

Profile

Data scientist with a strong background in user modeling, machine learning, data mining, natural language processing and software engineering. I enjoy uncovering patterns from unstructured data and developing experiments and algorithms which inform us about the nature of human-human and human-computer interaction.

Education

- 2007–2012 **Joint-PhD, Computer Science and Cognitive Science,**
University of Colorado Boulder, Boulder, Colorado.
Thesis: DISCUSS Toward a Domain Independent Representation of Dialogue
Committee: Wayne Ward (co-chair), Martha Palmer (co-chair), James Martin, Tamara Sumner, Sarel Van Vuuren, Barbara Fox
Areas of Focus: Natural Language Processing, Educational Technologies
- 2007–2010 **MS, Computer Science,**
University of Colorado Boulder, Boulder, Colorado.
- 1996–2000 **BS, Mathematical and Computer Sciences, Computer Science Specialty**
BS, Engineering, Electrical Specialty,
Colorado School of Mines, Golden, Colorado.

Professional Experience

- August 2013–present **Hapara, Lead Data Scientist, Project Lead,**
Analytics Team, Boulder, Colorado.
Lead research and development for Hapara Analytics team. Responsible for discovering, extracting and communicating actionable insights from live educational data using millions of interactions from Google Apps for Education. Architect and develop data ETL flows, text analytics/natural language processing and machine learning pipelines to enable user modeling, predictive analytics, social network analysis and data visualization. Create engaging and unique visualizations of data insights for end-user applications. Provide thought leadership and direction for future product roadmaps. Advise engineering teams on tech-transfer of research into implementation of production systems.
- November 2012–July 2013 **Avaya, Research Scientist,**
Avaya Labs Research, Westminster, Colorado.
Developed, investigated and prototyped new techniques for social media analysis, natural language processing, and information retrieval. Architected next-generation natural language processing pipeline for social media analysis. Led Avaya's research efforts around sentiment analysis and automated dialogue systems. Patented 15 claims.
- January 2012–August 2012 **University of Colorado Boulder, Research Assistant,**
Center for Computational Language and Education Research, Boulder, Colorado.
SHARP: Strategic Health IT Advanced Research Projects - Developed machine learning based components for extracting structured information from clinical notes using natural language processing. Transitioned cTAKES open source project to ClearTK infrastructure.

- Summer 2011 **Microsoft Corporation, Research Intern**,
Microsoft Research - Machine Learning Group, Redmond, Washington.
 Developed a machine learning and natural language processing based system for automatic generation and evaluation of educational questions from text. Managed data collection via Amazon Mechanical Turk. Published results in ACL conference.
- 2008-2012 **University of Colorado Boulder, Research Assistant**,
Center for Computational Language and Education Research, Boulder, Colorado.
 Developed the natural language dialogue system for MyST, a computer-based science tutor for elementary school students. Developed DISCUSS a semantically informed logical representation for modeling dialogue interactions. Managed a team of linguistic annotators in tagging dialogue transcripts with DISCUSS. Developed models for automatic question generation. Conducted Mechanical Turk Experiments to gauge efficacy of automated systems.
- 2006-2007 **Volunteers in Asia (VIA), Volunteer English Teacher**,
Pesantren Sunan Pandanaran, Yogyakarta, Indonesia.
 Taught conversational English to middle and high school students
- 2005-2006 **Intel Corporation (acquired from HP), VLSI CAD Tools Software Engineer**,
Fort Collins Design Center, Fort Collins, CO.
 Developed APIs, algorithms, and software tools to support VLSI design of high-end server microprocessors.
- 1999-2005 **Hewlett Packard Company, VLSI CAD Tools Software Engineer / Verification Engineer**,
Microprocessor Technology Lab, Fort Collins, CO.
 Developed software infrastructure to support VLSI design and architectural verification of high-end server microprocessors.

