

Hide menu

Lecture: Autocorrect and Minimum Edit Distance

Video: Intro to Course 2
1 min

Video: Week Introduction
55 sec

Video: Overview
1 min

Reading: Overview
3 min

Video: Autocorrect
2 min

Reading: Autocorrect
4 min

Video: Building the model
4 min

Reading: Building the model
3 min

Lab: Lecture notebook: Building the vocabulary
1h

Video: Building the model II
3 min

Reading: Building the model II
4 min

Lab: Lecture notebook: Candidates from edits
1h

Video: Minimum edit distance
3 min

Reading: Minimum edit distance
5 min

Video: Minimum edit distance algorithm
5 min

Reading: Minimum edit distance algorithm
3 min

Video: Minimum edit distance algorithm II
4 min

Reading: Minimum edit distance algorithm II
5 min

Video: Minimum edit distance algorithm III
3 min

Reading: Minimum edit distance III
4 min

Video: Week Conclusion
51 sec

Ungraded App Item: [IMPORTANT] Have questions, issues or ideas? Join our Community!
10 min

Reading: [IMPORTANT] Have questions, issues or ideas? Join our Forum!
2 min

> Week 1 > Building the model II

Building the model II

4.0 Calculating word probabilities

Calculate word probabilities

Example: “I am happy because I am learning”

$$P(w) = \frac{C(w)}{V}$$

$$P(\text{am}) = \frac{C(\text{am})}{V} = \frac{2}{7}$$

$P(w)$ Probability of a word

$C(w)$ Number of times the word appears

V Total size of the corpus

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total : 7

Note that you are storing the count of words and then you can use that to generate the probabilities. For this week, you will be counting the probabilities of words occurring. If you want to build a slightly more sophisticated auto-correct you can keep track of two words occurring next to each other instead. You can then use the previous word to decide. For example which combo is more likely, *there friend* or *their friend*? For this week however you will be implementing the probabilities by just using the word frequencies. Here is a summary of everything you have seen before in the previous two videos.

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear

yeah

dear

dean

... etc

Mark as completed



Like



Dislike



Report an issue

