Big data: architectures and data analytics

MapReduce - Exercises
Exercise #22

- Friends of a specific user
  - Input:
    - A textual file containing pairs of users (one pair per line)
    - Each line has the format Username1,Username2
    - Each pair represents the fact that Username1 is friend of Username2 (and vice versa)
    - One username specified as parameter by means of the command line
  - Output:
    - The friends of the specified username stored in a textual file
    - One single line with the list of friends

Exercise #22 - Example

- Input file
  - User1,User2
  - User1,User3
  - User1,User4
  - User2,User5

- Username parameter: User2
- Output file
  - User1 User5
Exercise #23

- Potential friends of a specific user
  - Input:
    - A textual file containing pairs of users (one pair per line)
    - Each line has the format Username1,Username2
      - Each pair represents the fact that Username1 is friend of Username2 (and vice versa)
    - One username specified as parameter by means of the command line
  - Output:
    - The potential friends of the specified username stored in a textual file
    - One single line with the list of potential friends
    - User2 is a potential friend of User2 if they have at least one friend in common

Exercise #23 - Example

- Input file
  - User1,User2
  - User1,User3
  - User1,User4
  - User2,User3
  - User2,User4
  - User2,User5
  - User5,User6

- Username parameter: User2
- Output file
  - User1 User3 User4 User6
Exercise #23 Bis

- Potential friends of a specific user
  - Solve problem #23 by removing the friends of the specified user from the list of its potential friends

Exercise #23 Bis - Example

- Input file
  - User1,User2
  - User1,User3
  - User1,User4
  - User2,User3
  - User2,User4
  - User2,User5
  - User5,User6

- Username parameter: User2
- Output file
  - User6
Exercise #24

- Compute the list of friends for each user
  - Input:
    - A textual file containing pairs of users (one pair per line)
      - Each line has the format
        - Username1,Username2
      - Each pair represents the fact that Username1 is friend of Username2 (and vice versa)
  - Output:
    - A textual file containing one line for each user. Each line contains a user and the list of its friends

Exercise #24 - Example

- Input file
  - User1,User2
  - User1,User3
  - User1,User4
  - User2,User5

- Output file
  - User1: User2 User 3 User 4
  - User2: User1 User5
  - User3: User1
  - User4: User1
  - User5: User2
Exercise #25

- Compute the list of potential friends for each user
  - Input:
    - A textual file containing pairs of users (one pair per line)
    - Each line has the format
      - Username1,Username2
    - Each pair represents the fact that Username1 is friend of Username2 (and vice versa)
  - Output:
    - A textual file containing one line for each user with at least one potential friend. Each line contains a user and the list of its potential friends
    - User1 is a potential friend of User2 if they have at least one friend in common

Exercise #25 - Example

- Input file
  - User1, User2
  - User1, User3
  - User1, User4
  - User2, User3
  - User2, User4
  - User2, User5
  - User5, User6

- Output file
  - User1: User2 User3 User4 User5
  - User2: User1 User3 User4 User6
  - User3: User1 User2 User4 User5
  - User4: User1 User2 User3 User5
  - User5: User1 User3 User4
  - User6: User2
Exercise #26

- Word (string) to integer conversion
  - Input:
    - A large textual file containing a list of words per line
    - The small file dictionary.txt containing the mapping of each possible word appearing in the first file with an integer. Each line contains the mapping of a word with an integer and it has the following format
      - Word\tInteger\n  - Output:
    - A textual file containing the content of the large file where the appearing words are substituted by the corresponding integers

Exercise #26 - Example

- Input files
  - Large textual file

```
TEST CONVERSION WORD TO INTEGER
SECOND LINE TEST WORD TO INTEGER
```

- Small dictionary file

```
1  CONVERSION
2  INTEGER
3  LINE
4  SECOND
5  TEST
6  TO
7  WORD
```
## Exercise #26 - Example

- **Output file**

<table>
<thead>
<tr>
<th>51762</th>
</tr>
</thead>
<tbody>
<tr>
<td>435762</td>
</tr>
</tbody>
</table>