

TRISTAN E. CHONG

San Francisco, CA
linkedin.com/in/tristaneuan

ong@tristaneuan.ch
github.com/tristaneuan

Software engineer with a focus on natural language processing.

PROFILE

Resourceful programmer with excellent communication skills and an insatiable hunger for knowledge. Formally trained in linguistics and computational linguistics, with on-the-job experience in natural language processing, software scaling, and distributed systems. Talented debugger who excels at cleaning up messes and making code work.

WORK EXPERIENCE

BETTERCOMPANY, San Francisco, CA
Software Engineer

Jul 2014 - Nov 2014

- Implementing app authentication via Facebook
- Extending a viral-loop event processor to extract and store information from user Facebook profiles and mobile phone contacts
- Porting the event processor from Node.js to Python
- Writing internal tools to interact with Amazon Web Services
- Creating an application that successfully identified prospective users for a given employer and geographical location
- Automating the sending of email invitations to targeted prospects based on criteria aligned with variable marketing strategies

WIKIA INC., San Francisco, CA

Jun 2013 - Jul 2014

Computational Linguist/Software Engineer, NLP

- Building a pipeline utilizing the Stanford CoreNLP software suite to parse tens of millions of pages of text
- Extending the capabilities of the parsing pipeline to scale efficiently and automatically on Amazon EC2 and S3, using distributed computing and concurrency principles
- Implementing components of a service-oriented architecture designed to extract and cache data on named entities, syntactic heads, coreference chains, dependency relations, and sentiment
- Writing ETL and load balancing modules in a Python library used for data science research
- Exploring various document summarization algorithms, and evaluating n-gram keyword extraction and sentiment analysis for potential business applications
- Designing a successful heuristic to infer the subject of a wiki using term frequency and weighted scoring
- Training latent Dirichlet allocation models with named entity data and using a distance metric to identify related pages as part of a recommendation system

FLUENTIAL INC., Sunnyvale, CA

Jan 2011 - Jun 2013

Computational Linguist

- Working as part of a team of linguists and engineers to develop machine translation software and spoken dialogue systems
- Writing context-free grammars in a variant of Backus-Naur Form for natural language parsing
- Testing and tuning support vector machines and semantic class taggers to achieve higher phrase classification accuracy
- Maintaining training and testing corpora, using Python and the Natural Language Toolkit to automate tasks such as crowdsourced data collection, text normalization, production of canonical forms via stemming, morphosyntactic operations, and creation of use cases for regression testing
- Writing interaction guides in YAML to manage dialogue states and conversation flow
- Localizing existing applications to different languages, regions, and target markets
- Generating and processing text-to-speech audio: silence trimming, noise removal, normalization, compression
- Translating TTS pronunciation dictionaries between proprietary formats and X-SAMPA

EDUCATION

M.S., Computational Linguistics, 2015

University of Washington, Seattle

- Graduated with a 3.74 GPA
- Studied speech technology, spoken dialogue systems, HPSG, HMMs, parsing, language modeling, smoothing
- Worked on projects involving POS taggers, context-free grammars and parsers, automatic summarization, grapheme-to-phoneme conversion, BNF+ ASR grammars

B.A., Linguistics and Anthropology, 2009

University of California, Los Angeles

- Graduated with a 3.54 GPA
- Earned a 3.91 GPA in Linguistics courses
- Recipient of National Merit Scholarship and Governor's Scholarship
- Studied phonetics, phonology, morphology, syntax, semantics, and pragmatics

PUBLICATIONS

Elwell, Robert, Tristan Chong, Kevin Cooney, and Chris Fife. "A Content-Based Recommendation System for Online Communities at High Scale."

Submitted to ACM Recommender Systems 2014

Elwell, Robert, Tristan Chong, and John Kuner. "A High-Scale Deep Learning Pipeline for Identifying Similarities in Online Communities."

Accepted by Taming Text, 2nd Edition

SKILLS

Python
AWS
Unix
HTML
English

NLTK
Apache Solr
Git
Regex
Spanish

Stanford CoreNLP
Django
Ruby
Vim
Mandarin

Gensim
Flask
SQL
Java
Ancient Greek