

Dave Farnham

linkedin: [linkedin.com/in/davefarnham](https://www.linkedin.com/in/davefarnham)

github: [dfarnham.github.io](https://github.com/dfarnham)

crates.io: crates.io/users/dfarnham

Statistical NLP, machine learning and computational linguistics developer specializing in custom language models, feature extraction, text analytics, parsing, classification and data modeling. I've implemented core IP for a variety of companies and products covering industries from Fortune 500 to startups in Education and Research (Essay Grading, Social Media Classification, Cross Language Informational Retrieval, Text Leveling, R&D Labs [AFRL, ARI]). An expert in Latent Semantic Analysis.

Laboratory for Atmospheric and Space Physics Research

Internal development for the IT department

University of Colorado Institute of Cognitive Science, consulting for Dr. Peter Foltz

Mathematics and ML framework used by lsa.colorado.edu

Pearson R&D, Cognitive Computing (AI Solutions), Data Scientist

Developed infrastructure libraries, API's, statistical analysis, and production models for both R&D technology pilots and company products.

- Infrastructure libraries for vector space embeddings & text processing. Feature extraction libraries written in Rust with Python interfaces.
- JSON API's over Docker deployed containers for R&D customer pilots.
- Complete rewrite & MLOps deployment of Pearson's production GSE (Global Scale of English) inference models (SciKit/Pandas/Rust) <https://www.pearson.com/english/about-us/global-scale-of-english.html>

Parallel Consulting, LLC (Remote to Petaluma CA), Data Scientist / R&D Architect

- Design & implementation of a custom feature extraction, data modeling & machine learning toolkit, command line tools & libraries for content assessment research.
- Technology components include an implementation of the proposed IETF JSON "ref" specification, Antlr (Java) grammar, Smile (Java), Weka, Docker, R, Groovy.

Oracle Corporation, R&D / Principal Member / Data Science

- Design & implementation of the high volume (100M+ daily) social media classification engine which performed customer specified NLP + Semantic matching in real time.
- Designed core NLP libraries supporting 20+ languages, core vector space library (Java/Jni/C++).

Collective Intellect, R&D (Acquired by Oracle)

- Start up, text processing infrastructure, embedded language models, NLP, statistical analysis.

Knowledge Analysis Technologies, Senior MTS (Acquired by Pearson)

- Start up / Boulder Incubator — Principal developer for IEA (Intelligent Essay Assessor).



Education

Colorado State University

Masters Program - Computer Science

BS - Computer Science, Minor - Mathematics