

# Lab6: Text Categorization & IR

---

- Use the data from Lab 5 for the following two problems
- TC: Design a classifier for each category
  - Test the unknown documents to determine their topics, a key containing the category information of all test documents is also provided
  - Define an evaluation scheme and report your classification results
  - Choose two different classifiers and compare their performance
- IR: Design a “search engine” that will take a query sentence, convert it into a query vector, compare it with each document vector in the repository with some distance
  - Test your “search engine” with a few “designed” queries that can be relevant or non-relevant to the documents in the repository
  - Rank the top five documents that are “close” to each of the above queries, and list their corresponding distances
  - Propose an evaluation scheme, and report you system performance