Elliott Indiran

Computational Linguist

github, linkedin: eindiran
site: https://eindiran.github.io/

eindiran@uchicago.edu
(805) 651-9333

Education

University of Chicago — Chicago, Illinois
Bachelor of the Arts in Linguistics with General Honors — 2011 - 2015

- 3.44 Cumulative GPA; 3.67 Fourth-Year GPA
- Relevant coursework: Introduction to Programming for Linguistics, Statistical Methods and Applications, Mathematical Modelling for Linguistics, Computer Science with Applications sequence, Computational Linguistics. Linguistics coursework in all major subfields (syntax, semantics, phonetics, phonology, morphology, psycholinguistics).
- Background in experimental design, creation and management of linguistic corpora, and analysis of natural language data.

Linguistic Society of America — Chicago, Illinois
Summer Institute at the University of Chicago — 2015

- Relevant coursework: Exploiting Web Data for Linguistic Research, Unsupervised Learning of Linguistic Structure, Advanced Probabilistic Modeling in R, Speech Technologies, Computational Psycholinguistics, Computational Lexical Semantics.

Work Experience

Computational Linguist at Promptu Systems — Menlo Park, California
Promptu Systems Corporation — October 2015 - present

- Worked on speech recognition projects in English, Italian, German, and Mandarin; both as a software engineer, writing components of the recognizer system, and as a linguist, handling language modeling and writing lexicons and grammars. This included constructing language model training datasets products in several significantly different problem domains.
- Wrote the component of the recognizer system responsible for handling lexicon lookups during entity registration, generating pronunciations wherever possible to avoid needing to perform G2P on out-of-vocab entries. With pre-G2P in place, the number of entities that needed G2P on first registration step dropped drastically, shaving seconds off of the recognizer's start-up latency.
- Led project on constructing an automated build system for language model training data, using CI techniques to build and test LMTDs, substantially reducing the frequency of bugs and regressions in the recognizer related to the language model and keeping over a dozen projects up-to-date while not under active development.
- Wrote and maintained a large number of developer tools for use in team; most notably I wrote a fully-featured Java application for recording speech. The recording app needed a lot of special features, so I designed, wrote and tested it entirely from scratch (using the Java Sound API). The app was designed to be useful for recording scripts of phrases, either in a car or in-office.
- Experience maintaining and refactoring codebases in C, Perl, Python and Java.
- Led a project to speed up the construction on language models. Rewrote the piece of software which was the bottleneck, reducing LM build times by as much as 20x for some datasets. Further decreased the build time of the LMTD by parallelizing the whole process, reducing build times by hours per day.
- Drafted a proposal for changing the build process of a major software component which allowed a substantial number of developer hours to be saved each time a new Acoustic scorer was built, saving upwards of several man-weeks per year.
- Wrote an Android app used internally for data collection and labeling of audio files.

Intern with the Linguistic Society of America Summer Institute Chicago, Illinois
Linguistic Society of America (LSA) June 2015 - August 2015

- Wrote the software responsible for managing grades and credit assignment at the institute, which was used to process class data for hundreds of participants across nearly 100 classes.
- Aided the directors of the institute in all technical matters; led tech support during registration.
- Managed a critical registration site for institute participants, ensuring that technical problems did not lead to issues for participants checking in.

Projects

Sitrus team-sitrus.github.io
2016 - present sitrus.io/app/

- Co-founded Sitrus with 2 friends, a platform with a number of tools for quantified-self data collection and analysis. We built out the tools with the goal of lowering the barrier-of-entry to experiment with quantified-self.
- Wrote the Paragon data collection tool, used to capture a user's data through an easy-to-use web interface in React + Django.
- Wrote the Sitrus query language, used as a frontend for querying and processing data from a relational database. The language can be with Postgres or SQLite, with the interpreter generating PostgreSQL and SQLite queries respectively.
- Developed a site that synthesizes quantified-self data from a variety of sources, allowing the user to unsilo their data and do data analysis and visualization for all of their data in one place. It includes integrations with a number of other apps including GitHub, Wakatime, RescueTime, Steam, FitBit and more.
- Designed a variety of data visualizations for the main quantified self dashboard.
- Wrote a database abstraction layer that allowed the backend to swap databases painlessly, supporting SQLite, Postgres, and MongoDB.
- Led Sitrus through Y Combinator's 2018 Startup School, graduating in November 2018. Elected as moderator of our group of 15 companies. Organized all tasks, meetings, homework, and discussions for our section.
- Developed the Sitrus Moods application for Android and iOS in React Native, available in the App Store and the Play Store.

Programming Language Projects More Info:
2017 - present github.com/eindiran

- Frequent contributor to a few Open Source projects, including the Coconut Programming Language and related tooling. Wrote large sections of the Coconut tutorial and the main Coconut documentation.
- pyknitout: A Python package that adds support for Knitout, a format for providing algorithmic knitting instructions to mechanical looms/knitting machines, to Python, allowing you to create Knitout files using very high level Python code.
- uforth: A small bootstrapping Forth system with a base written in x86-64 assembly and the nasm assembler.
- urbane: A brainf*ck interpreter written in C, with support for other brainf*ck-based languages including reversef*ck, wepmlrio, ???, f*ckbees, and P".

Other Projects More Info:
2017 - present github.com/eindiran

- AutoWater: a project to automate watering your houseplants using Arduino.
- orgmode-to-anki: automatically convert orgmode files into Anki flashcards. Useful for automating the generation of flashcards from Orgzly or emacs notes.
- Mancala: play mancala in the terminal against another human or against an AI that uses a min-max algorithm to prune possibilities of the game tree.

- interject: View and edit text in shell pipes while they run.
- libperceptron: A lean, zero-dependency library for implementing fast perceptrons in C.

Technical Skills

- Have used Java, Python, and C professionally in production systems. I have built and maintained a large continuous delivery pipeline using Jenkins, scripting with Groovy. Familiar with *nix utilities for system administration tasks, on Linux and OS-X.
- Python for scripting, data scraping and manipulation, and backend web-development (using Django).
- Databases: SQLite, PostgreSQL, MongoDB.
- App Development: Android (mobile), React Native (mobile), and Java (desktop).
- Familiarity reading: Javascript, C++, Perl.