

Jonathan Gordon

jgordon@jonathangordon.org

<https://jonathangordon.org>

RESEARCH INTERESTS

My research interests are broadly in artificial intelligence and computational linguistics. In my research, I use natural language processing to identify knowledge implicit in large collections of text, and I apply this knowledge to build innovative applications. I extract commonsense knowledge for machines to understand language and reason about the world, and I infer scientific conceptual structure to guide learners.

EDUCATION

2009–2014	PhD, Computer Science <i>University of Rochester</i>	Rochester, NY
	Thesis: “Inferential Commonsense Knowledge from Text” Advisor: Lenhart Schubert. Committee: James Allen, Daniel Gildea, Gregory Carlson.	
2007–2009	MS, Computer Science <i>University of Rochester</i>	Rochester, NY
2003–2007	BA, Computer Science <i>Vassar College</i>	Poughkeepsie, NY

RESEARCH EXPERIENCE

2014–present	Postdoctoral Researcher <i>USC Information Sciences Institute</i>	Marina del Rey, CA
	Performed research on diverse DARPA-, IARPA-, and NIH-funded projects involving natural language processing and information extraction.	
	<ul style="list-style-type: none">• Created abstract, symbolic knowledge to support natural language understanding and commonsense reasoning.• Investigated the multilingual, knowledge-driven interpretation of metaphors.• Inferred prerequisite relations among concepts in scientific publications and online videos and applied these to generate structured reading lists.• Supervised projects by MS students and visiting PhD students.	

- 2007–2014 **Research Assistant**
University of Rochester Rochester, NY
- As a graduate student, carried out NSF- and DARPA-funded research on knowledge extraction from text and helped to supervise junior students.
- Created tools for knowledge extraction from Web-scale text and measured the effect of text quality on knowledge extraction.
 - Investigated the use of targeted extraction to overcome the effects of reporting bias on what we learn about the world from text.
 - Created methods to “sharpen” underspecified, possibilistic factoids into explicitly quantified, partially disambiguated logical axioms for use in a general reasoner.
- Summer 2008 **Visiting Research Assistant**
USC Institute for Creative Technologies Marina del Rey, CA
- Created tools using latent semantic analysis (LSA) to automatically evaluate student essays for an intelligent tutoring system.
 - Ran human trials to train the system and test automated feedback.

PUBLICATIONS

BD2K Training Coordinating Center’s ERuDite: the Educational Resource Discovery Index for Data Science.

José Luis Ambite, Lily Fierro, **Jonathan Gordon**, Gully Burns, Florian Geigl, Kristina Lerman, and John Van Horn. IEEE Transactions on Emerging Topics in Computing (TETC) Special Issue on Scholarly Big Data, in review.

Linking Educational Resources on Data Science.

José Luis Ambite, **Jonathan Gordon**, Lily Fierro, and Gully Burns. International Semantic Web Conference, in review. Data: [doi:10.5281/zenodo.1214375](https://doi.org/10.5281/zenodo.1214375)

Democratizing Data Science through Data Science Training.

John Van Horn, Lily Fierro, Jeana Kamdar, **Jonathan Gordon**, Crystal Stewart, Avnish Bhattra, Sumiko Abe, Xiaoxiao Lei, Caroline O’Driscoll, Aakanchha Sinha, Priyambada Jain, Gully Burns, Kristina Lerman, and José Luis Ambite. In Proc. of the Pacific Symposium on Biocomputing, 2018.

Structured Generation of Technical Reading Lists.

Jonathan Gordon, Stephen Aguilar, Emily Sheng, and Gully Burns. In Proc. of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2017.

An Investigation into the Pedagogical Features of Documents.

Emily Sheng, Prem Natarajan, **Jonathan Gordon**, and Gully Burns. In Proc. of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2017.

BD2K ERuDIte: the Educational Resource Discovery Index for Data Science.

José Luis Ambite, Lily Fierro, Florian Geigl, **Jonathan Gordon**, Gully Burns, Kristina Lerman, and John D. Van Horn. In Proc. of the Fourth www Workshop on Big Scholarly Data: Towards the Web of Scholars (BigScholar), 2017.

Modeling Concept Dependencies in a Scientific Corpus.

Jonathan Gordon, Linhong Zhu, Aram Galstyan, Prem Natarajan, and Gully Burns. In Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), 2016.

A Corpus of Rich Metaphor Annotation.

Jonathan Gordon, Jerry Hobbs, Jonathan May, Michael Mohler, Fabrizio Morbini, Bryan Rink, Marc Tomlinson, and Suzanne Wertheim. In Proc. of the Third Workshop on Metaphor in NLP, 2015.

High-Precision Abductive Mapping of Multilingual Metaphors.

Jonathan Gordon, Jerry Hobbs, Jonathan May, and Fabrizio Morbini. In Proc. of the Third Workshop on Metaphor in NLP, 2015.

Inferential Commonsense Knowledge from Text.

