

Friday Aug 30, 2024

In-class Handout

COSC 426A NLP

Prof. Forrest Davis

Name:

Discuss and complete the following questions with the person nearest you. You **may** be asked to share your thoughts with the class.

1. Bontoc is a Malayo-Polynesian language spoken by around 41,000 people in the Philippines. Isolate the morphemes in the following words.

fikas	'strong'	fumikas	'he is becoming strong'
kilad	'red'	kumilad	'he is becoming red'
bato	'stone'	bumiato	'he is becoming stone'
fusul	'enemy'	fumiusul	'he is becoming an enemy'

2. For each problem specification below, give the relevant NLP task type. What would your input be? What would your output look like? Are there special things to consider in preprocessing? What might your training data look like?
 - a. You've been hired by the Colgate Maroon News to do some archival work. They want to know how many unique entities of interest (e.g., buildings, people, classes) have been mentioned in their news over the years.
 - b. You've been hired by a government watchdog organization to help them understand AI policy. Many staff members lack the time and training to read complicated AI policy, but they want to understand what has been done by the government. You've been tasked with automatically creating briefs.
 - c. You've been hired by a local high school to help their overburdened and underpaid English teachers. They want to identify sentences in student writing that are ungrammatical or vague to help speedup their feedback to students.
 - d. Discuss something you might want to use NLP for.