Programming Assignment #5

Citation Network Analysis

Scientific papers cite other relevant articles. Junior scientists usually struggle to find the most relevant papers. In this assignment, you will help (for now, only the computer scientists).

Your program takes a (large) JSON file that contains the information about the papers, a *keyword* and an integer *n*. It first searches the articles with titles containing the *keyword*, then it finds the papers that cited the papers in this step. We call them tier-1 papers. For tire-2 papers, we find the papers that were cited by tire-1 papers. More generally, tier-K papers are the papers that are cited in tier-(K-1) papers.

The output should be information of the papers *all* tiers up to level *n* ranked in order of importance. You decide how to define importance.

Instructions:

- You can use only Java Stream
- Use Maven to automate the build.

Submission Instructions:

- Create a project named “dblp-explorer” on your GitHub profile.
- Tag your submission version with “v.1”.
- Use GitHub markdown to create “design.md” file about how your design solves the problem efficiently.
- Use GitHub markdown to create “benchmark.md” and show the time and memory consumption of your program.

Example Files:

Citation-network V1: 629,814 papers and >632,752 citation relationships (2010-05-15).

Citation-network V2: 1,397,240 papers and >3,021,489 citation relationships (2010-09-13).

DBLP-Citation-network V3: 1,632,442 papers and >2,327,450 citation relationships (2010-10-22).

DBLP-Citation-network V4: 1,511,035 papers and 2,084,019 citation relationships (2011-01-08).

DBLP-Citation-network V5: [download from mirror site] 1,572,277 papers and 2,084,019 citation relationships (2011-01-08).

DBLP-Citation-network V6: 2,084,055 papers and 2,244,018 citation relationships (2013-09-29).

DBLP-Citation-network V7: [download from mirror site] 2,244,021 papers and 4,354,534 citation relationships (2014-05-25).

DBLP-Citation-network V8: 3,272,991 papers and 8,466,859 citation relationships (2016-07-14)

ACM-Citation-network V8: 2,381,688 papers and 10,476,564 citation relationships (2016-04-02)

ACM-Citation-network V9: 2,385,022 papers and 9,671,893 citation relationships (2017-01-20)

DBLP-Citation-network V9: 3,680,007 papers and 1,876,067 citation relationships (2017-07-03)

DBLP-Citation-network V10: 3,079,007 papers and 25,166,994 citation relationships (2017-10-27)

DBLP-Citation-network V11: 4,107,340 papers and 36,624,464 citation relationships (2019-05-05)