managed Beowulf cluster
 - Supported the use of Alchemy, a PVM-like application for parallelizing cognitive models
 - Analyzed latency and synchronization issues in Alchemy
 - Used: Network File System (NFS), Distributed Compiler (Distcc), Secure Shell Host (SSH), Bash Scripting

- Gaim (Pidgin) Instant Messaging Client** - <http://www.pidgin.im> **2005**

 - Analyzed the behavior of the OSCAR instant messaging protocol and official AIM (AOL Instant Messenger) clients and integrated file transfer capability via proxy server into Gaim
 - Collaborated with team of developers
 - Used: C, GDB (GNU Debugger), CVS, Ethereal Packet Analyzer

- Cerberus Artificial Vision Project** - <http://crescent.cs.tcu.edu/> **2005 – 2006**

 - Researched and implemented a distributed multi-camera artificial vision program
 - Communicated navigation data to Pioneer Robot via Wireless Network
 - Used: C, C++, GNU Make, Linux Socket Library, Video for Linux (V4L)

- TCU Catholic Community Website** - <http://www.catholic.tcu.edu/> **2003 - 2007**

 - Integrated with Google Calendar's XML capabilities
 - Updated look and feel of site to include drop-down menus and modern graphics
 - Used: ASP, DHTML, XML, JavaScript, MS Access, and Macromedia Studio

Grants and Awards

- Supported by Research Assistantship at Carnegie Mellon University (Fall 2007 – Present)
- Awarded \$1500 grant from Texas Christian University for Natural Language Research (Spring 2006)
- Mentored research assistants in writing an additional \$3000 of successful grant proposals (Spring 2006)
- Earned Dan Drew Scholarship from Upsilon Pi Epsilon national honor fraternity for the computing sciences (Spring 2006)
- Funded by Texas Christian University Dean's Scholarship for duration of undergraduate career
- Named Senior Scholar in Computer Science for Texas Christian University Class of 2007 (Spring 2007)

Involvement

Member – Association for Computational Linguistic	2008 - Present
Member – Association for Computing Machinery	2004 - Present
President – TCU Computer Science Society	2006 - 2007
Vice President – Upsilon Pi Epsilon Computer Science Honor Fraternity (TCU Chapter)	2006 - 2007
Eagle Scout – Boy Scouts of America	1990 - 2003