ITECH2302 Big Data Management Laboratory: Date-ordered Dictionary

Objectives:

- Sort Dictionary according to Key that contains a Date
- Jupyter notebook
- Try in pySpark

Activity 1

Sorting a Dictionary using Dates

1. Download the supplied datafile (*date_dict.csv*)
2. Load into pandas

These are the fields ('eid' is a unique index):

70-0.0 - 70-0.30	80-0.0 - 80-0.30	71-0.0 - 71-0.30	81-0.0 - 81-0.30
ICD10 codes	ICD10 dates	ICD9 codes	ICD9 dates

3. For each row create a date sorted dictionary
4. Create a dictionary that contains the results of frequencies of the ICD10 and ICD9 codes
5. Display this data as a Histogram
6. What is the most common ICD10 and ICD9 code?
7. How would you look for patterns in this data?

Activity 2

Try the same task using pySpark

Clues for date sorting dictionaries:-

From:
https://stackoverflow.com/questions/3977310/sorting-a-dictionary-with-date-keys-in-python

Method 1

```
def sortdict(d):
    for key in sorted(d): yield d[key]
for value in sortdict(mydict):
    # your code
    pass
```

Method 2

```
from collections import OrderedDict

ordered = OrderedDict(sorted(mydict.items(), key=lambda t: t[0]))
```