> term <- terms(lda, 7)

> (term <- apply(term, MARGIN = 2, paste, collapse = ", "))

Topic 1

"brexit, uk, the, n, britain, bn, eu"

Topic 2

"rt, brexit, n, eu, bn, bill, will"

Topic 3

"rt, brexit, the, eu, britain, i, vote"

Topic 4

"rt, eu, britain, uk, the, brexit, n"

Topic 5

"rt, the, uk, eu, n, bill, bn"

Topic 6

"brexit, rt, uk, britain, the, eu, i"

Topic 7

"brexit, n, divorc, will, i, rt, uk"

Topic 8

"rt, brexit, will, divorc, bn, bill, i"

Have list of queries

Every 24 hours get new queets related to queries

Put in database

Calculate sentiment and put in database: tweet, date, sentiment calculated

Export to CSV

Have website that let you select window of time

Help people make recommendations

Done:

Obtain text sources

-twitter

-using twitter api

Extract documents and move to corpus

Transformation

-standard text processing: convert to lower case, remove punctutation, remove numbers, remove stopwords, stemming

Extract Features

* Convert the text string into some sort of quantifiable measures
* Bag-of-words model
  + Each document is represented as a vector which counts the frequency of each term’s appearance in the document
  + Combine all the vectors for each document together and you create a *term-document matrix*
    - Each row is a document
    - Each column is a term

## Each cell represents the frequency of the term appearing in the Perform analysis

* Lots of approaches to take
* Basic
  + Word frequency
  + Collocation - words commonly appearing near each other
  + Dictionary tagging - locating a specific set of words in the texts
* Advanced
  + Document classification
    - Supervised
    - Unsupervised
  + Corpora comparison - comparing the content of different groups of text
  + Detecting clusters of document features - *topic modeling*
    - document

Collected sentences and put into vector

Use get\_sentiment function to assess sentiment of each word

Questions going forward:

How to utilize sentiment data