David A. Jurgens

CONTACT INFORMATION

UCLA Computer Science Dept.
4683 Boelter Hall
Los Angeles CA, 90025

Cell: (217) 412-85 73
Fax: (310) 206-8133
E-mail: jurgens@cs.ucla.edu
WWW: cs.ucla.edu/~jurgens

RESEARCH INTEREST

I am interested in statistical methods for extracting knowledge from text. My core research problems involve (1) automatically discovering the different meanings of a word, (2) identifying and comparing how concepts are related, and (3) methods for computing larger units of meaning, such as relations and sentences. My research has direct applications in many natural language processing tasks, and can also be applied to modeling human cognitive processes, new forms of information retrieval, and language understanding.

EDUCATION

University of California, Los Angeles, Los Angeles, CA

Ph.D. Candidate, Computer Science

- Thesis Topic: *Distributional Semantic Models for Extracting and Comparing Semantic Relations Between Words*
- Advisor: Professor Michael G. Dyer
- Area of Study: Natural Language Processing

Washington University in St. Louis

M.S., Computer Science, December 2004

- Thesis: *Road Extraction From Aerial Video Using Active Contours and Motion Cues*
- Advisor: Professor Robert Pless
- Area of Study: Computer Vision

B.A., Philosophy, May 2004

- *College Honors*,
- Second Major in Political Science
- Minor in Computer Science

RESEARCH EXPERIENCE

University of California, Los Angeles, Los Angeles, CA

Distributional Semantics, Word-Relation Extraction and Analogy
September 2006 - *present*
Advisor: Professor Michael Dyer

Distributional semantics makes the assumption that for two words their similarity in meaning is predicted by the similarity of the linguistic contexts in which each word appears. My research extends this model in three areas. First, I investigate how this model can be used to identify the different meanings a single word may have, e.g. "earth" as both a planet and as dirt. This work is primarily focused on the task of Word Sense Induction. However, I have also applied it to automated methods for discovering newsworthy events in a stream of news stories.

The two remaining research focuses are complementary. I investigate how to identify meaningful relationships between words; second, I examine to represent the relationships in such a way that they can be meaningfully compared. For example, a system

performing the first task might discover that "Titan *orbits* Saturn," and "Mercury *revolves around* the Sun." The second task would then assess how similar "orbits" is to "revolves around" to determine if the two words share an *analogous* relationship. My research aims discover and model these relationships through lexical and syntactic patterns between word pairs.
As a part of my research on distributional semantics, I am actively involved in the development of an open source package for building these models: The S-Space Package. This work is freely available to the research community and has been downloaded over a thousand times.

HRL Laboratories, Malibu, CA

Information and Systems Science Lab (ISSL), September 2010 – *present*
Principle Investigator: Dr. Tsai-Ching Lu

My research focuses on the intersection between network science and linguistics. I explore how information extraction can be used to enrich network models with new types of entities and relations between entities (e.g., clarifying the relationship between two individuals in a social network). Furthermore, I investigate new models for detecting communities within these enriched networks. As a part of this work, I also invented a new method for community detection that takes into account the relationships between members.

Sun Microsystems Laboratories, Burlington, MA

Project Darkstar, June – September 2006, 2007, 2008
Principle Investigator: Dr. Jim Waldo, *Mentor:* Seth Proctor

Projected Darkstar (PDS) is an application server for low-latency applications such as massively multiplayer online games, which need to scale to thousands of users while still responding quickly to messages. The over-arching research challenge is how to design a distributed system with the desired latency and scalability requirements. As a research intern on PDS, my work focused on three areas. First, I designed scalable, high-concurrency data structures for applications: a hash map, deque and queue. The second focus was on designing a pluggable profiling framework that allowed developers to identify performance issues. This work included a transactional logging framework, and new profiling reporting mechanisms. The third focus was on designing an efficient scheduler for processing the application workload. As a part of this, I invented one new latency-optimized scheduling algorithm and co-invented a scheduling data-structure, both currently submitted for patenting with a goal for later publication.

