**[Project Title]**

**Project Deliverable 1: Requirements Analysis**

**9/18/2022**

Group 03

*Contributors:* Arunabh Bhattacharya,

Kathleen Tiley,

Benjamin Weiss, and

Jonathan Williams

Table of Contents

[Overview 2](#_Toc114223828)

[Project Description 2](#_Toc114223829)

[Requirement for a Database System 2](#_Toc114223830)

[Web-Based User Interface Functionality 2](#_Toc114223831)

[Data 2](#_Toc114223832)

[Software 2](#_Toc114223833)

[Trend Analysis and Queries 2](#_Toc114223834)

[Trend Analysis Goals 2](#_Toc114223835)

[Trend Queries 2](#_Toc114223836)

# Overview

## Project Description

Businesses across many industries use database systems to organize data collected and determine trends, to obtain better insight into their customers and industry to make more informed business decisions. Electronic commerce websites often use data within their platform like sales, and reviews/ratings, to make decisions about what products to keep on their platform and what to recommend customers. Yet, the effectivity of this analysis is dependent on user engagement in their specific platform. Online communities can provide more information for businesses to supplement data collected to make better decisions. Social media platforms can provide a rich source of data as we can expect higher user engagement, and more information into the preferences of specific online communities. Goodreads.com is an online social media platform website where readers can rate, recommend, and share opinions on different books. Easily organized data like book ratings, books read by users, and user specific ratings, could be used to obtain better insight into the community’s sentiment to different books.

## Requirement for a Database System

A database system used should have user specific information to evaluate user sentiment to specific products, in this case books, which are computationally easy to analyze. Overall data on the books would also give insight into their specific perception in the community.

## Web-Based User Interface Functionality

The web based interface will use queries to interact with the database, and generate figures to illustrate the trends in the data selected by the user.

## Data

In 2017 a Ph.D. researcher at the University of California San Diego, Mengting Wan, scraped the largest currently available data from goodreads.com. Wan collected information of the books on the website, like associated generes, synopsis, user reviews. Additionally, Wan collected information on user-book interactions, and reviews organized by individual users. Wan has made this data openly available online through an open-source project on github.com [\*\*link\*\*]. The data is currently posted in a non-standard format similar to JSON, which the team will have to parse and clean to successfully set up the Oracle SQL database.

## Software

We will need to use Oracle SQL databases to make queries on the data and tease out possible trends related to user sentiment on books.

# Trend Analysis and Queries

## Trend Analysis Goals

## Trend Queries

**Phase 1**

**The summarized requirements from Prof:**

propose and understand an appropriate project topic, identify its main data management needs, explore and motivate its potential for interesting queries, and analyze the needed user functionality.

1. What are the main functions that the web-based user interface should provide?
2. How do the different functions work together? Sometimes there are dependencies between different functions.
3. Which real-world data are needed to support the functions identified before?
4. Can such real-world data be found in the Internet?
5. What (colloquial) queries are important for the application?
6. Which public domain and/or proprietary software is needed to perform the task? (The database system used must be CISE Oracle.)

(PDF file) that presents a clear and structured description and motivation of the project topic and requirements that the software solution should later fulfil.

The focus of this project is supposed to be on the database part and not so much on the application part. focus on database queries that evaluate large volumes of stored data. Of course, the application part must be highly functional, and the different user functions must cooperate nicely together. However, a fancy layout design of the user interface is not required but appreciated.

It is important that each group demonstrates in their deliverable that their application would really benefit from database support and that new information (such as trends) can be derived from the stored data. A simple retrieval of data from the database (that is, search) or the pure connection of different tables (that is, joins) are not sufficient. As an example, let us assume that a group selects a sales application as their project topic and stores many daily sales numbers in their database. Of course, one can search for sales data of interest in the database and display them in the user interface. But searching only identifies an interesting subset of all data stored in the database. DBMS are specialized for search tasks, and the respective SQL queries are relatively simply structured. This project aims at more interesting queries that, first, derive new information which is not explicitly stored in the database but can be derived from the data in the database by computations and, second, represent trends (see Section 3).

Arunabh-

Since we are required to do trend analysis and compute the changes in data over a period, based on the datasets, I could think of the following:

For point 5, some of the colloquial queries could be:

* What was the average number of books read during a certain period, ordered by genre?
* Which genres had the highest ratings during a certain period?
* Which genres received more reviews during a certain period?
* What was the percent of shelved books by genre?
* What was the ratio of read and shelved books by genre?