

INFO 6210

Data Management and Database Design

Database Project

Social Data News Site

Professor: Nik Bear Brown

In this project, you are assumed to be working for a company called *Nerd Analytics* and that you are completely in charge of the database. Your job is to create a social news site before June. You must model, gather, clean data and a particular domain (e.g. Games, Film, Databases, Cartoons, Baseball, Pokemon, Music, etc.) and integrate it with social media data about that domain (e.g. Twitter, Instagram, facebook, etc.) for). Each domain must have entities that represent people, places and things. For example, for games one must be able to model gamers, game developers, games, and addresses related to games or gamers. For music, one must be able to model music lovers, musicians and music companies, and addresses related to music.

Examples of social news sites are:

Fark <https://www.fark.com/>

Slashdot <https://slashdot.org>

Digg <http://digg.com/>

Reddit <https://www.reddit.com/>

Metafilter <https://www.metafilter.com>

Mix <https://mix.com>

Groups

This assignment can be done in groups of up to four.

Each person in a group must:

Represent and database data that represent different types of people, places and things. Note that two different people, places or things (i.e. two rows of data) are not two different types of people, places or things (i.e. two tables of data that represent different entities).

Gather social media data from at least one source per group member, that allows one to measure the questions asked of the project.

There are only two exceptions to the groups of up to four. Working on a hyperparameter database and working on a machine learning/AI jobs database.

Tags

Tagging is a process in which end users use free-form keywords to manually index content in an organic and distributed manner. Social tagging has rapidly become a popular practice in which users add free-form keywords to content in order to organize and categorize it. Social tagging is extensive on websites such as Social Media (YouTube, Twitter, Instagram, Snapchat, del.icio.us, Digg, Flickr, facebook, Google+, etc.).

Your database must be able to tag the social media data that you collect.

News and Trends

Your website must have a newsfeed that presents news relevant to your projected audience. This newsfeed must be updated automatically over time as new news comes in. It also must be able to show popular hashtags over time.

Design Requirements

Your submission must include:

- Sample data from every table.
- Data from a social media site.
- SQL for all of your inserts and queries.
- Any code and scripts you used.
- The TAs must be able to use execute the code and SQL.
- A brief README document explaining all of the files, the tests and their results and code.

Social Media (Twitter, Instagram, facebook, etc.) for Analysis

The project is to create a database, conceptual model, tables, data and queries that could support these twelve questions below for a particular domain

- i. What are people saying about me (somebody)?
- ii. How viral are my posts?
- iii. What posts are likely to be interesting to me?
- iv. What posts are like mine?
- v. What users post like me?
- vi. Who should I be following?
- vii. What topics are trending in my domain?
- viii. What keywords/ hashtags should I add to my post?
- ix. Should I follow somebody back?

Further to help us with the mess of social media tagging, the database will need to add tables that allow one to store syntactic and semantic information about tags.

Specifically, you will need to create tables for:

- I. Domain tags (tags in your domain)
- II. Synonyms (which tags are synonyms?)
- III. Mis-spellings (mis-spelled versions of words)
- IV. Semantic information (categories of tags)

Topics due:

Upload a paragraph with Database project idea, references, ER-diagram to Blackboard by March 12, 2019. This should include some Database/pictures along with an ER-diagram (conceptual model) that illustrate what you want to do. Only one person in group projects need to upload the report but the others in the group must upload a text file with the names of their group.

Progress reports:

In class session presentation of student Database project progress reports will be Project Progress two weeks before the final presentations. Progress presentations are about 10-15 minutes. You will also upload a progress report to Blackboard. Only one person in group projects need to upload the report but the others in the group must upload a text file with the names of their group.

Presentations due:

In class session presentation of student Database project progress reports will be Project Progress the last week of the semester before finals. Progress presentations are about 10-15 minutes. You will also upload a progress report to Blackboard. Only one person in group projects need to upload the report but the others in the group must upload a text file with the names of their group.

Projects Due:

Database Projects are due at the end of finals week. Only one person in group projects need to upload the project but the others in the group must upload a text file with the names of their group.

Grading Rubric:

The following breakdown will be used for determining the score for the Database project:

Assignment	Points
Topics/option choice/ ER-diagram	100
Progress report	100
Presentation	100
Database Project	1000

Submission of Assignments

Your submission must include the code, data, and diagrams along with a write-up. The students last names MUST be part of the zip file name that is uploaded. You MUST include your and group name or your name the name of the zip file you upload.

You will submit your assignments via BlackBoard. Click the title of assignment (blackboard -> assignment -> <Title of Assignment>), to go to the submission page. You will know your score on an assignment, project or test via BlackBoard. BlackBoard represents only the raw scores. Not normalized or curved grades.