Lab 3 – Solr Queries

Archita Pathak

University at Buffalo

Announcements

- Dry run complete
  - Submit before deadline
  - Remove your IP's access to Solr instance when grading script is running.
- Final deadline extended
  - Submission on 15th Sept, 2019 (Sunday); 11:59 PM.
  - Grading script will trigger exactly at midnight.
  - If you are sure public IP you submitted for dry run is going to remain same for final run as well, you don't need to submit again.

Running Queries - POIs

q=*:*

fq=verified:true

rows=0

facet.field:poi_id

Returns list of all poi_ids in the indexed documents along with the count.

Running Queries – Replies -1

For each POI and each date:

q = poi_id:18839785

fq = tweet_date:[2019-09-07T00:00:00Z TO 2019-09-08T00:00:00Z} & verified:true

facet.field = id

Returns ids of all the tweets posted by the POI on a particular day.

For grading, we will be looking at last 15 days.

Running Queries – Replies - 2

For each id:

q = replied_to_tweet_id: <id>

Partial crediting:

1. More than 0 dates when all the tweets have at least 20 replies
   - Check for consecutive dates in the list. Maximum penalty = 0.04
2. There were no dates when all tweets have at least 20 replies
   - Check if there were dates when at least more than half of the posted tweets have at least 20 replies. If yes, partial credit = 0.04, else total score for this POI = 0

Running Queries – text_xx fields

q=text_en:the

Rows=0

It should not return any documents if indexed correctly.

Grading Summary and Dry Run results

- Not binary
  - Lots of partial crediting
- Exception message will be present in log files. Please debug yourself.
- If you have followed all the steps correctly for setting up instances, you won't encounter connection exceptions.