Project Vidscape, June 2006 - July 2006
Principle Investigator: Dr. William A. Woods

I developed a distributed infrastructure for analyzing and annotating video segments for later text-based search. During the project, I implemented a new color clustering heuristic for image segmentation and created new representations for shape estimation as a time series. Our work focused on creating an intermediate logical representation for joining video interpretation with the Sun Labs text-based search engine. This project was ended prematurely due to a restructuring at Sun Labs.

Digital Arithmetic and Reconfigurable Architecture Laboratory

at the University of California, Los Angeles September 2005 - June 2006
Advisor: Professor Milŏs Ercegovac

I investigated the application of VLIW processors for performing serial most-significant digit first (MSDF) arithmetic using only integer operations. Using the results, I then developing a floating point emulation library for embedded VLIW processors implemented using only integer instructions and online floating-point MSDF arithmetic. The library is designed for space- and power-constrained embedded VLIW processors without a floating-point unit. This work was a part of a larger investigation for STMicroelectronics, where the floating-point emulation library was to be integrated with a specialized compiler for embedded processors.

Media and Machines Lab
at Washington University in St. Louis
Advisor: Professor Robert Pless

I investigated fully-automatic extraction of road maps from aerial video by using active contours, or "snakes," to represent both rural and urban roads. We developed a space-efficient spatio-temporal representation of motion estimates using optic flow and tensor fields. This allowed us to invent new heuristics for identifying traffic motion patterns in noisy motion data, which were used to build the final road map of the scene. All research algorithms were implemented using the Intel C++ OpenCV library.

Distributed Object Computing Lab
at Washington University in St. Louis
Advisor: Professor Ron K. Cytron

I designed and implemented a distributed framework for testing the feasibility of Real-Time Java in a production environment. Each distributed component was designed to operate in both the standard JVM and a Real-Time JVM. This allowed for dynamic scalability testing as new Real-Time components replaced default JVM components. We used AspectJ to weave in reference counting code extensions to replace existing garbage collection as a test for memory requirements and overhead. The work was presented at the 2004 DARPA PCES conference.

Publications

Refereed

David Jurgens, "An Evaluation of Graded Sense Disambiguation using Word Sense Induction." Proceedings of *SEM, the First Joint Conference on Lexical and Computational Semantics, 2012. ACL.

David Jurgens and Tsai-Ching Lu, "Friends, Enemies, and Lovers: Detecting Communities in Networks Where Relationships Matter" Proceedings of Web Science, 2012. ACM.

David Jurgens and Tsai-Ching Lu, "Temporal Motifs Reveal the Dynamics of Editor Interactions in Wikipedia." Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM), 2012. AAAI.

David Jurgens and Keith Stevens, "Measuring the Impact of Sense Similarity on Word Sense Induction." Proceedings of the EMNLP 2011 Workshop on Unsupervised Learning in NLP, 2011. ACL.

David Jurgens, "Word Sense Induction by Community Detection." Proceedings of the ACL 2011 Workshop TextGraphs-6: Graph-based Methods for Natural Language Processing, 2011. ACL.

David Jurgens and Keith Stevens. "Capturing Nonlinear Structure in Word Spaces Through Dimensionality Reduction." Proceedings of the ACL 2010 Workshop on GEometrical Models of Natural Language Semantics (GEMS), 2010.

David Jurgens and Keith Stevens. "HERMIT: Flexible Clustering for the SemEval-2 WSI Task." Proceedings of the ACL 2010 SemEval-2010 Workshop, 2010.

David Jurgens and Keith Stevens. "The S-Space Package: An Open-Source Framework for Word Space Algorithms." Proceedings of the ACL 2010 System Demonstrations, 2010.

David Jurgens and Keith Stevens. "Event Detection in Blogs using Temporal Random Indexing." In *Proceedings of the International Workshop on Events on Emerging Text Types (eETTs)*, pages 21-28, 2009.

