Paper Title

Student Name
UM6P

Karima Echihabi
UM6P

ABSTRACT

300 word description of the project

PVLDB Reference Format:
Student Name and Karima Echihabi. Paper Title, 15(10): XXX-XXX, 2022.
doi:XX.XX/XXX.XX

1 INTRODUCTION

This is where you motivate the problem.

2 PRELIMINARIES

This is where you put definitions, including a formal problem definition.

3 RELATED WORK

This is where you survey and cite the related work. Add all referenced work in the paper to the ref.bib file (using meaningful names not the default ones) and to cite it as follows [1].

4 THE XYZ APPROACH

This is where you describe the approch that solves the problem in section 2.

5 EXPERIMENTAL EVALUATION

This is where you describe the experimental setup and discuss results. Save plots in the img directory and make sure they look nice(Figure 1)

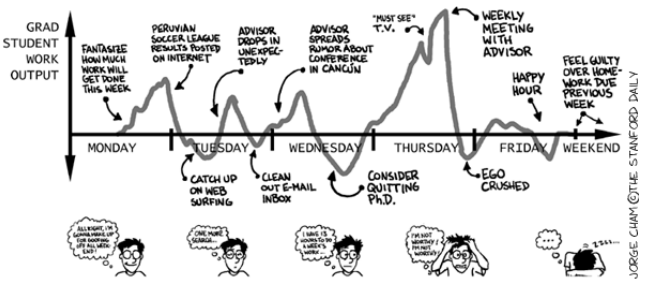

Figure 1: I promise to be productive today! [2]

Proceedings of the VLDB Endowment, Vol. 15, No. 10 ISSN 2150-8097.
doi:XX.XX/XXX.XX

6 CONCLUSIONS AND FUTURE WORK

One concluding paragraph summarizing the contributions and pinpointing any future work directions.

7 IDEAS AND QUESTIONS

7.1 Ideas

(1) The dataset discovery process is more likely to be iterative. At the beginning the user forms a general query for his information need. Retrieved datasets help the user better understand his need and hence better reformulate his query in the next iteration.
(2) Guarantee interactive-level response time.
(3) The relationship that we aim to capture between datasets oftentimes defines the structure of the query that search engine will accept.
(4) speed up the data science workflow.
(5) Similarity measures:
  (a) Text data: Hammin distance, Jaccard similarity, Jaccard containment ...
  (b) Numerical data: Correlation, cosine distance, euclidean distance ...
  (c) Binary data: ...
  (d) TF-IDF, Okapi BM25, LDA topic modeling
(6) Order based similarity measures (image retrieval)?

7.2 Questions

(1) If the query time is too small how can the user observe changes in query results as they arrive incrementally?
  + The user can observe changes in query results because when querying over Terabytes (or even Petabytes of data) the query time will increase significantly.

ACKNOWLEDGMENTS

We sincerely thank X, Y and Z.

REFERENCES

[1] K. Echihabi, K. Zoumpatianos, T. Palpanas, and H. Benbrahim. The Lernaean Hydra of Data Series Similarity Search: An Experimental Evaluation of the State of the Art. *PVLDB*, 12(2), 2018.
[2] PhDComics. Graduate Student Work Output. https://phdcomics.com/comics/archive.php?comicid=124, 2022.