Started on	Thursday, 17 October 2024, 2:26 PM
State	Finished
Completed on	Thursday, 17 October 2024, 2:33 PM
Time taken	6 mins 55 secs
Grade	**20.00** out of 20.00 (**100**%)

Question 1

Correct

Mark 2.00 out of 2.00

Stemming...

a. Provides a lower number of tokens per concept, but always more than one

b. Maintains the original text, keeping it readable

c. Allows to reduce words to their root form ✓

d. Consists of representing text as a vector using the Vector-Space Model

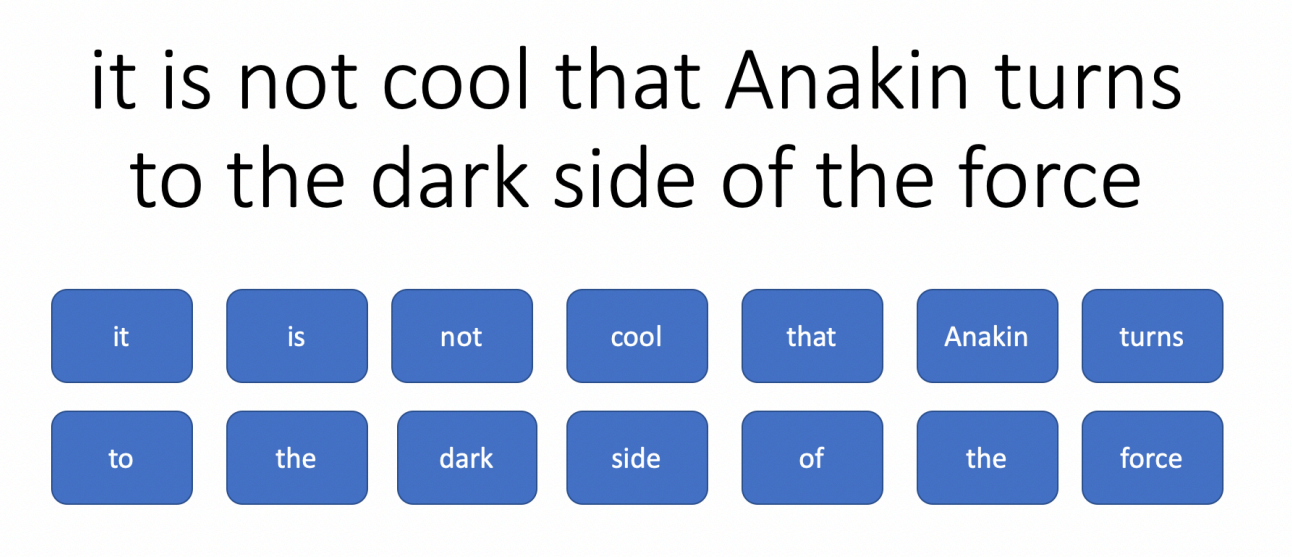
Your answer is correct.

Question 2

Correct

Mark 1.00 out of 1.00

What is this process called?



a. TF-IDF

b. Porter stemming

c. Tokenization ✓

d. Stopword removal

Your answer is correct.

Question 3

Correct

Mark 2.00 out of 2.00

Which is NOT a Readability Measure?

