Mark Conley

CS336 Principles of Information and Data Management

Sentiment Analysis

MongoDB

MongoDB is a NoSQL document oriented database that utilizes a key-value store behavior instead of a relation approach. Data is stored in JSON-like objects called BSON which allows for tremendous flexibility.

Benefits of using MongoDB over MySQL

- Schema-less design which allows the programmer to have more flexibility
- Allows for greater scalability and availability

Sentiment Analysis: mongoDB

- Used PyMongo to connect with mongoDB instance
- Created set of positive words and negative words from ‘positive words.txt’ and ‘negative words.txt’
- Get sentiment value for each review and created JSON file with id, review, and category

Sentiment Analysis: MySQL

Schema

- Id and review content
- Id, foreign key and, word count
- Sentiment value and word can be counted by referencing word

SQL Query

- Join each table and have words that have sentiment value and then count for each review

Conclusion

MongoDB vs MySQL

- PyMongo allows us to very easily build database when data is in JSON format
- Mongo is more efficient and faster in this case because joins in SQL are very expensive and time consuming
- Given the data is already in a SQL database the MySQL is much better because querying the database is very simple because of its relational structure.