

## Keith D. Stevens

---

CONTACT INFORMATION	<b>Institute for Scientific Computing Research</b> <b>Lawrence Livermore National Lab 7000 East Ave,</b> <b>Livermore, CA 94551</b>	<i>Mobile:</i> +1-805-433-2295 <i>E-mail:</i> <a href="mailto:kstevens@cs.ucla.edu">kstevens@cs.ucla.edu</a> <i>WWW:</i> <a href="http://kracken.cs.ucla.edu">kracken.cs.ucla.edu</a>
OBJECTIVE	Placement in a development and/or research position centered around distributed semantics and unsupervised learning, along with applying current theories on these topics to web search or other human oriented data exploration.	
RESEARCH INTERESTS	My research focuses on the unsupervised acquisition of word senses and the automated evaluation of related models. Word sense acquisition, or Word Sense Induction (WSI), attempts to use contextual information to learn distinct word senses as they appear in large corpora. My initial work centered around developing a simple framework for building and applying WSI models. My work now focuses on finding automated intrinsic and extrinsic evaluation metrics that explain how to improve these models.	
EDUCATION	<b>University of California, Los Angeles</b> , Los Angeles, California USA  Ph.D., Computer Science, In Progress <ul style="list-style-type: none"><li>• Thesis Topic: <i>Intrinsic and Extrinsic evaluation frameworks for Word Sense Induction models</i></li><li>• Adviser: <a href="#">Professor Michael G. Dyer</a></li><li>• Area of Study: Computational Linguistics</li></ul> M.S., Computer Science, October 2011 <ul style="list-style-type: none"><li>• Thesis Topic: <i>Extending Word Sense Induction with waiting and depending</i></li><li>• Adviser: <a href="#">Professor Michael G. Dyer</a></li><li>• Area of Study: Computational Linguistics</li></ul> B.S., Computer Science, December 2007 <ul style="list-style-type: none"><li>• Computer Science</li><li>• Minor in Mathematics</li></ul>	
RESEARCH EXPERIENCE	<b>Lawrence Livermore National Lab</b> , Livermore, California USA  <b>Word Sense and Topic Model Evaluation</b> June 2011 - <i>present</i> Advisor: <i>David Buttler</i>  I am currently evaluating three “topic models” with two recently developed coherence metrics that have been closely associated with human judgements. Our current results indicate that the coherence metrics, originally designed for Latent Dirichlet Allocation, can be applied to other techniques such as Non-negative Matrix Factorization and Singular Value Decomposition, as seen in Latent Semantic Analysis). Furthermore, we are evaluating new aggregate metrics for comparing entire models, rather than just individual topics.  Based on the above research, I am beginning to investigate the applicability of these coherence metrics as an intrinsic evaluation of Word Sense Induction (WSI) models. Current evaluation strategies for WSI focus on costly human judgements and do not provide a method for improving the models. Along these lines, I am also Consensus Clustering as an evaluation of the feature space used in WSI models and the stability	

of clustering models used to differentiate word senses.

### **Ontology Expansion**

June 2010 - September 2010

Advisor: *David Buttler*

With David Buttler, I have investigated methods of automatically enhancing existing concept hierarchies for the purpose of organizing unstructured text documents. This work is in coordination with my work in Distributional Semantics and will soon be used by several information retrieval tasks at the national lab.

**University of California, Los Angeles**, Los Angeles, California USA

### **Distributional Semantics**

September 2008 - June 2011

Advisor: *Michael G. Dyer*

Distributional semantics assumes that the semantics of words can be defined by the contexts in which they occur, an idea proposed by Firth in 1957. The learned semantics have been shown to approximate semantic judgements made by humans on psychological and standardized tests. My research in this field has three main goals.

First, models for distributional semantics have been developed in a variety of fields, such as Computational Linguistics, Cognitive Science, and Artificial Intelligence. In order to permit accurate evaluation of distinct models, a common framework is needed for their implementation and evaluation. To this end, I have collaborated with David Jurgens to develop the S-Space Package, a java based framework for implementing, testing, and extending distributional models.

Second, current distributional models are incapable of distinguishing between distinct meanings for a single word, for example, cat can refer to feline pets, large felines found in the wild, or a brand of construction vehicles. To learn these differences, I combine distributional models with clustering techniques. Initial research in this field, often known as Word Sense Induction, has already been shown to perform well according to one measure in a coordinated semantic task, SemEval, and to be applicable towards detecting the occurrence of news worthy events.

Lastly, current distributional models lack organization. Word semantics are unorganized and disconnected from each other. In contrast, concept hierarchies provide a rich set of relationships between the semantics associated with many words. I investigate methods of combining distributional models and Word Sense Induction models for the purpose of automatically building concept hierarchies. For this, I focus on two approaches: extending existing hierarchies by enhancing current semi-supervised techniques and inferring hierarchies automatically from word semantics learned through only distributional methods.

