

# Sarah Wiegreffe

Atlanta, GA  
swiegreffe6@gatech.edu  
sarahwie.github.io  
(843) 708-2133

---

Seeking a summer 2018 research internship in natural language processing.

## EDUCATION

---

**Georgia Institute of Technology (Georgia Tech)** **August 2017 - present**  
**PhD in Computer Science**  
**Advisors: Drs. Jacob Eisenstein and Jimeng Sun**  
GPA: 3.50/4.00

Research interest in natural language processing with applications in the clinical setting. Specific interests include machine reading, hierarchical representations and information extraction, latent variable models, and text generation.

**Honors College at the College of Charleston** **August 2013 - May 2017**  
**Bachelor of Science in Data Science, Summa Cum Laude**  
Minors in Mathematics and International Studies, Cognate in Economics  
GPA: 3.97/4.00

**University of Tartu, Estonia** **January - June 2015**  
Visiting Student, Faculty of Mathematics and Computer Science  
GPA: 4.75/5.00

## RESEARCH

---

**Computational Linguistics Lab, Sunlab at Georgia Tech** **August 2017 - present**

Researching latent variable generative models of text for the capturing of long-range dependencies. Developing methods for the extraction of meaningful textual information from hospital clinical notes for the prediction of patient diagnosis and treatment.

**Anderson Lab at the College of Charleston** **January 2016 - May 2017**

### **Project 1: Senior Thesis**

Researched extensions to Google's Word2Vec algorithm used to generate word embeddings for variable-length documents. Investigated performance of the algorithm when used directly as a classifier, and whether this technique, along with similarly created ensemble methods, could outperform benchmark preprocessing and machine learning pipelines on topic recognition tasks.

**Project 2: Joint with the Medical University of South Carolina Biomedical Informatics Center**

Researched automation of the process of identifying translational research patterns in medicine. Built datasets of unlabelled medical abstracts and trained machine learning algorithms to classify by translational type.

**Medical University of South Carolina  
Office of the Chief Informatics Officer**

**June - December 2015**

Designed a novel model forecasting the number of operating room cases in the university hospital on a given day, in order to improve hospital efficiency. Implemented the model as an automated visualization tool using Tableau which generates daily 30-day forecasts for departments across the hospital to assess.

**Institute of Computer Science, University of Tartu, Estonia**

**Spring 2015**

Developed a high-performing model to predict a subject's age by MRI brain scan. Implemented feature reduction via LASSO regression; identified brain volumes of note utilizing a hierarchy of features.

**PAPERS AND PRESENTATIONS**

---

Submission to NAACL 2017.

Wiegreffe, S., Anderson, P. and Obeid, J. *Can Classifications of Publications by Translational Categories be Automated?* Proceedings of the 2017 American Medical Informatics Association Joint Summits on Translational Science. Poster presentation, San Francisco, CA, March 2017.

Wiegreffe, S. and Anderson, P. *A Survey of Word2Vec Inversion Methods in Topic Recognition Tasks*. Bachelor's thesis. Presented at the College of Charleston, May 2017.

Talk and poster presentations on context-specific distributed language representations for sentiment classification at the College of Charleston Department of Computer Science and School of Science and Mathematics research symposiums (2016, 2017).

**AWARDS**

---

**CRA-W Grad Cohort for Women Invitee** (spring 2018).

**Phi Kappa Phi Graduate Fellowship** (2017).

**Data Science Major of the Year, Departmental Honors** (2017). College of Charleston.

**Grace Hopper Scholar**, the Anita Borg Institute (2016). Grace Hopper Celebration of Women in Computing Attendee (2015, 2016).

**William Aiken Fellow** (2013-2017). A fellowship representing the top 1% of students at the College of Charleston.

## PROFICIENCIES

---

**Languages** Python, SQL, Bash, Java, R, SAS, Octave. Fluency in French.

**Tools** scikit-learn, nltk, Pytorch, Dynet, Git, TeX, Tableau, Oracle RDBMS, MongoDB

## TEACHING AND EMPLOYMENT

---

**Graduate Research Assistant, Georgia Institute of Technology** **Fall 2017 - present**

**Teaching Assistant, Charleston Digital Corridor** **Fall 2014**  
MongoDB database course for adult professionals.

**Math Lab Tutor, College of Charleston Center for Student Learning** **Fall 2014**  
Subjects including calculus, statistics, business math, precalculus and algebra.

**Medical University of South Carolina Biomedical Informatics Center** **Summer, Fall 2014**  
Used relational database and SQL querying skills to test and improve the data warehouse for South Carolina's hospitals. Worked in a team environment to develop a testing process and write scripts to test web applications against the database.

**Women in Computing Club at the College of Charleston** **2014 - 2017**  
Club president (2017), Vice-president (2016) and Treasurer (2015-2016). Worked to promote diversity in computer science, host professional development workshops, and conduct community outreach.

## OTHER PROJECTS

---

***Identifying Anonymous Rules on Social Media*** **Georgia Tech, Fall 2017**  
A linguistic analysis of anonymous social media sites using Twitter data.

***DocRank*** **College of Charleston, Spring 2017**  
A medical provider ranking algorithm and webapp, using Coley's algorithm, for healthcare API company PokitDok.

***Migration Trends*** **College of Charleston, Spring 2016**  
A RShiny visualization tool for tracking refugee migration patterns using data from the UNHCR.