**Group-3**

**Week-1**

1. We have 4600 article id’s and each article is cited for 100 times.
2. Sort the articles by year and title and then perform the statistical analysis.
3. Under statistical analysis we need to collect authors, citations where it has been published, intent(background, method and other types) and influential or non-influential.

**Part-2:**

Collect all the abstracts and separate them year by year and perform the analysis on each topic

**Tasks Done:**

1. Collected all the article id’s from the CSV file and appended it to the URL.
2. Read each URL from the CSV file and saved the data into a JSON file.

**Attendance:**

1. Akshitha Maddi
2. Akshatha Kamath
3. Vyshnav Kasa
4. Varun Chandra Dasari

You need to create several tables in the database

1. Frist table: article information

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Article\_id | Venue | Title | Year | Abstract | Doi | fieldsOfStudy | Topics (separate the topic by coma) | Download url (pdf\_url) | Total citation count | Total download | Influential citation count |
|  | SIGIR |  |  |  |  |  |  |  |  |  |  |
|  | SIGIR |  |  |  |  |  |  |  |  |  |  |
|  | SIGIR |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

1. Author information

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Article\_id | Author\_id | Author name | Author\_rank | Author\_url | Author\_institution | Author publication count (how many publications has the author published) |
|  |  |  | First |  |  |  |
|  |  |  | Second |  |  |  |
|  |  |  | Third author |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

1. Citation information

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Article\_id | Citing\_paper\_id | Citing\_paper\_doi | Citing\_paper\_title | Citing\_paper\_title | Citing\_paper\_venue | Citing\_paper\_year | Intent  (every intend one row) | Is influential |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

1. Reference information

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Article\_id | Reference\_paper\_id | Reference \_paper\_doi | Reference \_paper\_title | Reference \_paper\_title | Reference \_paper\_venue | Reference \_paper\_year | Intent  (every intend one row) | Is influential |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

**Week-2**

1. Use the url <https://api.semanticscholar.org/v1/paper/10.1145/511793.511795> and replace the article id’s to download the json data.
2. Parse the required data from the JSON File
3. Create a database in MySQL
4. Create tables using the above structure.
5. Save the data into the tables.

**Tasks for each member this week:**

**Akshitha Maddi :** Collecting all the articles and downloading the Json Data

**Akshatha Kamath:** Parsing the JSON data

**Varun Chandra and Vyshnav Kasa:** Creating and loading data into database.

**Attendance:**

1. Varun Chandra Dasari
2. Vyshnav Kasa
3. Akshatha Kamath
4. Akshitha Maddi

Analysis can be conducted:

(1) author analysis - the most influential authors and their co-authors. Number of articles for each author/their co-authors, and the topics (The purpose is to understand how a scholar in this area can impact the field through co-authorship/collaobration/mentoring)

(2) topic change over 2 years, 4 years, 5 years, and more. And then we can visualize the change of the topics. We need to do topic analysis of title and abstracts. However, I prefer the topic is more than just keywords, should be key phrases.

(3) citation analysis. The citation and download analysis can be performed and associate with (1). but can also be compared with journal impact factors. Considering the acceptance rate, we may want to ask some journal editors about their journal acceptance rate to make some comparison.

(4) Citation content analysis: what are the intents distribution of the citations (each year)? What are the influential citations distribution?