Russell Horton : Curriculum Vitae

300 Bothin Rd, Fairfax, CA 94930 · 415 787 2281 · h@hhh.hm
https://www.linkedin.com/in/linguistics/

EDUCATION

2012 - *MA Linguistics*, The University of California, San Diego
2002 - *AB Linguistics*, The University of Chicago, *Honors, Phi Beta Kappa*

EMPLOYMENT

GitHub
Senior Security Analyst, Platform Health Team
2018 - present

- Lead the full lifecycle of abuse risk assessment and mitigation for numerous products, including Actions, Codespaces and Copilot. Consult on product design for abuse resistance, set requirements for telemetry, build and ship rules and models to flag abuse, train account review analysts on abuse patterns.

- Work closely with product teams, Engineering, Legal, Trust and Safety, OpSec, SIRT, Support, Billing and other teams to build and operate an integrated abuse response. Abuse types include commercial spam, data exfiltration, hate speech, billing fraud, theft of resources, account take overs, malware, phishing, student account fraud, impersonation, coordinated inauthentic activity, and more.

- Built, shipped and maintain our team's machine learning service as well as most production models for abuse classification. Python, Flask, sklearn.

Verizon OnCue
2016 - 2017 *Lead Data Scientist*
2015 - 2016 *Data Scientist*

- Led team of data scientists and engineers building a recommender system for TV and movies backed by viewing data from millions of customers on a catalog of hundreds of thousands of items. Spark, Scala, Python.

- Invented and prototyped data-driven TV products. Built system to identify entities and concepts in closed-caption data and link them to Wikipedia-derived concept graph.

EcoHealth
Senior Data Scientist
2014 - 2015

- Led cross-functional team developing emerging disease detection dashboard, funded by Defense Threat Reduction Agency grant. Intensive text analysis and proposition extraction from news reports and social media text fed machine learning pipeline to identify anomalies in human and animal health. Worked closely with epidemiologists and public health scientists to identify textual markers of potential emerging pandemics. Python, Scikit-learn, MeteorJS, Scala, Stanford NLP.

- Wrote Annie (https://github.com/ecohealthalliance/annie): A Python NLP annotation framework that creates a unified representation of diverse annotations such as tokenization, ngrams, POS tagging, entity linking, time and date recognition, proposition extraction, etc.

Reverb / Wordnik
Computational Linguist
2008 - 2014

- Built entity / concept disambiguation framework that formed the core document representation for our recommender system for news and web content. Scala, Python.

The University of Chicago
Programmer Analyst
2004 - 2009

- Machine learning and text mining, full-text search, retrieval and display.
- Built Philomine, an early application for machine learning on natural language text, allowing humanities researchers to perform classification and clustering experiments using a variety of algorithms via a simple web interface.

PUBLICATIONS

with Mark Olsen and Glenn Roe, "Something Borrowed: Sequence Alignment and the Identification of Similar Passages in Large Text Collections", Digital Studies / Le Champ numérique Vol 2, No 1, 2010.

with Timothy Allen, Stéphane Douard, Charles Cooney, Robert Morrissey, Mark Olsen, Glenn Roe, and Robert Voyer, "Plundering Philosophers: Identifying Sources of the Encyclopédie", Journal of the Association for History and Computing, vol. 13, no. 1, Spring 2010.

with Shlomo Argamon, Jean-Baptiste Goulain, and Mark Olsen, "Vive la Différence! Text Mining Gender Difference in French Literature", Digital Humanities Quarterly, Spring 2009, Volume 3, Number 2.
with Robert Morrissey, Mark Olsen, Glenn Roe, and Robert Voyer, "Mining Eighteenth Century Ontologies: Machine Learning and Knowledge Classification in the Encyclopédie", Digital Humanities Quarterly, Spring 2009, Volume 3, Number 2.

CONFERENCE PAPERS

with Les Henderson, "Sequence Alignment and Similarity in Biology and the Humanities", Digital Humanities and Computer Science 2010, Northwestern University, Novemeber 20th 2010

with Mark Olsen and Glenn Roe, "PAIR: Pairwise Alignment for Intertextual Relations", Annual Meeting of the Society for Digital Humanities -- Société pour l'étude des médias interactifs - Carleton University, Ottawa, May 25-27, 2009.

with Charles Cooney, Mark Olsen, Glenn Roe, and Robert Voyer, "Deconstructing Machine Learning: A Challenge for Digital Humanities", Digital Humanities 2008, University of Oulu, Oulu, Finland, June 25-29, 2008

with Charles Cooney, Mark Olsen, Glenn Roe and Robert Voyer,"PhiloMine: An Integrated Environment for Humanities Text Mining", Digital Humanities 2008, University of Oulu, Oulu, Finland, June 25-29, 2008

with Charles Cooney, Mark Olsen, Glenn Roe, and Robert Voyer, "Hidden Roads and Twisted Paths: Intertextual Discovery using Clusters, Classifications, and Similarities", Digital Humanities 2008, University of Oulu, Oulu, Finland, June 25-29, 2008

with Charles Cooney, Mark Olsen, Glenn Roe, and Robert Voyer, "Feature Creep: Evaluating Feature Sets for Text Mining Literary Corpora", Digital Humanities 2008, University of Oulu, Oulu, Finland, June 25-29, 2008

with Charles Cooney, Robert Morrissey, Mark Olsen, Glenn Roe, and Robert Voyer, "Re-engineering the tree of knowledge: Vector space analysis and centroid-based clustering in the Encyclopédie", Digital Humanities 2008, University of Oulu, Oulu, Finland, June 25-29, 2008

Shlomo Argamon, Jean-Baptiste Goulain, Russell Horton and Mark Olsen, "Discourse, power and écriture féminine: Text mining gender difference in 18th and 19th century French literature", Digital Humanities 2007, University of Illinois, June 2007

with Robert Morrissey, Mark Olsen, Glenn Roe and Robert Voyer,"Mining Eighteenth Century Ontologies: Machine Learning and Knowledge Classification in the Encyclopédie", Digital Humanities 2007, University of Illinois, June 2007.

with Charles Cooney, Mark Olsen, Glenn Roe, and Robert Voyer, "Extending PhiloLogic", Digital Humanities
2007, University of Illinois, June 2007

INVITED PRESENTATIONS

with Hu, Diana, "Scaling Verizon IPTV Recommendations with Scala and Spark" (slides). RecSys Large Scale Recommender Systems Workshop, RecSys 2016, MIT, Cambridge, Massachusetts.

"'Ever since nineteen, had a perfect rhyme scheme': A corpus study of English rap rhyme" Networks and Network Analysis for the Humanities: Reunion Conference, The University of California at Los Angeles, Oct 20 - 22 2011.

with Mark Olsen. Sequence Alignment, Shared Services, and Digital Humanities, Project Bamboo Workshop, Tucson, Arizona, January 2009.

with Robert Morrissey, "The ARTFL Project: From words to works", The Dilemmas of Digitization, Oxford University, May 22-24, 2008.

PATENTS

with Hu, Si Ying D. and Medapati, Suri B, "Automatic Media Summary Creation Systems and Methods" -- United States 20150300, March 2015

with Allen, Timothy and McKean, Erin, "Data Mining for Free Range Definitions" -- United States WORD 1000- 1, February 2011