CS387 Project: Requirements and Analysis

Team

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

The Problem

We aim to create a web-based research analytics system for CS, similar but not limited to CSRankings. The interface provides a metrics-based ranking of top computer science institutions worldwide. This information is useful for students and professionals who are interested in associating themselves with academic institutions in the field of computer science. The interface is also useful for referring to articles and relevant conferences using citation graphs and co-authorship graphs. This information is best represented in the form of a Graph DB.

Entities/Nodes

- Institute
    - Attributes
        - id
        - name
        - country
        - metrics
            - total number of members (people associated with the institute)
            - total number of publications
            - total number of citations
            - publications per member (#publications/#members)
            - citations per member (#citations/#members)
    - id, tuple(name, country) are unique (i.e., we allow institutes to be present in multiple countries and treat them separately. For example: Google India and Google Zurich)
- Author
    - Attributes
        - id
        - name
        - country
        - metrics
            - number of publications
            - number of citations
            - h-index (maximum value of h such that the given author/journal has published at least h papers that have each been cited at least h times)
    - id is unique
- Article
    - Attributes
        - id
        - title
        - year
        - metrics
            - number of citations
    - id is unique
- Venue
    - Attributes
        - id
        - name
        - acronym
        - type (conference/journal/workshop)
        - metrics
            - average number of citations

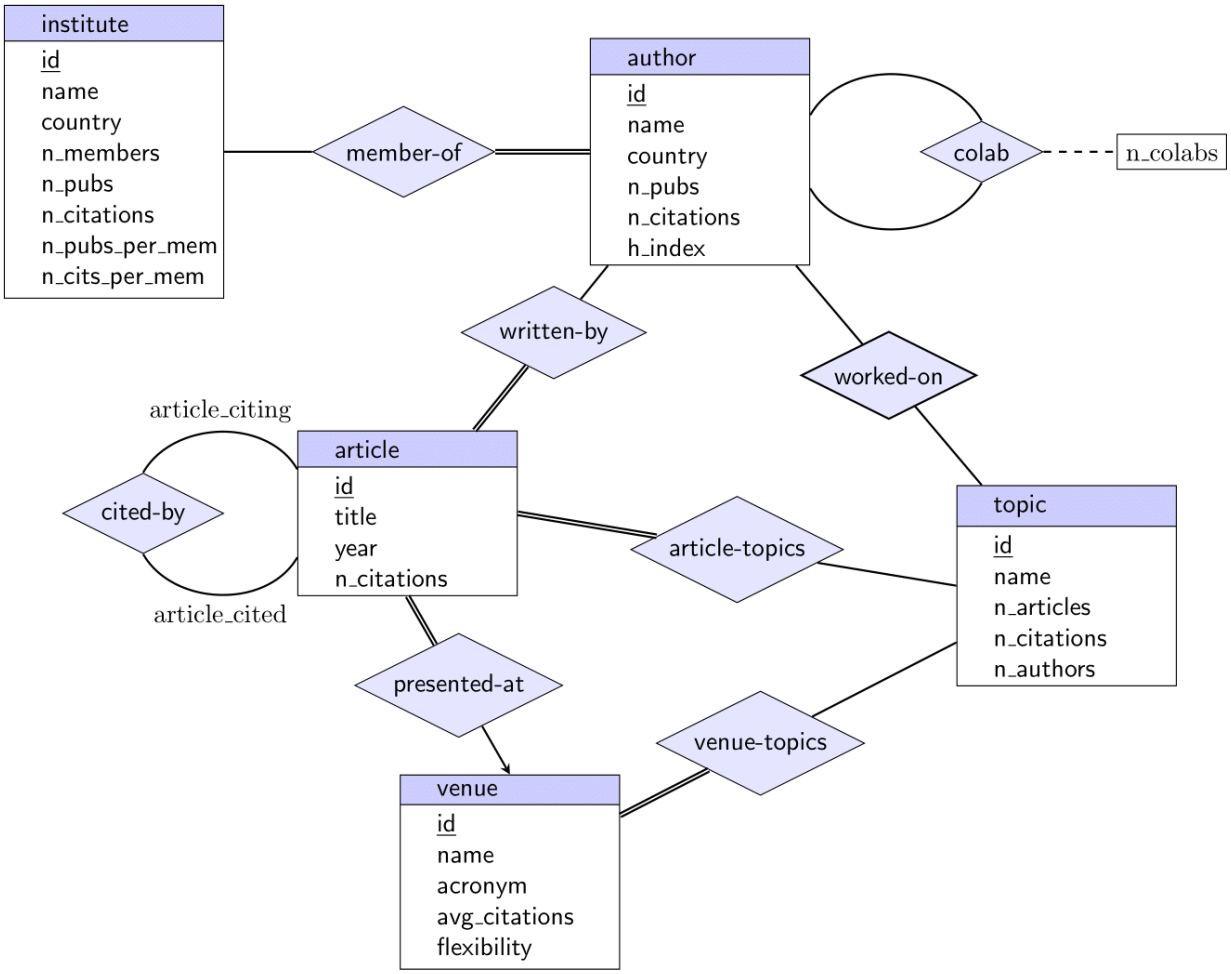
                - flexibility (#articles published in the venue that has at least 1 topic not present in the venues list of topics / #articles published in the venue)
        - id, name both are unique
- Topic
    - Attributes
        - id
        - name
        - metrics
            - total number of articles having this topic
            - total number of citations using the above articles
            - total number of authors having this topic in their top-5 topics
    - id, name both are unique

Metrics are to be computed using Spark and stored in the DB, instead of being computed each time the entity is fetched.
If data is updated, or the user defines custom parameters for the metrics (eg. time period), then these metrics will be recomputed.

