**POS Tag in Software Documentation**

What are POS tags:

Noun, Verb, Adjective, etc

Word Embedding: Turn all existing words into vectors

Most\_Similar() most\_similar(flower) = rose, flowers, flowe

* Check if recommed words are similar in terms meaning and context
* Check if the word tag is correct

Import pandas as pd

C = 2

Analyzing human gene in disease A from conference A – author …..

* Analyze gene in dieseae B conference B – author ….

Software documentation: Recommend other blocks of verbs or nouns or phrases to current verbs ()

Proof read about ML, DL and build simple recommendation systems

NLP tools: NLTK, StanfordNLP (Java), Spacy, GoogleSyntax Net, …,

(Word2Vec, Fasttext) => Different Word Embedding ()

**Goal: (after applying word embedding tools) – mainly for normal English**

* Come up with a way to evaluate the word embeddings (check similar words, check correct tags)
* Come up with a better word embedding?
* Analyze current issues with word embedding for software documents
* Find approaches or best practices that can improve current word embeddings for software documentation!!!!!!

**A few approaches for Research:**

* It solves an existing problem (if current tools are not sufficient and you come up with new or improved version of that algorithm)
* Review: you apply all current tools in a field and do an Analysis about the tools performance
* Use current or new tools to tackle a real-world problem that hasn’t been tackled using those tools before

**Prerequisite**

* Basic understanding of NLP and common libraries
* Some knowledge of Machine Learning
* Git and Github for project management
* Either python or Java is fine.