**James A. Scharf**

22 Garey Drive | Chappaqua, NY 10514 | Mobile: (914) 522-2930 | [jscharf8@jhu.edu](mailto:jscharf8@jhu.edu)

**EDUCATION**

**The Johns Hopkins University** – Baltimore, MD *Expected* *May 2021*

* Bachelor of Arts, Political Science
* Minor, Computer Science

**Horace Greeley High School**– Chappaqua, NY *June 2017*

**RESEARCH EXPERIENCE**

**Center for Language and Speech Processing (CLSP), Johns Hopkins University** – Baltimore, MD *Summer 2020 – Present*

*Dr. David Yarowsky’s Low-Resource Languages Lab ([LoReLab](https://www.cs.jhu.edu/~arya/yarowsky-lab/))*

* Utilized Facebook’s Fairseq Sequence-to-Sequence (Seq2Seq) library for the automatic segmentation of letter clusters
* Currently implementing a sequence-tagging approach to multilingual syllabification with a Conditional Random Field (CRF)
* Scraping and parsing foreign language webpages into structured data with Pandas, Selenium and BeautifulSoup
* Aided in the creation of the largest dataset of pronouns in the world in 1000+ languages *(publication pending submission)*
* Created visualizations for pronoun pro-drop data with Seaborn and Matplotlib
* Proposed a novel Universal Grapheme-to-Phoneme system to translate orthography to the International Phonetic Alphabet
* Normalizing datasets for downstream usage
* Attended tutorials and workshops at [ACL 2020](https://acl2020.org/) about the computational identification of media bias, fact verification, machine translation and feature representation

**Department of East Asian Studies, Johns Hopkins University** – Baltimore, MD *Summer 2020 – Present*

*Computational Research Assistant for Dr. Giovanna Maria Dora Dore*

* Extracted statistical information regarding word usage in 700 news articles about Hong Kong protests
* Generated plots with Seaborn and Matplotlib to compare news coverage by year
* Cleaned and combined news datasets with Pandas

**Department of Computer Science, Johns Hopkins University** – Baltimore, MD *Fall 2019*

*Dr. David Yarowsky’s Computer Science Workshop*

* Member of graduate and undergraduate team to build a “universal dictionary” of 1000+ languages
* Used HTML5, JavaScript (Node.js and client-based) and SQL to build a mobile web app to collect low resource language data
* Resulted in a poster presentation at the UNESCO Language Technologies for All (LT4All) conference

**Department of Political Science, Johns Hopkins University** – Baltimore, MD *September 2018 – October 2019*

*Research Assistant for Prof. Adam Sheingate*

* Used Python and R to analyze F.E.C. data related to elections from 2010 to 2016 to determine campaign expenditure patterns; created necessary visualizations for presentations
* Built natural language classifier using the Scikit-learn machine learning framework, pattern matching and query-expansion
* Authored extensive documentation to communicate methodology and provide accountability
* Submitted a research note to the Electoral Studies journal

**PUBLICATIONS AND PRESENTATIONS**

David Yarowsky, Arya D. McCarthy, Garrett Nicolai, Winston Wu, Aaron Mueller, Dylan Lewis, Yingqi Ding, Abhinav Nigam, Emre Ozgu, Debanik Purkayastha, James Scharf and Kenneth Zheng, “[A 1000-language Collaborative Universal Dictionary and Universal Translator](https://www.cs.jhu.edu/~arya/yarowsky-lab/#research-interests),” presented at the UNESCO Language Technologies for All (LT4All) conference

*Publication submitted and pending review in Election Studies for my work with Dr. Sheingate, as described above.*

**EMPLOYMENT EXPERIENCE**

**Department of Computer Science, Johns Hopkins University** – Baltimore, MD *Spring 2020*

*Course Assistant for Information Retrieval and Web Agents*

* Hired assistant for a course in machine learning, information extraction and web scraping
* Helped students during office hours with projects such as sense disambiguation classifiers, word embeddings and vector models of information retrieval
* Graded homework assignments

**The Johns Hopkins University News-Letter** – Baltimore, MD*October 2017 – January 2020*

*Staff Writer – News & Features*

* Covered news and events ranging from comedy shows to the rising crime rate and its impact on the college community
* Networked to find exclusive and primary sources for publication; have interviewed people such as Bill Miller, donor of $75 million to Dept. of Philosophy at Johns Hopkins
* Wrote for the News & Features, Science & Technology and Arts sections and routinely made the front page
* Articles are available at: <https://www.jhunewsletter.com/staff/james-scharf>

**ODN (formerly Open Data Nation) –** New York, NY *Summer* *2019*

*Data Science Intern*

* Programmed statistical analyses of roadway and car crash data with Pandas and Scikit-learn
* Created maps and visualizations using Geopandas, Geoplot, Seaborn and Matplotlib
* Cleaned and prepared data for machine learning models

**Town of New Castle –** New Castle, NY *Summer 2018*

*Intern*

* Researched possible additions to the Town Code by providing comparative memos about other towns’ rules and regulations
* Worked with database and municipal software

**Institute for Data Intensive Engineering and Science –** Baltimore, MD *Spring 2018*

*Intern – Web Development*

* Developed front-end web applications
* Redesigned educational astrophysics webpages using HTML5, JavaScript and WordPress

**Chappaqua Public Library –** Chappaqua, NY *2016 - 2017*

*Computer Page – Computer Skills Teacher and I.T.*

* Conducted one-on-one PC, Mac, Android, iPhone, and iPad teaching sessions with town residents
* Planned and organized Python coding workshops for youths
* Selected and installed new Windows 10 workstations

**INTERESTS & OTHER**

* **Computer Skills:** Proficiency in MS Office, R, HTML5/JavaScript, Python, C, C++, Java, Unix/Linux, CRM
* **Awards:** Eagle Scout
* **Language Skills:** Knowledge of Latin, basic Spanish (reading), Russian (reading)
* **Volunteering:** JHU Tutorial Project (Spring 2019), Vice President of Faculty Outreach JHU Political Science Steering Committee
* **Github:** <https://github.com/JamesScharf>
* **Professional Membership:** Association for Computational Linguistics (ACL)