Static Code Analysis for Plain English

**Goal:** To harness the speed of a computer to check large volumes of text stored in different formats.

- Identify technical terms and obscure words
- Find dead URLs, missing citations, unused references
- Find simple grammatical mistakes, sections that are difficult to read, run-on sentences
- Show word usage frequency, common sentence structure
Who would use this tool?

- Maintainers of large documents like websites and technical references
- Curators of diverse documents such as research papers from several fields
- Students
- Anyone without access to an editor

How?

- Web interface for uploading documents and viewing statistics
- Command line tool for crawling large databases, executable from scripts and other tools
Architecture:

- Libraries – extract text from a wide variety of documents, interact with a database of references
- Interface
  - browser side (JavaScript, cookies, etc.) for cross platform support
  - server side (Java, Python, etc.) for well established external libraries

Challenges:

- Scope – flexibility vs. domain specificity
- User Friendliness – well formatted feedback, few false positives