a. SMOG Index

b. Zipf's Index ✓

c. Gunning-Fog Score

d. Flesch-Kindcaid Reading Ease

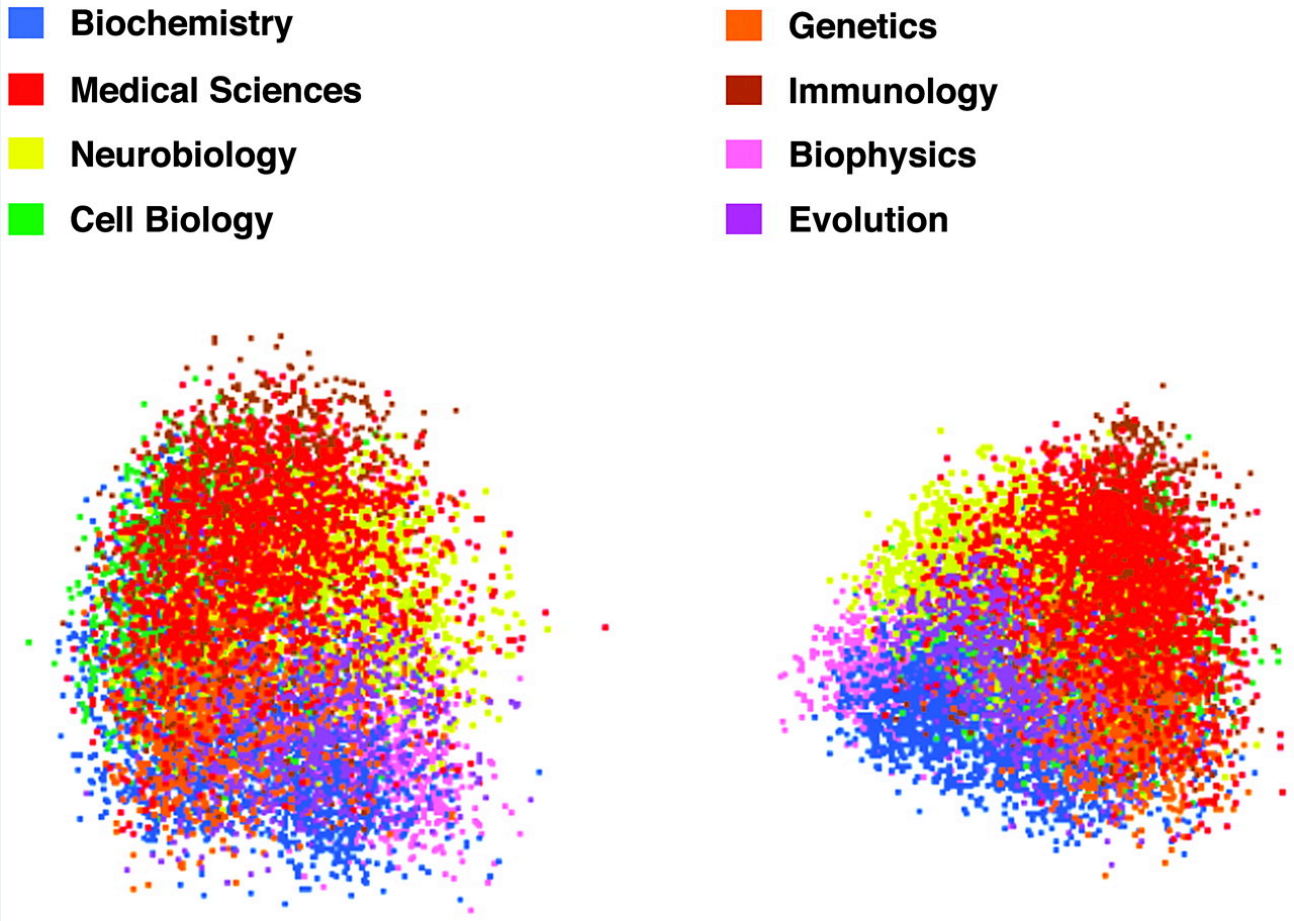
Your answer is correct.

Question 4

Correct

Mark 2.00 out of 2.00

The image below could be the result of...



a. Nothing, really. It's just a bad Vis.

b. Latent Semantic Analysis ✓

c. TF-IDF

d. Porter Stemming

Your answer is correct.

Question 5

Correct

Mark 1.00 out of 1.00

Why is it difficult to create effective text visualizations?

- a. Because text only consists of words themselves and not possible relationships among them
- b. Because text is pre-attentive
- c. Because text consists of abstract concepts ✓
- d. Because text is usually of low dimensionality

Your answer is correct.

Question 6

Correct

Mark 1.00 out of 1.00

What doesn't tokenization require?

