# CS2510: Data Structures and Algorithms – Assignment 3

Muhammad Yaseen Khan
DCS, MAJU

Fall – December 9, 2018

**Instructions.**

(i) Do not cheat. (ii) The assignment is consisting of four questions. (iii) You have to submit one `***.java` file that accommodate answers to all questions. (iv) Those questions where a descriptive answer is required, you can do in the same `***.java` file, as/in multi-line comments. (v) Read the first and third instructions again. (vi) This is an individual assignment. (vii) `***` in the instructions iii and iv means your student id. (viii) Failing to comply the instructions is same as failing to get assignment graded.

**Scenario.**

You are given three files, `rekhta-poets-1.tsv`, `rekhta-poets-2.tsv` and `rekhta-poets-3.tsv`. Each of these files contain data of more than 4,000 Urdu poets and their URLs (as scrapped from rekhta.org). Every line of these files contains the information separated by a tab space. The sample format of data is following:

| | |
|---|---|
| اے۔جی جوش | /poets/a-g-josh?lang=ur |
| عابدہ عروج | /poets/aabida-urooj?lang=ur |
| عادل رضا منصوری | /poets/aadil-raza-mansoori?lang=ur |
| آفاق صدیقی | /poets/aafaque-siddiqui?lang=ur |
| آفتاب الدولہ لکھنوی قلق | /poets/aaftaab-ud-daula-lakhnavi-qalaq?lang=ur |
| آفتاب رئیس پانی پتی | /poets/aaftab-rais-panipati?lang=ur |
| آغا نثار | /poets/aaga-nisaar?lang=ur |
| آغا اکبر آبادی | /poets/aagha-akbarabadi?lang=ur |

The specific point to note is the URLs are same for the poets in all files, however, the additional information exist in query string, such that the URLs in file which has information (name of poets) in Urdu language has a concatenated constant-string `?lang=ur`, similarly, the URLs in the file of Hindi data has concatenated constant-string `?lang=hi`, and for the file carrying the names in English there is no such query string. An example is given below:

| | |
|---|---|
| قمر جلالوی | /poets/qamar-jalalvi?lang=ur |
| क़मर जलालवी | /poets/qamar-jalalvi?lang=hi |
| Qamar Jalalvi | /poets/qamar-jalalvi |

**Objective.**

Make a (one) consolidated dictionary for the poets data, such as the processed URLs (the hints for processing is defined in following) will be set as key, while the ArrayList having the poet names (Urdu, Hindi, and English)

will be set to the its respective URL. For doing so, you need to declare a hash map such as `HashMap<String, ArrayList<String>> poetDic = new...`. The structure of the final entries in the dictionary is as following:

```
{
    "/poets/a-g-josh": ["  اے۔جی جوش ", " ए जी जोश , "A G Josh"],
    "/poets/allama-iqbal": [" علامہ اقبال ", " अल्लामा इक़बाल , "Allama Iqbal"],
    "/poets/mirza-ghalib": [" مرزا غالب ", " मिर्ज़ा ग़ालिब , "Mirza Ghalib"],
}
```

**Processing Hints.**

- Declare a single dictionary.

- Strip off query string in URL i.e. `?lang=ur` and `?lang=hi` from the line while reading, this will give you the key.

- Process a single file at a time. This will populate all of the keys in the dictionary with one name.

- Then proceed to other files, for which you just have to process your key. Access the value through the key and then append the new name in it. Update the old value against the same key with the new list.

**Submit timely on your Google Classroom**
**The due date is December 12, 2018**