Technical Skills

Specialties: Machine Learning, User Modeling, Natural Language Processing / Text Analytics, Data Mining, Predictive Modeling, Information Retrieval

Big Data: Spark (including SparkSQL, MLib, Spark Streaming and GraphX), Hadoop

Programming Languages: Python (Scikit-Learn, Pandas, Numpy, SciPy), Scala, Java, SQL, Go, C#, C/C++, Javascript, Matlab, R

Machine Learning Frameworks: Scikit-Learn, Mallet, Weka

Natural Language Processing frameworks: ClearTK, UIMA, UIMAFit, DKPro, NLTK

Information Retrieval frameworks: Lucene, Elastic Search

Visualization: D3.js, NVD3, Matplotlib, Bokeh

Crowdsourcing and Annotation: Amazon Mechanical Turk API, Experimental Design, Statistical Agreement

Web Development: REST Endpoint Design, Angular JS, Django, Google AppEngine, Cloud Computing

Open Source Contributions

- Co-owner ClearTK - Toolkit for developing statistical natural language processing components on UIMA
- Contributor cTAKES - Clinical Text Analysis and Knowledge Extraction System

Publications

Steven Bethard, Philip Ogren, and Lee Becker. ClearTK 2.0: Design patterns for machine learning in UIMA. In *LREC 2014, Ninth International Conference on Language Resources and Evaluation*, May 2014.

Stephen T. Wu, Vinod C. Kaggal, Dmitriy Dligach, James J. Masanz, Pei Chen, Lee Becker, Wendy W. Chapman, Guergana K. Savova, Hongfang Liu, and Christopher G. Chut. A common type system for clinical natural language processing. *Journal of Biomedical Semantics*, 4, January 2014.

Dmitriy Dligach, Steven Bethard, Lee Becker, Timothy Miller, and Guergana K. Savova. Discovering body site and severity modifiers in clinical texts. *Journal of American Informatics Association (JAMIA)*, October 2013.

Lee Becker, George Erhart, David Skiba, and Valentine Matula. Avaya: Sentiment analysis on twitter with self-training and polarity lexicon expansion. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), in conjunction with the Second Joint Conference on Lexical and Computational Semantics (*SEM 2013)*, 2013.

Lee Becker. *DISCUSS: Toward a Domain Independent Representation of Dialogue*. PhD thesis, University of Colorado Boulder, November 2012.

Lee Becker, Sumit Basu, and Lucy Vanderwende. Mind the Gap: Learning to Choose Gaps for Question Generation. In *The 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Montreal, Canada, June 2012.

Lee Becker, Martha Palmer, Sarel van Vuuren, and Wayne Ward. Learning to Tutor Like a Tutor: Ranking Questions in Context. In *The 11th International Conference on Intelligent Tutoring Systems*, Crete, Greece, June 2012.

Lee Becker, Martha Palmer, Sarel van Vuuren, and Wayne Ward. Question Ranking and Selection in Tutorial Dialogues. In *The 7th Workshop on Innovative Use of NLP for Educational Applications*, Montreal, Canada, June 2012.

Caleb Phillips, Lee Becker, and Elizabeth Bradley. Strange Beta: An Assistance System for Indoor Rock Climbing Route Setting. *Chaos*, 22(1), March 2012.

Wayne Ward, Ron Cole, Daniel Bolanos, Cindy Buchenroth-Martin, Edward Svirsky, Sarel van Vuuren, Tim Weston, Jing Zheng, and Lee Becker. My Science Tutor: A conversational multimedia virtual tutor(submitted for review). *Journal of Educational Psychology*, 2012.

Lee Becker, Martha Palmer, Sarel van Vuuren, and Wayne Ward. Evaluating questions in context. In *AAAI 2011 Fall Symposium on Question Generation*, Arlington, VA USA, November 2011.

Caleb Phillips, Lee Becker, and Elizabeth Bradley. Strange Beta: An assistance system for indoor rock climbing route setting using chaotic variations and machine learning. Technical Report CU-CS-1088-11, University of Colorado Boulder, Boulder, CO USA, October 2011.

