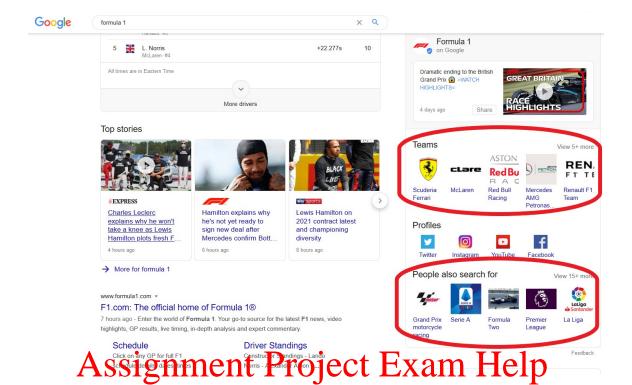
Reading List

- 1. RHEEM: Enabling Cross-Platform Data Processing. VLDB 2018.
- 2. LSH Ensemble: Internet-Scale Domain Search. VLDB 2016.
- 3. Table Union Search on Open Data. VLDB 2018.
- 4. HoloClean: Holistic Data Repairs with Probabilistic Inference. VLDB 2017.
- 5. Auto-Detect: Data-Driven Error Detection in Tables. SIGMOD 2018.
- 6. <u>Heterogeneous Recommendations: What You Might Like To Read After Watching Interstellar.</u> VLDB 2017.
- 7. Maverick: Discovering Exceptional Facts from Knowledge Graphs. SIGMOD 2018.
- 8. Knowledge Exploration Using Tables on the Web. VLDB 2016.

Questions

- 1. What are the main two categories of collaborative filtering approaches for recommendation systems?
- 2. What is the difference between a user-based and an item-based collaborative filtering? In which scenarios one would be preferred ever the other? Exam Help
- 3. Scalability is an issue in the face of heterogeneous recommendation because it is not feasible to calculate multi-hop similarities between items (e.g., the relationship between y1 and y3 in the following ratings: x1 likes y1, x2 likes y2, x3 likes y2, x3 likes y3). What are the possible techniques to be come this problem?
- 4. Explain how is mining for exceptional facts in knowledge graphs different than outlier detection? $Add \ We Chat \ powcoder$
- 5. For finding unionable attributes for a large corpus of relational tables, the state-of-the-art approach applies three statistical tests to determine if a pair of attributes is unionable. Those are set domains test, semantic domains test, and natural language domains test. Explain in which cases each test excels and fails.
- 6. Optimizers in databases rely on cardinality estimates of intermediate results to produce execution plans. RHEEM follows a similar approach by estimating the cardinality of the output of each execution operator in the plan. Sometimes these estimates are not accurate, which affects the execution time given the initial execution plan. What can be done to solve this problem?
- 7. One approach to detect errors in single columns in relational tables is called Potter's Wheel. It relies on finding common patterns from values in a given column and predict values that significantly differ from these patterns as errors. This approach however falls short in multiple cases. In which cases does Potter's Wheel fail? What may be a good strategy to overcome this problem?
- 8. For the given search result on Google for "Formula 1", the following panel is shown (attached). Give an overview of how the results in the red capsules are obtained.



https://powcoder.com

Add WeChat powcoder