

## Econ vocab terms with multiple words (words with spaces OK in vocab\_list but not in documents?)

So looking through the given stuff for a3, I noticed that many of the vocab terms for a3.track\_topic have multiple words within them. Through my current understanding of the code we are expected to produce (and know how to produce), it would be a lot more work if we had to account for multi word groups in docstrings within the vocab lists, and not let it detract from smaller subsets of that vocab. Like how "euro" and "euro zone" are both vocab words, we wouldn't want to count both maybe. Honestly, I don't have a full grasp on all the magic stuff happening that is provided, so I might just be wrong, and our code is producing the expected result, but I feel like this might throw the code off.

#pin

a3

~ An instructor (Kevin Cook) thinks this is a good question ~

Updated 15 minutes ago by Prof. Lee and [REDACTED]

### the instructors' answer, *where instructors collectively construct a single answer*

You're asking a lot of questions about the "big picture", which I'm very inspired to see!

In setting up A3, we made a couple of compromises between "being real" and "being easier for students in CS1110", and the situation with track\_topics() is a case in point.

It is true that the vocab\_list can contain items like "small business", even though "small business" cannot be considered a single word in docs\_list according to the precondition. It is reasonable to have a test case for this situation, and for a document "we own a small business", according to the specification, there will be no "hits" in that document for "small business".

Here's the reason for the set-up:

1. There are economic terms in The Economist's list that have spaces --- my current favorite is "animal spirits"--- where the individual words "animal" and "spirits" don't make sense as economic terms. That's why terms with spaces are allowed as items in vocab\_list. So, we're trying to be realistic on the vocab\_list front.
2. At the same time, on the doc\_lists front: students may have found it challenging to deal with whitespace issues in the convert functions. So, I decided to make processing documents in track\_topics easier, allowing students to treat all whitespace as word separators, recognizing that this sacrifices some of the correctness of the topic-tracking analysis.

Updated 1 hour ago by Prof. Lee and Kevin Cook

### followup discussions *for lingering questions and comments*