- a. A good list of heuristics
- b. A good stemming framework ✓
- c. An appropriate language model
- d. A good library

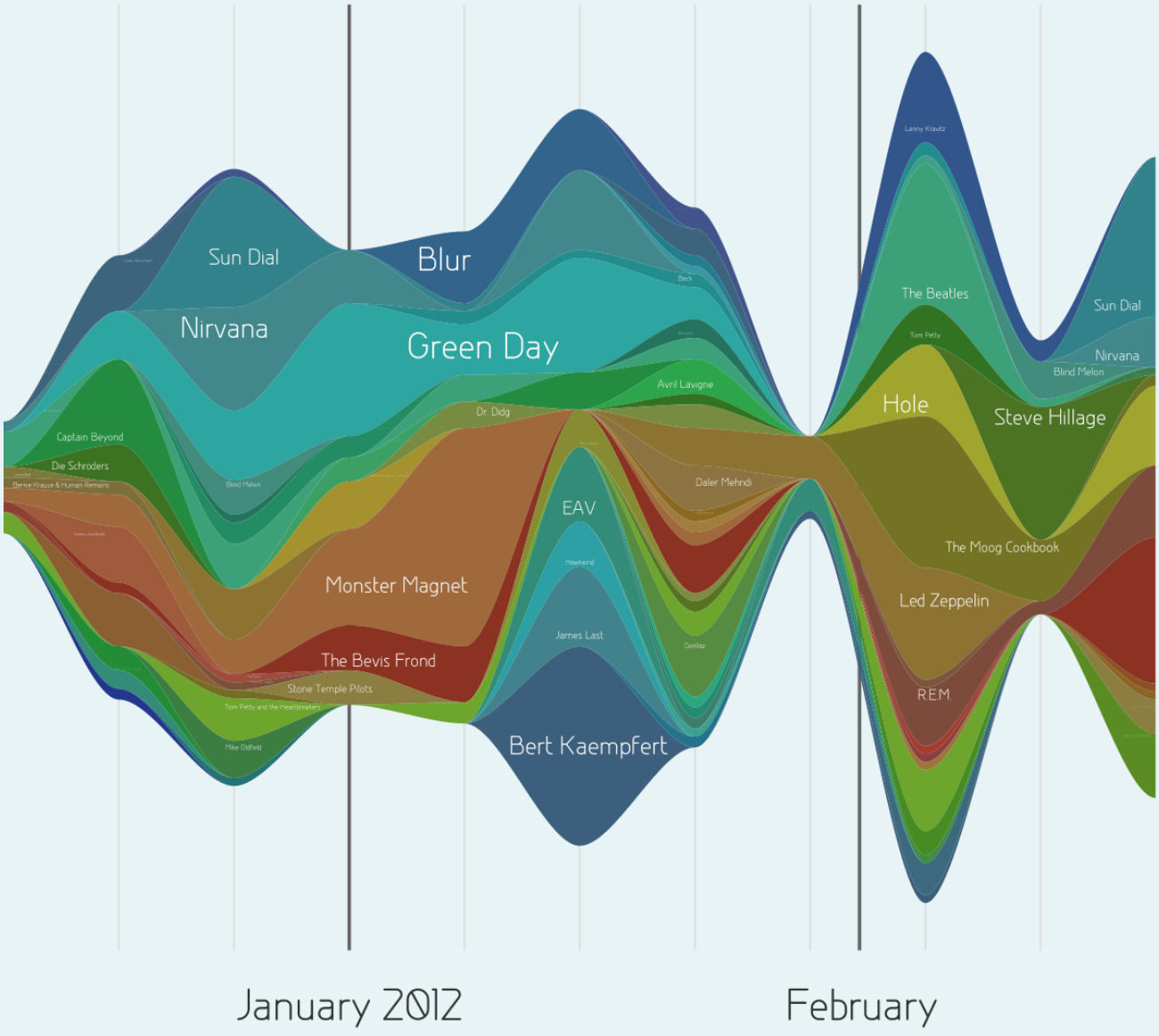
Your answer is correct.

Question 7

Correct

Mark 2.00 out of 2.00

Steamgraphs for visualizing text...



a. May be useful to show overall trends ✓

b. Are as good as Tag Clouds for representing general trends over time

c. Almost always violate the expressiveness principle because of the distortion they cause

d. Are excellent for representing hierarchical relationships and textual patterns

Your answer is correct.

