

Assignment 3. Big Data platforms and tools – 2: Graph Databases

In this assignment, you should provide experiments and analysis concerning graph databases

1. Principles of graph databases:
 - a. Analyze the most important features of graph DB
 - b. Provide a brief synopsis of this analysis as part of your report (up to 1 page)
 - c. Annotate a research paper “Use of Graph Database for the Integration of Heterogeneous Biological Data”; include it into your report
2. Neo4j
 - a. Install the software and familiarize yourself with Cypher and Browser
https://www.quackit.com/neo4j/tutorial/neo4j_installation.cfm,
<https://neo4j.com/docs/cypher-manual/current/introduction/>
 - b. Prepare a small illustration of running Neo4j, use a small graph of your choice
 - c. Prepare to discuss it in class, March 15
3. Movie dataset
 - a. Consider the TMDb dataset (from Assignment 2)
 - b. Formulate a multiple domain graph (of your choice of complexity)
 - c. Create a Neo4j model for this task
 - d. Expand (reasonably) the graph model in Neo4j
 - e. Run queries of your choice.
 - f. Present the results in the final report.

This is a group assignment.

We will discuss it in class on March 15 (n. 1, 2) and March 22 (after the exam) March 29.

Submit your full report to the Dropbox by March, 28, 11:59pm.