

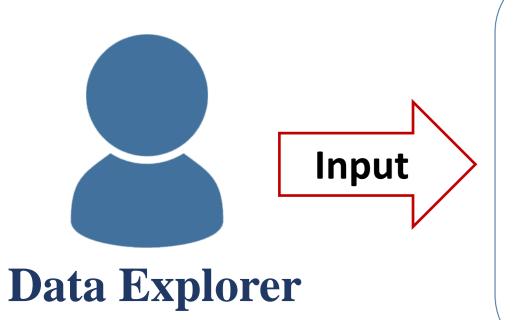


Continuous Adaptive Graph Modeling Using RecGraph and LSH-Forest for Data Exploration and Recommendation

Ibraheem Taha^{1,2,3}, Yannis Ioannidis^{1,2}, Torben Bach Pedersen³, Alkis Simitsis^{1,2}, Matteo Lissandrini³ ¹Athena Research Center, Greece; ²National and Kapodistrian University of Athens; ³Aalborg University, Denmark

PROBLEM STATEMENT

We aim to design a graph data modeling method to provide helpful, and diverse recommendations to data explorers.



Continuous and Adaptive Graph Modeling

- Continuous: online, fast, and interactive.
- Adaptive: dynamic, and flexible.

Using:

- RecGraph [1].
- LSH-Forest [2].

Recommendations: • Dataset. Output

• Action.

• Visualization.

Difficulties with Juneau [3]:

- Individually, constructing data profiles for objects, influencing processing efficiency and scalability.
- Using several threads, and DBMSs which overloads resources, and providing limited forms of results.

OUR SOLUTION

(1) Document of Strings

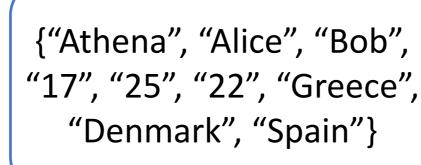
Treating the table as a document of strings.

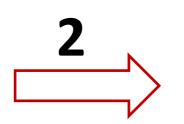
Name	Age	Country
Athena	17	Greece
Alice	25	Denmark
Bob	22	Spain



(2) LSH-Forest

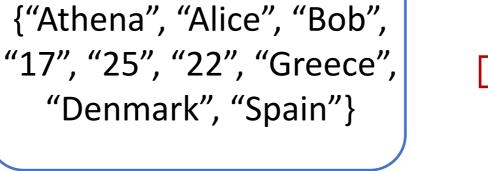
Employing LSH-Forest for adaptive k similar tables.

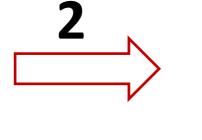




(3) Single DBMS

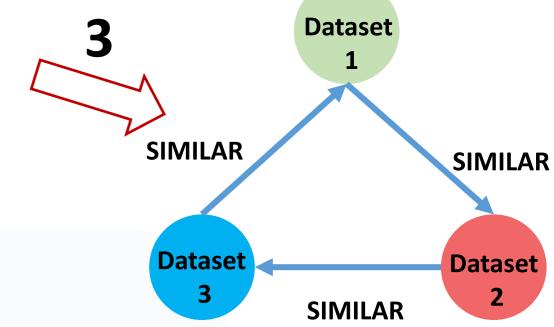
Modeling objects and their connection in a single DBMS.



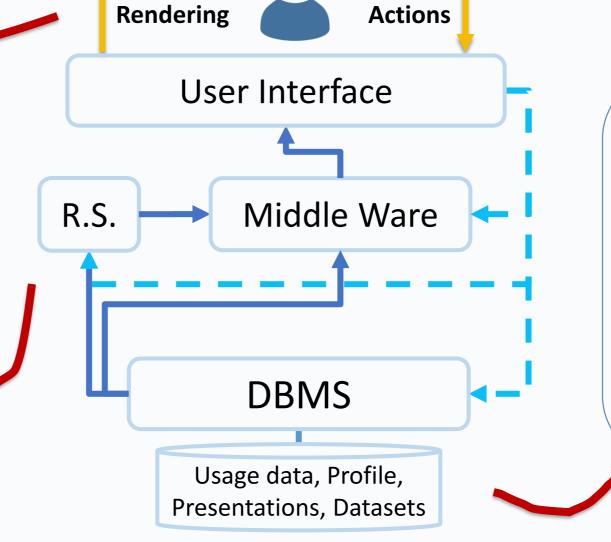




Bucket 1



- (5) Tackles the complexity and heterogeneity of data exploration environment.
- (1) Improves, efficiency, and scalability of data processing.
 - (2) Enhances the time into sub-linear; efficiency.



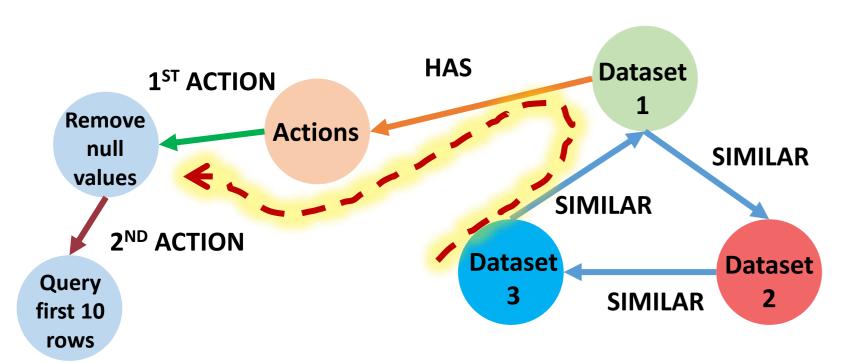
- (3) Addresses resources utilization.
- (4) Improves efficiency, scalability, and flexibility of data search, and retrieval.



Data Exploration Architecture

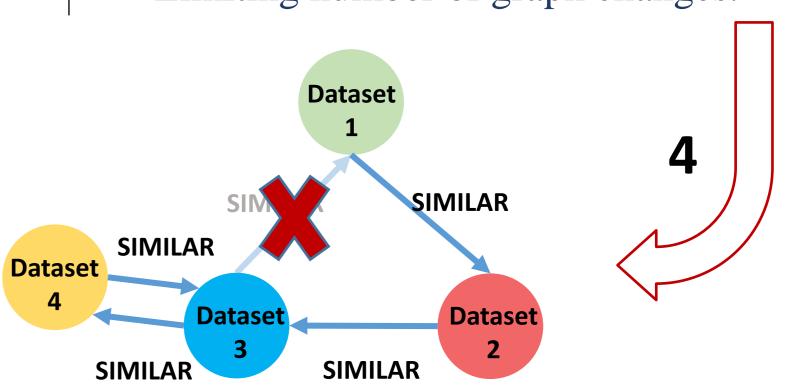
(5) Several Types of Recommendations

Providing several types of recommendations.



(4) Dynamic, and Partial Updates

Limiting number of graph changes.



[1] M. Kyriakidi, G. Koutrika, Y. Ioannidis. Recommendations as Graph Explorations. In RecSys, 2020. [2] Bawa, Condie, and Ganesan, "LSH Forest: Self-Tuning Indexes for Similarity Search," WWW '05, pp. 651-660, ACM. [3] Zhang and Ives, "Finding related tables in data lakes for interactive data science," SIGMOD '20, pp. 1951-1966, ACM.