Robert Pless and David Jurgens. "Road extraction from motion cues in aerial video." In *Proceedings of the ACM Conference on Geographic Information Systems*, pages 31-38, 2004.

Forthcoming

David Jurgens and Ioannis Klapaftis, "SemEval-2013 Task 10: Graded Word Sense Induction: SemEval-3 Task Proposal." *to appear in SemEval-2013*.

David Jurgens, Saif M. Mohammad, Peter D. Turney, and Keith J. Holyoak. "SemEval-2012 Task 2: Measuring Degrees of Relational Similarity: SemEval-3 Task Proposal." *to appear in SemEval-2012*.

Manuscripts

David Jurgens and Keith Stevens. "Word Ordering with Reduced Dimensionality for Sense Induction." Technical Report 090020, University of California Los Angeles, 2010.

David Jurgens. "Road Extraction From Aerial Video Using Active Contours and Motion Cues." Masters Thesis, Computer Science Department, Washington University in St. Louis, 2005.

Awards and Fellowships

UCLA Computer Science Department Fellowship, September 2005 – June 2006

First place and Most Substance award, `$CONTEST`, 2010, a research presentation competition for the UCLA Computer Science Department. The 2010 contest was themed "Motivations for Research."

First place , `$CONTEST`, 2011, a research presentation competition for the UCLA Computer Science Department. The 2011 contest was themed "Research impact in Science."

Mentoring

Undergraduate Students

Sky Lin, Grace Park, and Alex Nau, 2009
I worked closely with three students as a part of the CS 199 Course: Directed Research in Computer Science. As their mentor, I helped them identify research topics in distributional semantics. Throughout the spring quarter, I chaired weekly group meetings

with the students. Their resulting research was added to the S-Space Package.

Moon Bae, 2009
I worked with Moon to design a project that bridged his two interests: natural language and vision.

Sean August Scot Moon, 2010
I worked with Sean on a linguistics-focused project for analyzing dependency parsers.

Alexander Honda, 2010
I worked with Alexander on examining feature importance in word-space models. This work was done as a part of the CS 199 Course: Directed Research in Computer Science.

Vasishta Jayanti and David Cohen, 2010 – *present*
I worked with Vasishta and David on extending lexical ontologies with word-space models. This work was done as a part of the CS 199 Course: Directed Research in Computer Science, and continues as independent studies.

Teaching Experience

University of California, Los Angeles, Los Angeles, CA

Engineering 183 - Engineering and Society
Teaching Fellow, *Fall 2009, Spring 2010, Spring 2011 quarters*

This course addresses the professional and ethical considerations in the practice of engineering. Emphasis is placed on the impact of technology on society and the development of moral and ethical values. This course also serves as the technical writing course for the School of Engineering. Lectures cover ethical topics, while teaching assistant-led sections cover technical writing.

I taught a weekly three-hour discussion on technical writing and ethics. As the teaching assistant, I was the primary instructor for all writing content for the course. I helped the instructors develop a new curriculum as well as created supplemental materials for other teaching assistants to use. The resulting materials and lesson plans were aggregated into a wiki for future teaching assistants to use. In addition, I organized weekly meetings for current quarter's teaching assistants to assess teaching strategies and collaborate on lesson plans.

Computer Science 136 - Computer Security
Teaching Assistant, *Spring 2009 and Winter 2010 quarters*

The class teaches real-world securty analysis, using the DETER testbed for experiments. Students use existing knowledge of operating systems, computer networks, and UNIX programming to perform practical labs to analyze and fix insecure systems. Topics include filesystem permissions, firewalls, buffer overflows, pathname attacks, SQL injection, intrusion detection, man-in-the-middle attacks, and replay attacks.

My work focused on teaching one two-hour discussion per week. In addition, I designed numerous supplemental materials and mini-exercises to introduce students to the lab topics.

