

# CIS5300 - Speech and Language Processing - Homework 1 Notes

Jonathon Delemos - Mark Yakstar

August 27, 2025

## 0.1 Regular Expressions (RE) Tools

The assignment highlights three core python regex functions:

- `re.search()` *Finds the first substring with the given pattern*
- `re.findall()` *Extracts all non-overlapping matches.*
- `re.sub()` *Takes one string and replaces with the desired one.*

Regular expressions are a powerful tool that should be used to transform text, locate patterns, extract structured information.

## 0.2 Data Structures

This assignment relied heavily on python dictionaries and lists to manipulate data at various levels.

### Dictionaries

- `word freqs =` - keys are immutable strings, values are integers
- loop through requires `items()`, *for word, count in word freq.items():*
- `return word: count for word, count in freq.items() if count >= threshold` List comprehensions combined with dictionaries were prevalent

## 0.3 Questions

It would be nice to speak with the professor. His lectures aren't posted anywhere..