CS387: Project Proposal

Team

Name	Roll Number
Adarsh Kumar	19D180003
Adithya Bhaskar	190050005
Devansh Jain	190100044
Harshit Varma	190100055

Type of project

Information system / DB application
We aim to create a web-based research analytics system for CS, similar but not limited to
CSRankings. We plan to test our system using data from the DBLP dataset (author and
publication data) and the Semantic Scholar dataset (citations data).
Please see the description and references sections for more details.

Description / Proposed Website Interface

(Note: The below list is currently not exhaustive)

- `(/institutes)`
  Institute rankings similar to [1], [2].
  Institute names will be clickable and will redirect to `/institutes/<institute_id>`
  - User should be able to filter by (brackets contain the input type)
    - Institute name (text input)
    - Venues of publications (selection)
    - Country (selection)
    - Time period (range)
  - User should be able to sort by predefined metrics (after applying the above filters)
    - Number of faculty
    - Average citations

- `(/institutes/<institute_id>)`
  Display author rankings for the given institute similar to `/authors`, along with some aggregated metrics for the institute

- `(/authors)`
  Display a table with author rankings similar to [3].
  Author names will be clickable and will redirect to `/authors/<author_id>`
  - User should be able to filter by (brackets contain the input type)
    - Author name (text input)
    - Venues of publications (selection)
    - Country (selection)
    - Time period (range)
  - User should be able to sort by predefined metrics (after applying the above filters)
    - Number of publications
    - Number of citations
    - h-index
    - i10-index

- `(/authors/<author_id>)`
  Show the metrics for the author, along with the publications.
  Visualize the variation of these metrics wrt time using charts.
  Show overall research interests as a pie chart.
  Show co-authorship graph for the author, like [4, 7]. (nodes are authors, weighted undirected edge if they've collaborated, with weight = no. of co-authored articles)
  Filters remain the same as those for `/authors`

- `(/articles/<article_id>)`
  Contains details like venue of publication, citations, etc. about the article.
  Also show a citation graph (nodes are articles, directed edge (v1, v2) if v1 cites v2) for the article, like [4, 6].
  An additional search interface will be provided where the user can enter the article's DOI and search for the article directly.

Relevant concepts

- Theory
  - Feature specification
  - ER Design
  - Normalization
  - Conversion to schema
- Implementation
  - Relational DBs
  - Graph DBs
  - Front-end, back-end designs and queries
  - Data visualization
  - Raw data processing

Plan / Timeline

We plan to follow the same timeline as specified in the Moodle document.

Project Deliverable	Approx deadline (to be finalized on Moodle)
Team formation, first sketch of the project requirements defining scope.	Mar 7th
Complete requirements & Analysis: Detailed enumerated english requirements, Complete ER diagram, forms, use cases in detail.	Mar 20th
Design Doc: Normalized schema, Indexes, DDL, Data load scripts, SQL for transactions, Detailed forms, (optionally) Controller logic	Apr 1st
Detailed Test Plan including load testing	Apr 1st
Final report (everything so far + test results + load testing results + conclusions)	Apr 20th
Demo and viva	TBA

References

- [1] https://csrankings.org/
- [2] https://airankings.org/
- [3] https://airankings.professor-x.de/
- [4] https://timwoelfle.github.io/Local-Citation-Network/
- [5] https://scholar.google.com/
- [6] https://www.connectedpapers.com/
- [7] https://app.vosviewer.com/docs/examples/