

## 1. Consider these documents:

|      |   |
|------|---|
| Doc1 | breakthrough drug for schizophrenia         |
| Doc2 | new schizophrenia drug                      |
| Doc3 | new approach for treatment of schizophrenia |
| Doc4 | new hopes for schizophrenia patients        |

- a. Draw a term-document count matrix for the document collection
- b. Calculate the inverse document frequencies of the terms in the document collection
- c. Draw a term-document TF-IDF weighted matrix for the document collection
- d. Find the document in the collection that is the closest to the new document "schizophrenia drug" by using cosine similarity between the TF-IDF vectors of the documents

## 2. Read, understand and re-run the example implementation of sentiment analysis of movie reviews by using binary or TF-IDF vectors at

<https://drive.google.com/file/d/1DyDKHSg6yTpDNQgv5bMfZKi-VDvCHVkl/view?usp=sharing>