## **PUBLICATIONS**

### **Referred**

Keith Stevens, Terry Huang and David Buttler, "The C-Cat Wordnet Package: An Open Source Package for modifying and applying Wordnet." *Systems Demonstration of the Global Wordnet Conference 2012*, 2012.

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 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*, 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.

### Manuscripts

David Jurgens and Keith Stevens. “Word Ordering with Reduced Dimensionality for Sense Induction.” Technical Report 090020, University of California Los Angeles, 2009.

Keith Stevens. “Extending Word Sense Induction with waiting and depending.” Masters Thesis, University of California Los Angeles, 2010.

### In Progress

Keith Stevens and Others. “Applying Topic Coherence metrics over many models and many topics.”

Keith Stevens and Others. “Evaluation Induced Word Senses with Topic Coherence Metrics.”

Keith Stevens and Others. “Understanding Word Sense Induction with Consensus Clustering.”

### TEACHING EXPERIENCE

#### University of California, Los Angeles, Los Angeles, California USA

##### Computer Science 32 - Introduction to Computer Science II

Teaching Assistant, *Winter 2009 Spring 2009*

This course introduces students to basic object oriented design a fundamental data structures. Topics include C++ classes, C++ templates, pointer management, trees and traversal methods, inheritance, and recursion. As the teaching assistant, I was responsible for teaching a weekly 2 hour discussion section that reviewed class lectures and broke down class assignments.

##### Computer Science 35L - Software Construction Laboratory

Teaching Assistant, *Fall 2008 Fall 2009*

This course introduces students to fundamental concepts for developing software in a unix environment. Topics include Bash and Python scripting, The GNU debugger, basic C programming, version control, network security, and basic operating system tools. As the teaching assistant, I was responsible for all lectures and lab assignments.

### **Computer Science 111 - Introduction to Operating System Principles**

Teaching Associate, *Spring 2010 Fall 2010*

This course introduces students to the basic theories behind modern operating systems, with a focus on theories behind the Linux Kernel. Topics include concurrency, device management, file systems, memory management, process and thread management, Remote Procedure Calls, scheduling, security and protection, and shell design. As the teaching assistant, I applied theories taught in class to a six lab assignments that covered virtual memory, process management, kernel modules, shell design, file systems, and remote applications.

### **Computer Science 132 - Modern Compiler Construction**

Teaching Associate, *Winter 2011*

This course introduces students to the theories and practices behind compiler design. Students learn topics focusing on parsing, type checking, register allocation, and virtual machines. As the teaching assistant, I applied theories taught in lectures to practical examples of compilers and assisted students with lab assignments, along with grading of all lab assignments.

### **Engineering 183 - Engineering and Society**

Teaching Associate, *Winter 2011*

This course introduces students to professional and ethical considerations experienced in engineering positions. 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 covered all writing requirements for students and provided a series of writing exercises designed to improve their ability to concisely describe, evaluate, and solve ethical dilemmas in Engineering.

#### PROFESSIONAL EXPERIENCE

#### **Lawrence Livermore National Lab**, Livermore, CA

*Student Research Intern, June 2010 to September 2010*

Mentor: *David Buttler*

- Implemented an existing taxonomy expansion algorithm for a highly distributed framework based on Hadoop and HBase
- Implemented several methods for condensing an existing taxonomy
- Evaluated several methods for integrating taxonomy induction and condensation techniques
- Evaluated several methods for enhancing taxonomy expansion and condensation methods such that the resulting taxonomy is customized for a particular domain of knowledge
- Developed a new, java based interface for the widely used WordNet taxonomy

#### **Google, Inc.**, Kirkland, WA

*Software Engineering Intern, June 2009 to September 2009*

Mentor: *Zhen Lin*

- Enhanced a document selection system with the addition of semantic analysis. This analysis provided a more detailed representation of the document, as it changed over several points in time. This analysis allowed the application to inform users of more relevant aspects of the page.
- Enhanced crawling framework for the document selection system, improving savings by up to 50% usage of external resources.
- Restructured document selection system to be more scalable and abide by internal policies.

*Software Engineering Intern, January 2008 to September 2008*

Mentor: *Hizakazu Igarashi*

- Built a highly scalable, stateless server for responding to a wide variety, and high volume, of requests.
- Developed several plugins for the designed server allowing access to a wide variety of data sources.
- Improved serving time of the server by addition of caching responses in memory
- Improved response time of a query formulation system with the use of a Trie, and integration as a plugin for the server.
- Heavily untested, documented, and debugged same server.