Engineering 111 - Introduction to Finance and Accounting for Engineers
Teaching Assistant, *Winter 2009 quarter*

Engineering 111 is the second course of a three-part series designed to introduce engineers to topics they would encounter in a full MBA program. The course focused on a wide range of topics: critical components of finance and marketing; research and

practice as they impact the management of technology commercialization; and internal and external marketing and financing of high-technology innovation. Concepts included present value, future value, discounted cash flow, internal rate of return, return on assets, return on equity, return on investment, interest rates, and the cost of capital.

I taught two one-hour discussion sections per week, prepared exams, and developed homework and supplemental material for students. I also organized two workshops for SAS programming in the business environment.

Engineering 110 - Introduction to Technology Management and Economics for Engineers
Teaching Assistant, *Fall 2008 quarter*

Engineering 110 the first course of a three-part series designed to introduce engineers to topics they would encounter in a full MBA program. The course focused on the fundamental principles of micro- and macro-economics as they relate to technology management and how individuals, firms, and governments impact successful commercialization of high technology products and services.

I taught two one-hour discussion sections per week, prepared exams, and developed homework and supplemental material for students.

Computer Science 161 - Artificial Intelligence
Teaching Assistant, *Winter 2008 quarter*

CS161 introduces students to a survey of artificial intelligence topics as well as the Lisp programming language. Topics include search algorithms, problem spaces, first order logic, expert systems, natural language processing, genetic algorithms, probability, decision trees and neural networks. Projects are designed to reinforce selected material with specific programming exercises.

My teaching responsibilities consisted of teaching two two-hour sections per week. I also was responsible for grading homework and exams.

Computer Science 132 - Compiler Construction
Teaching Assistant, *Fall 2006, Winter 2007, Fall 2007, and Spring 2008 quarters*

As an upper-division undergraduate course, CS132 builds on many prerequisite courses of formal language and automata theory and software development. This course focuses on the stage-based construction of a compiler, starting with subset of Java as input and finally producing MIPS assembly. Topics include LL and LALR parsing, type checking, semantic analysis, liveness analysis, object and memory layout, register allocation, and other topics of the instructor's choosing.

My teaching responsibilities consisted of leading one two-hour discussion session per week in addition to proctoring multiple office hours. I designed grading rubrics and homework test cases as well as graded homework and exams.

Computer Science 111 - Operating Systems Principles
Teaching Assistant, *Spring 2007 quarter*

CS111 presents an in-depth overview of many operating system fundamentals. Student projects include both application-level and kernel-level development. Topics include processes and scheduling, threads and synchronization, file systems, virtual memory, distributed systems and security.

My teaching responsibilities were to update the projects as well as design and implement new project ideas. I also taught one two-hour section per week and graded homework and exams.

Washington University in St. Louis, St. Louis, MO

Computer Science 102 - Introduction to Computer Science II
Teaching Assistant, *Fall 2002 through Spring 2004 semesters*

CS 102 is the second semester introduction into software development and computer science topics. The course emphasizes more sophisticated uses of object-oriented concepts and techniques for managing communication among software components. Students are introduced to packages, file I/O, parsing, graphical user interfaces, exception handling, threads, concurrency, synchronization, and network programming. Algorithms and data structures are presented as needed to support discussion of these topics.

My teaching responsibilities included holding weekly office hours. During several semesters I designed many of the homework assignments and developed all reference implementations. In the fall of 2003 and spring of 2004, I was the head teaching assistant, which had the responsibility of designing all grading rubrics and leading the graders during weekly grading sessions.

Computer Science 519 - Computer Vision
Teaching Assistant, *Spring 2004 semester*

This course covers the theory and practice of extracting information from images and video. Topics include the classical concepts of the geometry of image and video capture, creation of image and video mosaics, and techniques for creating descriptions of 3D objects and scenes. The course also includes an overview of current vision research topics, including video textures, non-Lambertian surface modeling, and multi-camera and catadioptic imaging systems.