Jonathan Gordon. PhD Thesis. University of Rochester, 2014.

Reporting Bias and Knowledge Acquisition.

Jonathan Gordon and Benjamin Van Durme. In Proc. of the CIKM Workshop on Automated Knowledge Base Construction (AKBC), 2013. (Best Paper Award)

WordNet Hierarchy Axiomatization and the Mass-Count Distinction.

Jonathan Gordon and Lenhart Schubert. In Proc. of the IEEE International Conference on Semantic Computing (ICSC), 2013.

Using Textual Patterns to Learn Expected Event Frequencies.

Jonathan Gordon and Lenhart Schubert. In Proc. of the NAACL Workshop on Automatic Knowledge Base Construction and Web-Scale Knowledge Extraction (AKBC-WEKEX), 2012.

Towards Adequate Knowledge and Natural Inference.

Lenhart Schubert, **Jonathan Gordon**, Karl Stratos, and Adina Rubinoff. In Proc. of the AAAI Fall Symposium on Advances in Cognitive Systems, 2011.

Episodic Logic: Natural Logic + Reasoning.

Karl Stratos, Lenhart Schubert, and **Jonathan Gordon**. In Proc. of the IEEE International Conference on Knowledge Engineering and Ontology Development (KEOD), 2011.

Discovering Commonsense Entailment Rules Implicit in Sentences.

Jonathan Gordon and Lenhart Schubert. In Proc. of the EMNLP Workshop on Textual Entailment (TextInfer), 2011.

Quantificational Sharpening of Commonsense Knowledge.

Jonathan Gordon and Lenhart Schubert. In Proc. of the AAAI Fall Symposium on Commonsense Knowledge, 2010.

Learning from the Web: Extracting General World Knowledge from Noisy Text.

Jonathan Gordon, Benjamin Van Durme, and Lenhart Schubert. In Proc. of the AAAI Workshop on Collaboratively-built Knowledge Sources and Artificial Intelligence (WikiAI), 2010.

Evaluation of Commonsense Knowledge with Mechanical Turk.

Jonathan Gordon, Benjamin Van Durme, and Lenhart Schubert. In Proc. of the NAACL Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, 2010.

Weblogs as a Source for Extracting General World Knowledge.

Jonathan Gordon, Benjamin Van Durme, and Lenhart Schubert. In Proc. of the Fifth International Conference on Knowledge Capture (K-CAP), 2009.

PRESENTATIONS

Distribution and Inference.

With Jerry Hobbs. DSALT: Workshop on Distributional Semantics and Linguistic Theory at ESSLI, 15–26 August 2016. Submitted.

Towards Learning World Knowledge Suitable for Inference.

Reasoning with Text Workshop, 18–19 February 2011, USC Institute for Creative Technologies. Invited.

TEACHING EXPERIENCE

- 2008–9, 2013 **Teaching Assistant**
University of Rochester Rochester, NY
- Assisted with computer science courses for undergraduate and graduates students, including lecturing, writing programming projects and labs, supervising undergraduate TAs, and grading homework and exams.
- CSC 444: Logical Foundations of Artificial Intelligence, Fall 2009 and Fall 2013
 - CSC 108: Introduction to Computers, Fall 2008
 - CSC 170: Introductory Computer Programming, Spring 2008
- Fall 2004 **Lab Coach**
Vassar College Poughkeepsie, NY
- Ran introductory Computer Science labs, helping students program in Java and holding weekly office hours for questions.

ADDITIONAL WORK EXPERIENCE

- 2005–2007 **Assistant System Administrator**
Vassar College Poughkeepsie, NY
- Helped to maintain three labs of research workstations, department servers, and faculty computers.
 - Worked on transitioning from Solaris and Mac OS X to Linux and moving servers to running on Xen virtual machines.
 - Created a custom Web-based solution for monitoring system status.

PROFESSIONAL ACTIVITIES

- Reviewer, Empirical Methods in Natural Language Processing (EMNLP), 2018.
- Program committee, Cognitum Workshop on Cognitive Knowledge Acquisition and Applications, 2015, 2016, 2018.
- Program committee, Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2018.
- Program committee, AHA! Workshop on Information Discovery in Text, 2014.
- Reviewer, AAAI Fall Symposium on Commonsense Knowledge (CSK 2010).
- Site organizer, North American Computational Linguistics Olympiad (NACLO), 2017.
- Meeting organizer, Seminar in Knowledge, Inference, and Language, Department of Computer Science, University of Rochester, 2009–2014

SKILLS

Programming	Especially in Python; experience with Java, C++, Common Lisp, and other languages; Git.
NLP/ML tools	NLTK, Scikit-learn, Gensim, Word2vec, Mallet, parsing.
Systems	Linux and macOS; Docker containers; SQL databases and Elastic-Search; parallel processing with Torque and Slurm.
Publication	Web (HTML and CSS) and print (L ^A T _E X).

Los Angeles, May 2018