

Machine Learning and Natural Language Processing Interview Questions

March 14, 2016

- What is POS (Part of Speech) tagging? What is the simplest approach to building a POS tagger that you can imagine?
- How would you train a model that identifies whether the word "Apple" in a sentence belongs to the fruit or the company?
- Design a recommendation system.
- What is latent semantic indexing and where can it be applied?
- How would you build a system that auto corrects text that has been generated by a speech recognition system?
- What is regular grammar? Does this differ in power to a regular expression and if so, in what way?
- How would you build a POS tagger from scratch given a corpus of annotated sentences? How would you deal with unknown words?
- How would you find all the occurrences of quoted text in a news article?
- How would you build a system to translate English text to Greek and vice-versa?
- How would you build a system that automatically groups news articles by subject?
- What are stop words? Describe an application in which stop words should be removed.
- How would you design a model to predict whether a movie review was positive or negative?
- What is entropy? How would you estimate the entropy of the English language?
- What is the TF-IDF score of a word and in what context is this useful?

- How does the PageRank algorithm work?
- What is dependency parsing?
- What are the difficulties in building and using an annotated corpus of text such as the Brown Corpus and what can be done to mitigate them?
- Are you familiar with WordNet or other related linguistic resources?
- Do you have any experience in building ontologies?
- Do you speak any foreign languages?
- What tools for training NLP models (nlTK, Apache OpenNLP, GATE, MALLET etc.) have you used?