# Problem 15. Twitter Analysis

|  |  |
| --- | --- |
| Program: | analysis.py |

The file Health-News-Tweets.zip is d dataset that contains health news from more than 15 major health news agencies such as BBC, CNN, and NYT. Each file is related to one Twitter account of a news agency. For example, bbchealth.txt is related to BBC health news. Each line contains tweet formatted as id|date and time|tweet. The separator is '|'.

Write a program named analysis.py that reads data from a file and extracts from the tweets ***twitter mentions***  and ***hashtags***.

# Problem 16. Course List

|  |  |
| --- | --- |
| Program: | courselist.py |

The file Final\_Draft\_Midterm\_Sem1\_19\_20.pdf is a PDF file of the college’s midterm examinations schedule which often posted on the website. You work for a data analytics company that that is building a database of courses in different colleges in Oman. Your boss asked you to create a database of IbriCT courses from the file. The database should consist of two fields: **course id** and **course name.** Write a program named courselist.py that extracts from the text of the given PDF the required data (i.e., course ID and course name).

**Hint** Extracting text from a PDF can be done outside Python. If you want to work with PDF files using Python check <https://realpython.com/pdf-python/> for examples of some libraries that do the task.