Relations/Edges

- Author – Institute
    - An institute can have >= 0 authors
    - An author can have >= 1 institutes (“Independent” will be a default institute)
- Author – Article
    - An article can have >= 1 authors
    - An author can have >= 0 articles
- Author – Topic
    - An author can have >= 0 topics (determined by the top k topics the author has published articles in)
    - A topic can have >= 0 authors
- Article – Venue
    - An article can have = 1 venue
    - A venue can have >= 0 articles
- Article – Topic
    - An article can have >= 1 topics (“Miscellaneous” won’t be allowed)
    - A topic can have >= 0 articles
- Venue – Topic
    - A venue can have >= 1 topics (this is predefined/fixed)
    - A topic can have >= 0 venues
- Article – Article
    - This defines the citation graph
    - An article can be cited by or cite >= 0 other articles
- Author – Author
    - This defines the co-authorship/collaboration graph
    - This relation/edge will have an attribute which contains the number of times this pair of authors have co-authored an article
    - An author can collaborate with >=0 other authors

Conceptual ER Diagram



User Interface

- `(/home)`
  Contains tabs for `/institutes, /authors, /venues, /topics`

- `(/institutes)`
  Shows a table with columns corresponding to the institute entity's attributes
  Institute names will be clickable and will redirect to `/institutes/<institute_id>`
  Rows will be sorted according to a default metric (citations per member)
  User will be able to:
  - sort by any of the defined metrics (dropdown)
  - filter by the country (dropdown)
  - filter by topics (dropdown): only articles having these topics will be used for re-computing the metrics
  - search by name (text input)
  - specify a time range (a slider with 2 movable ends): metrics will be re-computed after this using only the articles which were published in this time range

- `(/institutes/<institute_id>)`
  Show the institute details (name, country, metrics)
  Show a table with all the members of the institute, member names redirect to `/authors/<author_id>`
  Charts:
  - bar graph: #publications vs year
  - bar graph: #citations vs year

- `(/authors)`
  Shows a table with columns corresponding to the author entity's attributes
  Author names will be clickable and will redirect to `/authors/<author_id>`
  Rows will be sorted according to a default metric (total citations)
  User will be able to:
  - sort by any of the defined metrics (dropdown)
  - filter by the country (dropdown)
  - search by name (text input)
  - filter by topics (dropdown): only articles having these these topics will be used for re-computing the metrics
  - specify a time range (a slider with 2 movable ends), metrics will be re-computed after this using only the articles which were published in this time range

- `(/authors/<author_id>)`
  Show author's details (name, country, institutes), metrics, top-5 topics, and a table containing his publications (name, year, citations)
  Charts:
  - bar graph: publications vs year

    - bar graph: citations vs year
    - pie chart: number of publications vs topics (top-5 only)
    - line graph: h-index vs year
    - co-authorship graph for the author (nodes are authors, weighted undirected edge if they've collaborated, with weight = no. of co-authored articles)

- `(/articles/<article_id>)`
  Shows article entity's attributes along with the article's venue, topics and metrics.
  Charts:
    - bar graph: #citations vs year
    - citation graph for the article (nodes are articles, directed edge (v1, v2) if v1 cites v2)

- `(/venues)`
  Shows a table with columns corresponding to the venue entity's attributes and the list of topics for that venue
  Rows will be sorted according to a default metric (average citations)
  Venue names will be clickable and will redirect to `/venues/<venue_id>`
  User will be able to:
    - sort by any of the defined metrics (dropdown)
    - filter by the type (dropdown)
    - filter by topics (dropdown)
    - search by name (text input)

- `(/venues/<venue_id>)`
  Show venue details (name, acronym, type, topics, metrics)
  Charts:
    - bar graph: #publications vs year

- `(/topics)`
  Shows table with topic names and their metrics
  Rows will be sorted according to a default metric (average citations)
  Topic names will be clickable and will redirect to `/topics/<topic_id>`
  User will be able to:
    - sort by any of the defined metrics (dropdown)

- `(/topics/<topic_id>)`
  Shows topic name and metrics
  Charts (can be used to visualize a topic's popularity over time):
    - bar graph: #publications with this topic vs year
    - bar graph: #citations of publications with this topic vs year

Intended Classes of Users

- Public user (no login required): full access to the user interface described above except for actions which modify the data
- Venue admin: can login to a venue's page and add articles