자료구조 프로젝트

Jong-Kyou Kim, PhD

2016-06-03

제 1 절 개요 Overview

- 한 번에 한 단어를 tweet 하는 시스템이 있다고 가정하자. Suppose that there is a system by which users can tweet a word at a time.

- 이 시스템에는 다음과 같은 정보가 있다. This system has the following information

    - 사용자 정보 user profile
    - 사용자간 친구관계 friendship relation
    - 사용자별 tweet Tweets by each user

- Due date: 2016-06-24 23:59:59 (Fri)

제 2 절 목표 Objective

- Advanced search

    - Which word is tweeted the most by all users
    - Which user tweeted the most words
    - Find users who tweeted a word (e.g., '연세대')
    - Find all people who are friends of at least one of the above found users

- Advanced delete

    - Delete all mentions of a word
    - Delete all users who mentioned a word
        * This requires updating friendship database

- Graph algorithms

    - Top 5 Strongly connected components (most number of users)
    - Top 5 Shortest path (weight = number of friends)

제 3 절 Data format

3.1 User profile (user.txt)

- User profile is composed of three parts. Identification number (string), sign-up date (string), and screen name (string).

```
433072426
Sat Dec 10 03:28:31 +0000 2011
qhals0086

174670917
Wed Aug 04 14:30:51 +0000 2010
soloist_shin

...
```

3.2 Friendship (friend.txt)

- Friendship information is stored as a tuple. User id (string) and his/her friend id (string).

```
313426093
323060839

313426093
107145933

...
```

3.3 Word tweet (word.txt)

- Word tweet is composed of three parts: User id (string), tweet date (string), and a word (string)

```
433072426
Sat Mar 17 14:31:34 +0000 2012
그건

433072426
Sat Mar 17 14:31:34 +0000 2012
혼자만의
```

```
433072426
Sat Mar 17 14:31:34 +0000 2012
착각~저게


...
```

제 4 절   Interface

```
0. Read data files
1. display statistics
2. Top 5 most tweeted words
3. Top 5 most tweeted users
4. Find users who tweeted a word (e.g., '연세대')
5. Find all people who are friends of the above users
6. Delete all mentions of a word
7. Delete all users who mentioned a word
8. Find strongly connected components
9. Find shortest path from a given user
99. Quit
Select Menu:
```

4.1   Read data files

```
Total users: xxx
Total friendship records: xxx
Total tweets: xxx
```

4.2   Statistics

```
  Average number of friends: xxx
  Minimum friends: xxx
  Maximum number of friends: xxx


  Average tweets per user: xxx
  Minium tweets per user: xxx
  Maximu tweets per user: xxx
```

제 5 절　Submit items

- Submit Github URL in which the following information is stored

  - Source code

  - Manual

  - Report

    * What data structure you chose and why

    * What is your expected performance

    * How would you improve the system in the future

  - Self evaluation form

제 6 절　Self evaluation

- Submit a github account (10)

- Commit source code displaying menu (10)

- Commit the first draft of manual (10)

- Read data files (20)

- Statistics (20)

- Top 5 most tweeted words (10)

- Top 5 most tweeted users (5)

- Find all users who mentioned a word (10)

- Find all users who are friend of the above user (5)

- Top 5 strongly connected components (10)

- Find shortest path from a user (id) (10)

제 7 절　Evaluation criteria

- Documentation (30)

- Bug fixing efforts (20)

- Code quality (20)

- Large size data test $\rightarrow O(n \log n)$ (30)