Requirements & Eng Standards Lightning Talk

William Sengstock, Kelly Jacobson, Zach Witte, Austin Buller, Sam Moore, Dan Vasudevan, Jacob Kinser

Problem Statement

- **Project:** Mining and Evaluating Verb tags and Other Important POS tags inside Software Documentation.
- To improve Part of Speech (POS) tagging to make information mining more effective when applied to software languages.
- Plan to improve usage for software languages
    - Java
    - Python
    - C
    - etc.

Requirements & Constraints

Research Requirements/Constraints	Programming Requirements/Constraints
- Topics to research - Data pre-processing techniques - Word vectorization - Word embedding and word clustering - Algorithms - POS tagging techniques - Supervised/unsupervised Learning - Each member of the project is expected to research on their own and cite credible sources	- Models will be written in Python using Jupyter Notebook (constraint) - Research and experiment with different NLP models - Data used for research will come from approved sources, such as www.kaggle.com (constraint) - Evaluate the performance of each model compared to manual review

Engineering Standards

- ISO-IEC 9001: Quality Management
  - Helps ensure quality throughout the project, as well as continuous improvement
- ISO-IEC 830: Software Requirements Specifications
  - Influences structure of project, along with communication between users
- ISO-IEC 12207: Software Life Cycle Processes
  - Influences project design, maintains product throughout stages

Intended Users and Uses

- End goal is that we improve NLP for software documentation
- Improve recommendations for software documentation
    - Researching python language, recommend relevant python documents
    - Improving relevance features, finding similar documents
- Anyone looking for software documents could benefit from it
- Analyzing for similar code blocks across different sources