**Networked and Embedded Systems Laboratory**, Los Angeles, CA

*Student Research Intern, March 2007 to December 2007*

Mentor: *Professor Mani Srivastava*

- Designed multiple modules for use on a Micaz/Mica2 sensor board using the SOS embedded operating system, which is based on a Linux architecture.
- Reduced the transmission rate of nodes by defining a time series representing sensor values.
- Designed a unit test framework for the SOS operating system.

## SERVICE

Recent contributor to several open-source software projects, including:

- [Vim-LaTeX](#) suite
- [Vimperator](#) and [Pentadactyl](#) Firefox extensions
- [Git](#) distributed version control system
- [Mercurial](#) distributed version control system
- Personal projects archived at <http://hg.tedpavlic.com/>

Frequent contributor to [Wikipedia](#).

- Significant contributions to articles on control theory, electronics, and signals and systems.

**OSU FIRST Robotics Team**, The Ohio State University, 2000–2004

- Introduced middle school and high school students to science and technology by participating with them in national robotics competitions.
- Led 2002 team to regional silver medal *Engineering Inspiration Award*.
- *Lead Team Mentor*, 2002–2004
- *Component Design Team Lead Mentor*, 2001–2002

Institute for Electrical and Electronics Engineers (IEEE), Member, 2002–present

- IEEE Control Systems Society (2004–present)
- IEEE Computer Society (2009–present)
- IEEE Intelligent Transportation Systems Society (2011–present)
- IEEE Systems, Man, and Cybernetics Society (2011–present)

- IEEE Robotics and Automation Society (2011–present)

Animal Behavior Society, Member, 2011–present

Director of Computers, [Engineers’ Council](#), [The Ohio State University](#), 2002

[Linux Virtual Server Project](#), 1999–2000

- Early member of the team that formed the open-source project that is now an important load balancing solution for the Linux software platform.

[Greater Columbus Free-Net](#), 1995–1997

- Provided technical support services.

CompuTeen Bulletin Board System, 1993–1995

- Administrated dial-up bulletin board system.
- Founded and administrated TeenLiNK, an international electronic mail network that spread through the United States, Canada, and Australia and delivered mail over a series of electronic dial-up drop offs.

APPLICATION AREAS	Autonomous and Unmanned Vehicles, Flexible Manufacturing Systems, Distributed Power Generation, Intelligent Lighting, Power Demand Response, Microgrids, Smart Grids
HARDWARE AND SOFTWARE SKILLS	<p>Analog and Digital Electronics:</p> <ul style="list-style-type: none"> <li>• Bipolar and FET implementations of continuous and switched amplifiers, modulators, converters, and filters</li> <li>• Computer-Aided Design Tools: Cadence OrCAD, NI Multisim, SPICE, pst-circ</li> </ul> <p>Embedded and Real-time Systems:</p> <ul style="list-style-type: none"> <li>• Software and hardware development with several MCU and DSP platforms (e.g., Motorola MCU’s, Texas Instruments DSP’s, Atmel ATmega MCU’s, Microchip PIC MCU’s, and others)</li> </ul> <p>Instrumentation, Control, Data Acquisition, Test, and Measurement:</p> <ul style="list-style-type: none"> <li>• <a href="#">dSPACE</a> hardware (e.g., RTI1104) and Control Desk software, <a href="#">Simulink</a>, <a href="#">LabVIEW</a> and other <a href="#">National Instruments</a> control and data acquisition hardware and software (e.g., MIO, SMIO, DSA, DMM, and others), Hewlett-Packard and Agilent bench-top equipment</li> </ul> <p>Computer Programming:</p> <ul style="list-style-type: none"> <li>• C, C++, Java, JavaScript, NetLogo, Pascal, Perl, PHP, Lisp, UNIX shell scripting (including POSIX.2), GNU make, AppleScript, SQL, MySQL, MATLAB, Maple, Mathematica, and others</li> </ul> <p>Version Control and Software Configuration Management:</p> <ul style="list-style-type: none"> <li>• DVCS (Mercurial/MQ, Git/StGit), VCS (RCS, CVS, SVN, SCCS), and others</li> </ul> <p><a href="#">MATLAB</a> skill set:</p> <ul style="list-style-type: none"> <li>• Linear algebra, Fourier transforms, Monte Carlo analysis, nonlinear numerical methods, polynomials, statistics, <math>N</math>-dimensional filters, visualization</li> <li>• Toolboxes: communications, control system, filter design, genetic algorithm and direct search, signal processing, system identification</li> </ul> <p>Software Verification:</p> <ul style="list-style-type: none"> <li>• KeY, PRISM, KeYmaera</li> </ul>

Information/Internet Technology:

- Networking (UDP, TCP, ARP, DNS, Dynamic routing), Services (Apache, SQL, MediaWiki, POP, IMAP, SMTP, application-specific daemon design)

Productivity Applications:

- T<sub>E</sub>X (L<sup>A</sup>T<sub>E</sub>X, B<sub>B</sub>T<sub>E</sub>X, PSTricks), Vim, most common productivity packages (for Windows, OS X, and Linux platforms)

Operating Systems:

- Microsoft Windows family, Apple OS X, IBM OS/2, Linux, BSD, IRIX, AIX, Solaris, and other UNIX variants

## EXPERTISE

Mathematics:

- Applied Mathematics, Real and Complex Analysis, Measure Theory, Differential Geometry, Game Theory, Graph Theory, Combinatorics

Control Theory and Engineering:

- Linear and Nonlinear Systems Theory, Feedback, Variable Structure Systems and Sliding Modes, Distributed and Intelligent Control, Dynamic Optimization, Biomimicry, Bio-inspiration, Hybrid and CyberPhysical Systems

Communications and Signal Processing:

- Probability, Random Variables, Stochastic Processes, Information Theory, Estimation, Networks

Computer Science and Engineering:

- Model Checking (automated, distributed, hybrid, probabilistic), Hybrid Automata, Software Verification, Component-Based Reusable Software

Natural Sciences (Biology, Neuroscience, Psychology, Anthropology):

- Behavioral Ecology, Foraging Theory, Altruism, Impulsiveness, Evolution

## REFERENCES AVAILABLE TO CONTACT

**Dr. Kevin M. Passino** (e-mail: [passino.1@osu.edu](mailto:passino.1@osu.edu); phone: +1-614-312-2472)

- Professor, Electrical and Computer Engineering, The Ohio State University
- ◊ 205 Drees Laboratories, 2015 Neil Ave., Columbus, OH 43210
- ★ *Dr. Passino was my graduate adviser.*

**Dr. Bruce W. Weide** (e-mail: [weide.1@osu.edu](mailto:weide.1@osu.edu); phone: +1-614-292-1517)

- Professor and Associate Chair, Computer Science and Engineering  
The Ohio State University
- ◊ 395 Drees Laboratories, 2015 Neil Ave., Columbus, OH 43210
- ★ *Dr. Weide is a co-PI on the NSF grant that funds my current postdoctoral position.*

**Dr. Ian M. Hamilton** (e-mail: [hamilton.598@osu.edu](mailto:hamilton.598@osu.edu); phone: +1-614-292-9147)

- Assistant Professor, Evolution, Ecology, and Organismal Biology and Mathematics  
The Ohio State University
- ◊ 300 Aronoff Laboratory, 318 W. 12th Avenue, Columbus, OH 43210
- ★ *Dr. Hamilton has been a valuable interdisciplinary resource to me.*

**Dr. Andrea Serrani** (e-mail: [serrani.1@osu.edu](mailto:serrani.1@osu.edu); phone: +1-614-292-4976)

- Associate Professor, Electrical and Computer Engineering  
The Ohio State University
- ◊ 205 Drees Laboratories, 2015 Neil Ave., Columbus, OH 43210
- ★ *Dr. Serrani was a member of my doctoral committee.*

- Dr. Paolo A. G. Sivilotti** (e-mail: [sivilotti.1@osu.edu](mailto:sivilotti.1@osu.edu); phone: +1-614-292-5835)
- Associate Professor, Computer Science and Engineering, The Ohio State University
  - ◊ 395 Dreese Laboratories, 2015 Neil Ave., Columbus, OH 43210
  - ★ *Dr. Sivilotti is a co-PI on the NSF grant that funds my current postdoctoral position.*
- Dr. Richard J. Freuler** (e-mail: [freuler.1@osu.edu](mailto:freuler.1@osu.edu); phone: +1-614-688-0499)
- Professor of Practice, Mechanical and Aerospace Engineering  
The Ohio State University
  - ◊ 244 Hitchcock Hall, 2070 Neil Ave., Columbus, OH 43210
  - ★ *Dr. Freuler coordinates the Fundamentals of Engineering for Honors program in which I served as an instructor early in my academic career.*
- Dr. George H. Staab** (e-mail: [staab.1@osu.edu](mailto:staab.1@osu.edu); phone: +1-614-292-7920)
- Associate Professor, Mechanical and Aerospace Engineering  
The Ohio State University
  - ◊ W192 Scott Laboratory, 201 W. 19th Ave., Columbus, OH 43210
  - ★ *Dr. Staab is the faculty adviser for the OSU FIRST robotics and engineering outreach group of which I was a four-year member and team leader.*
- Dr. Clayton Daigle** (e-mail: [Clayton.Daigle@silabs.com](mailto:Clayton.Daigle@silabs.com); phone: +1-512-532-5935)
- Mixed-Signal Engineer, Silicon Laboratories, Austin, TX
  - ★ *Dr. Daigle was my direct supervisor when I worked for National Instruments as an analog hardware R&D engineer.*