Wayne Ward, Ron Cole, Daniel Bolanos, Cindy Buchenroth-Martin, Edward Svirsky, Sarel van Vuuren, Tim Weston, Jing Zheng, and Lee Becker. My Science Tutor: A conversational multi-media virtual tutor for elementary school science. *ACM TSLP: Special Issue on Speech and Language Processing of Children's Speech for Child-machine Interaction Applications*, 7(1), August 2011.

Lee Becker, Wayne Ward, Sarel van Vuuren, and Martha Palmer. DISCUSS: A dialogue move taxonomy layered over semantic representations. In *IWCS 2011: The 9th International Conference on Computational Semantics*, Oxford, England, January 2011.

Lee Becker, Rodney D. Nielsen, Ifeyinwa Okoye, Tamara Sumner, and Wayne Ward. What's next? Target concept identification. In *Proceedings of the 3rd Workshop on Question Generation*, Pittsburgh, PA USA, June 2010.

Lee Becker, Wayne Ward, and Sarel van Vuuren. DISCUSS: Enabling detailed characterization of tutorial interactions through dialogue annotation. In *EDM 2011: Proceedings of the 3rd International Conference on Educational Data Mining.*, Pittsburgh, PA USA, June 2010.

Lee Becker, Rodney D. Nielsen, and Wayne H. Ward. What a pilot study says about running a question generation challenge. In *The 2nd Workshop on Question Generation*, Brighton, England, July 2009.

Rodney D. Nielsen, Lee Becker, and Wayne H. Ward. TAC 2008 CLEAR RTE system report: Facet-based entailment. In *Proceedings of the Text Analysis Conference*, Gaithersburg, Maryland, November 2008. National Institute of Standards and Technology.

Patents

Analyzing Document Revisions to Assess Literacy, Lee Becker, Jan C Zawadzki, Hapara, Inc., US Patent Pending, No. 14/643,678, filed March 10, 2015

Recommending Literacy Activities in View of Document Revisions, Lee Becker, Jan C Zawadzki, Hapara, Inc., US Patent Pending, No. 14/643,685, filed March 10, 2015

Determining Author Collaboration from Document Revisions, Lee Becker, Jan C Zawadzki, Hapara, Inc., US Patent Pending, No. 14/643,690, filed March 10, 2015

Impact Prediction of Social Media Interaction, David Skiba, George Erhart, Lee Becker, Avaya, Inc., US Patent 20150172145, filed December 18, 2013, Issued June 18, 2015

Natural Language Processing (NLP) And Natural Language Generation (NLG) Based On User Context For Enhanced Contact Center Communication, Reinhard Klemm, George Erhart, Lee Becker, David Skiba, Avaya, Inc., US Patent 20150163358, filed December 11, 2013, Issued June 11, 2015

Anti-Trending, David Skiba, George Erhart, Lee Becker, Valentine C. Matula, Avaya, Inc., US Patent 20150161216, filed December 11, 2013, Issued June 11, 2015

Aggregated Multi-Topic Agent Desktop, David Skiba, George Erhart, Lee Becker, Valentine C. Matula, Avaya, Inc., US Patent 20150139415, filed November 19, 2013, Issued May 21, 2015

Pseudo Agent Matching, David Skiba, George Erhart, Lee Becker, Avaya, Inc., US Patent 20150139416, filed November 20, 2013, Issued May 21, 2015

Deep Language Attribute Analysis, David Skiba, George Erhart, Lee Becker, Valentine C Matula, Avaya, Inc., US Patent 20150134325, filed November 14, 2013, Issued May 14, 2015

Dialog Compatibility, David Skiba, George Erhart, Lee Becker, Valentine C Matula, Avaya, Inc., US Patent 9020135, filed September 10, 2013, Issued April 28, 2015

Unspoken Sentiment, David Skiba, George Erhart, Lee Becker, Valentine C Matula, Avaya, Inc., US Patent 20150073775, filed September 11, 2013, Issued March 12, 2015

Automatic Domain Sentiment Expansion, Lee Becker, David Skiba, George Erhart, Valentine C Matula, Avaya, Inc., US Patent 20150073775, filed September 11, 2013, Issued March 12, 2015