Question 8

Correct

Mark 2.00 out of 2.00

Latent Semantic Analysis

a. Allows us to understand the relative importance of a term in a collection of documents

b. Aims at removing unnecessary words (stop words) from the original text

c. Does not consider the relationships among concepts

d. Extracts and represents the contextual-usage meaning of words by statistical computations applied to a large corpus of text ✓

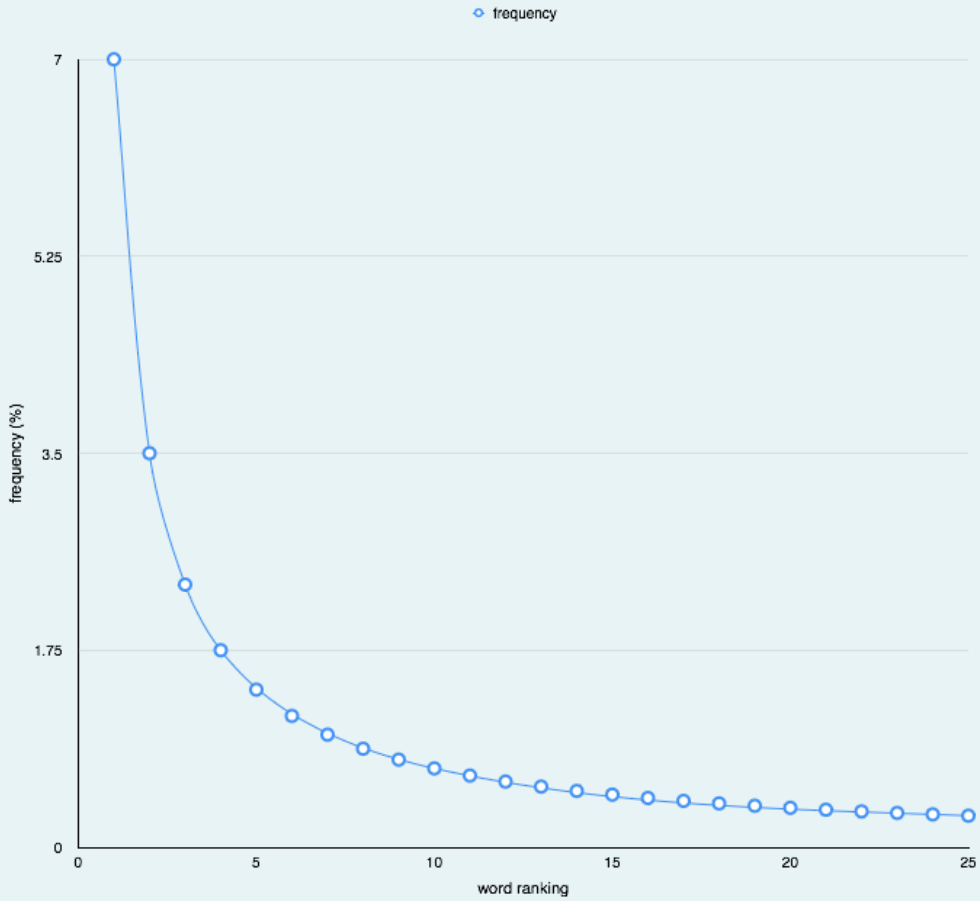
Your answer is correct.

Question 9

Correct

Mark 2.00 out of 2.00

"20% of words appear 80% of the time". This is...



a. A pattern that has been established for the English language and doesn't apply to any more

b. Zip's Law ✓

c. A lie

d. True for just a couple of examples in the literature, especially older books.

Your answer is correct.

Question 10

Correct

Mark 2.00 out of 2.00

Tag Clouds...

a. Provide an optimal visual encoding (size / position)

b. Make it very easy to find particular words

c. Keep text structure

d. Have an inacurate size encoding (long words are inherently larger, which may influence perception) ✓

Your answer is correct.

Question 11

Correct

Mark 1.00 out of 1.00

Which technique is more appropriate to show patterns in a story?

a. Word Trees ✓

b. Word clouds

c. Heat Maps

d. Sparkclouds

Your answer is correct.

Question 12

Correct

Mark 2.00 out of 2.00

This is an example of Tokenization

"The horror movie appealed to John" >> "Horror movie appealed John"

Select one:

True

False ✓