IDENTIFYING DISCIPLINE-SPECIFIC VOCABULARY ON ENGINEERING EXAMS

Chirag Variawa, Susan McCahan
University of Toronto
mccahan@mie.utoronto.ca

Abstract —The goal of engineering examinations is to measure the academic performance of our students with respect to learning outcomes. On exams, the students are often asked contextualized questions using vocabulary that might be unfamiliar.

A study is being conducted that investigates the accessibility of language on engineering exams with the goal of making language clearer for all students. Specifically, a Term-Frequency Inverse-Document Frequency (TF-IDF) algorithm is used to characterize words on a given engineering exam.

By comparing data across different test cases, the TF-IDF algorithm appears to accurately distinguish discipline-specific vocabulary from non-discipline-specific vocabulary. These results inform an approach that could maintain the integrity of engineering exams and create more accessible assessment tools.

Keywords: Universal Instructional Design, Inclusivity, Assessment.

1. BACKGROUND

As the student population diversifies, instructors are seeing students face learning barriers due to inaccessible language regardless of existing English proficiency. There is evidence in the field of higher education claiming that increasing accessibility reduces learning barriers [1]. For example, the word “hood” might be inaccessible for a proficient English speaker accustomed to describing that automobile component as the “bonnet”. Since these words are not domain-specific to engineering, they are not taught in class [2]. This causes a particular problem for final exams because they are now testing the corpus of non-domain-specific vocabulary in addition to mastery of course material: and this can compromise the validity of the assessment.

The overall goal of this project is to identify inaccessible language, but the current area of research is to separate domain-specific and non-domain-specific words. This helps us focus in on the vocabulary that may be difficult for students and which is not explicitly taught in the discipline.

Inaccessible language is defined as vocabulary not explicitly taught, nor used in everyday speech but is assumed to be known by the student, including words like: succinct, sleuth, blob, etc. Domain-specific language is a corpus of vocabulary that contains technical jargon relevant to the area of study, and is therefore explicitly taught.

2. ONGOING WORK

To help us find characteristic words within groups of documents, an algorithm from the field of computer science called “Term Frequency Inverse-Document Frequency” (TF-IDF) is deployed. This algorithm is used to tag all of the words in a sample exam from Chemical Engineering.

The preliminary data shows that the TF-IDF algorithm is capable of accurately tagging characteristic words in the sample exam to help us distinguish domain-specific from non-domain-specific vocabulary. Specifically, across four test cases the best outcome was achieved when the input exam was compared against a dataset containing all exams from engineering in that particular year.

Ongoing work will include replicating the effectiveness of this method across different input exams to generate a corpus of domain-specific vocabulary. After removing domain specific terms from the search, the methodology will be used to investigate and characterize the other terms, such as inaccessible terminology, on engineering final exams.

References