Dialog Reorder, David Skiba, George Erhart, Lee Becker, Valentine C Matula, Avaya, Inc., US Patent 8964962, filed October 1, 2013, Issued February 14, 2015

Cross-Domain Topic Expansion, David Skiba, Valentine C. Matula, Lee Becker, George Erhart, Avaya, Inc., US Patent 20150006460, filed June 27, 2013, Issued January 1, 2015

Semantic Translation Model Training David Skiba, George W. Erhart, Lee Becker, Valentine C. Matula, Avaya, Inc., US Patent 20150006143, filed June 27, 2013, Issued January 1, 2015

Cross-lingual Seeding of Sentiment David Skiba, George W. Erhart, Lee Becker, Valentine C. Matula, Avaya, Inc., US Patent 20140303981, filed September 9, 2013, Issued October 9, 2015

Social Media Provocateur Detection and Mitigation David Skiba, Valentine C. Matula, Lee Becker, George Erhart, Avaya, Inc., US Patent 20140304343, filed October 7, 2013, Issued October 9, 2014

Automatic Negative Question Handling David Skiba, Valentine C. Matula, Lee Becker, George Erhart, Avaya, Inc., US Patent 20140304343, filed September 6, 2013, Issued October 2, 2014

Adaptively Presenting Content Based on User Knowledge, Sumit Basu, Lucy Vanderwende, and Lee Becker, Microsoft Corporation, US Patent 20130157245, Filed December 15, 2011, Issued June 20, 2013

Random code generation using genetic algorithms, Zachary Smith, Lee Becker, and David Heckman, Hewlett-Packard Company, US Patent 6918098, Filed July 16, 2002, Issued July 12, 2005

Professional Activities

- Organizer Front Range NLP Meetup 2013-Present
- Member Association for Computational Linguistics(ACL), 2009-present
- Member Association for the Advancement of Artificial Intelligence (AAAI), 2011-present
- Member International Artificial Intelligence in Education Society, 2009-present
- Reviewer Google Science Fair 2015 (brought in as expert in NLP and machine learning)
- Reviewer NLP BEA 2015: The 10th Workshop on Innovative Use of NLP for Building Educational Applications, June 04, 2015
- Reviewer NLP BEA 2014: The 9th Workshop on Innovative Use of NLP for Building Educational Applications, June 26, 2014
- Reviewer FLAIRS 2012: The 25th Florida Artificial Intelligence Researchers Society Conference (Applied NLP Track), May 23-25, 2012, Marco Island, Florida.
- Reviewer Discourse and Dialogue: Special Issue on Question Generation, 2011
- Reviewer ACL 2011: The 49th Conference of the Association of Computational Linguistics in the area of Natural Language Processing Applications, June 19-24, 2011, Portland, OR.
- Reviewer FLAIRS 2011: The 24th Florida Artificial Intelligence Researchers Society Conference (NLP Track), May 18-20, 2011, Palm Beach, Florida.
- Program Committee Third Workshop on Question Generation, 2010, Pittsburgh, Pennsylvania.
- Program Committee The Fifth International Workshop on Question Generation, 2012, Intelligent Tutoring Systems, Chania, Crete, Greece.

Languages

775 37th Street • Boulder, CO 80303

☎ 720.382.8688 • ✉ lee.becker@gmail.com • <http://www.leebecker.com>

5/6

Indonesian **intermediate**

*language classes taken at the University of Colorado and Wisma Bahasa,
Yogyakarta, Indonesia
4 years of study*

Spanish **basic**

References

Rowan Wing Data Scientist, Hapara

Dr. Martha Palmer Professor Linguistics, Computer Science, Institute of Cognitive Science, University of Colorado Boulder - martha.palmer@colorado.edu

Dr. Wayne Ward Principal Research Scientist - Boulder Language Technologies and Research Professor Computer Science, University of Colorado Boulder - wward@btek.com

Dr. James Martin Professor Computer Science, Linguistics, Institute of Cognitive Science, University of Colorado Boulder - james.martin@colorado.edu