Justin Sybrandt

Ph.D. Candidate

Email: Justin@Sybrandt.com Website: Justin.Sybrandt.com
GitHub: JSybrandt Phone: (484) 354-8692

Objective Putting unstructured information to good use remains a challenge, especially regarding scientific text. With every day comes thousands of newly published papers, which can lead to missed findings and, in turn, delayed progress. In order to aid researchers, my work involves understanding this information through text-based and graph-based machine learning methods in my thesis project: Moliere.

Education

2016-Present, Clemson University

- Ph.D. in computer science. (GPA:4.0)
- Focus on machine learning, text mining, and literature-based discovery.

2012 - 2016, Grove City College

- BS in computer science, minor in mathematics.
- Graduated Summa Cum Laude with a GPA of 3.85.
- Top of class in computer science with a major GPA of 3.95.

Honors and Awards

- Member of Upsilon Pi Epsilon CS Honor Society.
- Recipient of the GAANN DAISE PhD. fellowship.
- Recipient of a KDD'17 travel award.
- Member of the Kappa Mu Epsilon National Mathematics Honor Society.
- Member of the Alpha Tau Mu chapter of Mortarboard.
- President of the Grove City Chapter of the ACM.

Publications

Are Abstracts Enough for Hypothesis Generation?

- In Submission.
- Explored the effect corpus size and quality have on topic-based HG systems.
- Found full-text documents offer a marginal improvement, increase runtime by $45\times$

Large Scale Validation of Hypothesis Generation Systems

- In Submission.
- Developed new mathematical models to anticipate strong HG connections.
- Devised a method to validate hypothesis generation systems at large scale.

MOLIERE: Automatic Biomedical Hypothesis Generation System

- Accepted for oral presentation at KDD 2017 (Acceptance Rate 8.8%).
- Data mining 24.5 million medical documents to form a large multi-modal network.
- Models hypotheses using Latent Dirichlet Allocation.
- Awarded "Best Final Project and Poster Presentation" in Data Science student conference.

Rapid Replication of Multi-Petabyte File Systems

- Presented at PDSW 2015 as a Work In Progress.
- Presented at the ACM student poster session at SC15.
- Built Distsync, a distributed tool capable of replicating large GPFS file systems.
- Deployed system with NERSC to facilitate large, high speed data transfers.

Development Skills and Technologies

- C++
- SQL
- Linux
- Python
- C
- VIM
- Java
- Scala
- LaTeX
- Bash
- Julia
- Git

Experience

Summer 2018, Ph.D. SWE Intern, Google

- Designed a graph-mining solution for identifying product attributes.
- Implemented solution at scale via Google-internal distributed systems.
- Performed comprehensive validation, ensuring classifier performance across product categories.

Summer 2017, Graduate Research Assistant - Los Alamos National Lab

- Developed high performance software in julia for non-negative matrix factorization.
- Extended MOLIERE to water resources research with the computational environmental science group (EES-16).

2015-2016, Programming Intern - Vigilant Cyber Systems, Inc.

- Developed a visualization library in Scala using ScalaFX.
- Independently managed time when working remotely.
- Balanced senior-level course work with development.

Summer 2015, Student Researcher - UC Berkeley & NERSC

- Designed and implemented a tool to quickly synchronize multi-petabyte General Parallel File Systems.
- Presented a poster at the ACM Student Poster Session at SC'15.
- Presented a work in progress paper at the Parallel Data Storage Workshop.

Summer 2014, Student Researcher - Grove City College

- Added distribution preferences to Data Stream Management Systems (DSMS).
- Simulated new DSMS features in Python.
- Studied modern DSMS through recent research papers.
- Presented a poster at the Grove City student poster session.

2012-2014, Programming Intern - Gravic Inc.

- Worked on a six person team developing tools for administering exams.
- Collaborated with corporate partners to design features which allowed our products to share data.
- Gained familiarity with project management while extending the Remark VB.NET code base.

Research Interests

- Machine Learning
- Text Mining
- Hypothesis Generation
- Natural Language Understanding
- Graph Mining
- Artificial Intelligence

Teaching Experience

- Spring 2018, Guest Lecture: Applied Data Science
- Fall 2017 - Spring 2018, Project Manager: Seminar in Professional Issues II