Kristen Sheets

(415) 233-3032 sheetskristen@gmail.com sheetskristen.github.io

EDUCATION

M.S. in Computational Linguistics 2018 - 2020
Brandeis University; Waltham, MA
Coursework: Statistical & Neural Methods for NLP, Computational Semantics, Linguistic Annotation for Machine Learning, Data Structures & Algorithms, Information Retrieval, NLP Systems for Named Entity Recognition
Teaching assistantships: Linguistic Typology, Syntactic Theory (*for Prof. Lotus Goldberg*)

B.A. in Linguistics 2010 - 2014
University of California, Santa Cruz; Santa Cruz, CA
Honors: Dean's Undergraduate Research Award, Dean's Honors
Teaching assistantships: Syntactic Structures (*for Prof. Maziar Toosarvandani*)

SKILLS

Expertise:	Natural Language Processing, Machine Learning, Language Data Annotation, Linguistics
Programming:	Python (NLTK, PyTorch, pandas, sklearn, spaCy, Tensorflow), Java, Haskell, Bash, SQL
Tools:	CoreNLP, Elasticsearch, Neo4J, OpenFST, Thrax, Tesseract, Wikidata
Natural languages:	English (native), German (fluent), French (proficient)

WORK EXPERIENCE

Spotify 06/2020 - 08/2020
Linguistics, PhD Intern; Boston, MA
- Designed and built a semantic network representing lexical relations and domain knowledge between music descriptors.
- Wrote grammars and built FSTs for parsing natural language utterances for query fulfillment.

Lab for Linguistics and Computation 09/2018 - 08/2020
Graduate Research Assistant; Brandeis University, Waltham, MA
- Researched inherent polysemy and created a semantic type classifier, supervised by Prof. James Pustejovsky.

Charles River Analytics 02/2020 - 05/2020
Natural Language Processing Intern; Cambridge, MA
- Created a custom entity and relationship extraction pipeline for use on scientific and technical papers, includes OCR pipeline, design and implementation of symbolic functional grammar systems for entity parsing and extraction.

Basis Technology 06/2019 - 08/2019
Linguistic Data Intern; Cambridge, MA
- Created process for natural language processing tool evaluation and executed on a large comparison project in collaboration with R&D.
- Sourced novel multilingual data sets leveraging Wikipedia, census data and other publicly available resources, augmenting and writing custom data processing scripts.

Ericsson Emodo *(formerly Placecast)* 01/2016 - 07/2018
Solutions Engineer, Technical Analyst; San Francisco, CA
- Served as a technical expert to external stakeholders for custom integrations, including Fortune 100 enterprise partners.
- Designed and automated data integration by writing and maintaining process controllers for data ingestion and analysis.

Google 07/2014 - 01/2016
Localization Project Coordinator; Mountain View, CA
- Managed end to end global localization production, including triaging functional and linguistic issues for multiple product lines including Gmail, Google Apps for Work, and Google Play across 68 languages.
- Trained new Localization Project Coordinators, managing ramp and development for 5 new team members.

SELECTED PROJECTS

Named Entity Recognition for Podcasts
Developed a custom ontology, extending from the OntoNotes NER annotation guidelines and trained two models, a linear CRF and a neural CNN-biLSTM-CRF model for named entity recognition from podcast transcripts.

Semantic Type Classifier
Using the WordNet annotated SemCor corpus, designed and trained a classification model using contextual word embeddings (BERT) to identify a word's CoreLex basic semantic type, as a proof of concept for identifying trends in polysemy.