NLP for Aircraft Maintenance

Steven Li, Karen Wong, Kiko Whiteley

Sponsor: LtCol Thomas Kline - U.S. Marine Corps Forces Pacific

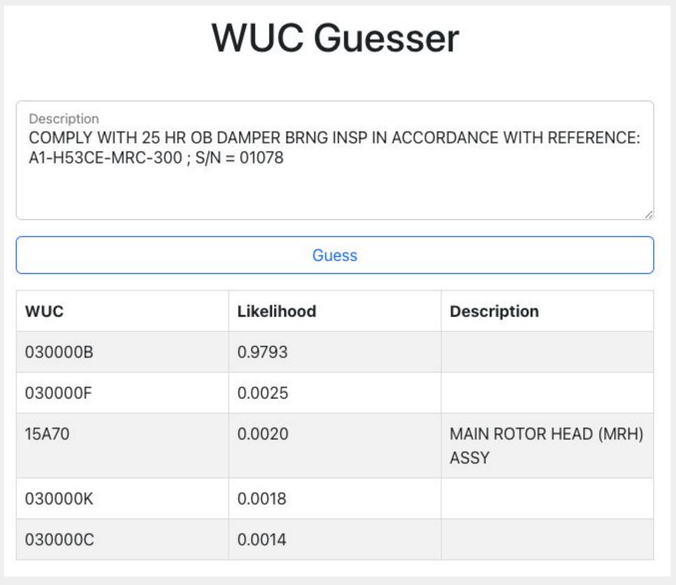
ICS 496
Fall 2022

BUSINESS PROBLEM

Need for increasing **accuracy** and **specificity** of maintenance data input by engineers as well as mitigating **barrier to entry**. Currently, Work Unit Codes ("WUCs") must be **manually chosen** by traversing a large tree, making some inputs **erroneous or too broad**. Existing processes are time-consuming to acclimate to and regularly use.

SOLUTION

A Python web application using natural language processing on historical data and provided WUC trees to convert user inputs into WUCs.



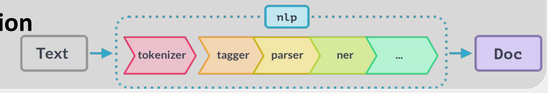
- Takes in user input description
- Predicts possible matching WUCs
- Displays the likelihood percentage along with the description of each WUC

INFORMATION & COMPUTER SCIENCES
UNIVERSITY *of* HAWAIʻI *at* MĀNOA

IMPLEMENTATION

Data Transformations

- Retrieve historical data text files, WUC trees, appendices and **convert them to data**
- **Remove anomalies** and use regular expressions and spaCy to do **mappings**
- Expand abbreviations, contractions, number to words
- **Lemmatization**
  - "CHARGING" ⟶ "CHARGE"
- **Tokenization**



Word Embeddings

fastText for word vector representation

- Considers context ⟶ "engine" and "generator" have similar vectors
- Come from neural networks
- Character n-grams learn word structure and "rare" words

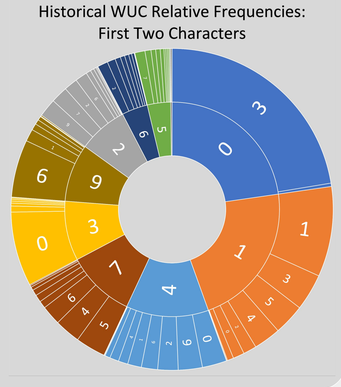
fastText's built-in classifier

- Binary tree classifier where each leaf is a WUC
- Issue: may not consider the hierarchical nature of WUCs

Custom Classifier

Weighted-average sequential, supervised

- Leverages **hierarchical structure** of WUC tree
- Allows for **fine-tuning** of relative steps' weights
- Partitions groups based on real-world similarities
- Sub-models utilize fastText's optimized supervised classifier



TECHNOLOGY STACK

NLTK, spaCy, fastText
Natural Language Processing

Flask
Back-End

HTML, JS, CSS
Web Front-End

CHALLENGES

- Abundance of different WUCs
- Some WUCs are more accurate than others
- Description variations based on user convention/personal preference
- High processing power demand to train models

NEXT STEPS

- Gather user feedback on working information
- Continue to train models for each Type Equipment Code (TEC)
- Deploy Docker containers
- Share with report