My teaching responsibilities involved holding weekly office hours, designing homework rubrics and grading assignments.

Industrial Experience

Microsoft, Bing
Research Software Development Intern, June 2010 - September 2010, Redmond, WA

I investigated how identify fault patterns in time series of noisy features from performance monitoring sensors. The goal was to extract higher-order relationships not reflected in a single monitoring sensor. The resulting research was integrated into post-mortem analyses of service-impacting issues. In addition, I developed a validation tool for verifying the correctness of transformation scripts, which are used by Bing services to converted data between different versions of the internal Bing API.

Microsoft, MSN Video
Software Development Intern, July 2009 - September 2009, Redmond, WA

I invented and subsequently developed a new video fingerprinting technology to catalog MSN Video clips. This work was integrated with a workflow that added closed captioning to MSN Video clips, which had significant value additions of searching inside the video's spoken text and provided additional content tags for search. In addition, I

created natural language pattern extraction technology to expand the set of video tags based on website content.

Moscience, Inc.

Lead Artificial Intelligence Architect, February 2008 - June 2008, Los Angeles CA.

I was a member of a five-person start-up based on presenting internet content as a continuously browsable media stream. My work focused on created a semantic clustering algorithm to relate pop-culture data. Connections between pop-culture concepts, celebrities and movies were combined into a novel graphical format for an internet-as-appliance browser. My contribution generated a working prototype that was essential in securing a second round of angel funding.

amazon.com

Software Development Engineer, February 2005 - September 2005, Seattle, WA

I was part of a software team that supported both the amazon.com product catalog and website content build system, as well as software tools used by the business clientéle for interacting with these systems. I integrated additional vendors and data sources into the amazon.com website content, which generated an estimated 750K USD in company revenue per year. In addition, I created new software tools for improving developer and business user productivity, resulting in a net savings of 8 hours of developer time and 15 hours of nontechnical work per week. Lastly, I designed a new alarming system for team-maintained services to increase system stability and reduce downtime.

Service

Program Committees

*Sem, First Joint Conference on Lexical and Computational Semantics, Montreal, Canada

Computer Science Graduate Student Committee

Co-founder of a UCLA graduate student group dedicated to improving the academic and social lives of graduate students. My contributions include

1. organizing and chairing a graduate student town hall panel session for new students (Fall 2010);
2. mentoring incoming Masters and PhD students (2008 - *present*);
3. organizing parts of the 2010 prospective student visit day;
4. creating an annual survey of the research software developed at UCLA, to increase collaboration (Spring 2009);
5. starting a weekly tea-time, which is now attended by over 150 students and faculty (Fall 2007);
6. and organizing the bi-annual department picnics and coordinating with corporate sponsors (Spring 2009, 2010, Fall 2009).

Software Systems

S-Space Package is an open source package written in Java for building and evaluating word space algorithms. The package includes reference implementations of frequently cited algorithms, specialized data structures for natural language processing, and multi-threaded matrix implementations for concurrent algorithms. The project is available at code.google.com/p/airhead-research/.

I founded this project and have been a primary contributor along with with Keith Stevens. This package is being developed as a part of my research, and makes all of the software used for my research freely available.

Project Darkstar is an open source application server for designing massively multi-player online applications that require high concurrency and low latency. This project was formerly supported by Sun Microsystems; after the company's acquisition by Oracle, the project was forked and rebranded as Red Dwarf. The project is now available at reddwarf.sourceforge.net.

I worked on Project Darkstar during my time at Sun Labs; my contributions were focused on scalable data structures, system and application profiling, and scheduling.

Patents

"Method for stage-based cost analysis for task scheduling," with Seth T. Proctor and James Megquier. USPTO Patent application 20090328046, 2009.

"Methods and apparatus for window-based fair priority scheduling," with Seth T. Proctor and David R. Chase. USPTO Patent application 20080235693, 2007.

References

Available upon request