COMP39/9900 Computer Science/IT Capstone Project
School of Computer Science and Engineering, UNSW

Project Number: P11

Project Title: A System for Predicting Information Evolution in Social Networks

Project Clients: Jiaojiao Jiang

Project Specializations: Software development; Big data analytics and visualization.

Number of groups: 4

Background:

A "meme" is a piece of transmissible information that spread and evolve analogously to genes through communication. The goal of this project is to develop a system that can predict the evolution of information within social networks. By analyzing the spread and transformation of data, this system aims to provide insights into how information evolves over time, enabling better understanding and forecasting of trends, virality, and influence patterns within digital communities.

Requirements and Scope:

1. Data Acquisition: Building pipelines to collect and store data from social networks such as Twitter, Facebook, Instagram, and Reddit.

2. Data Processing: Cleaning and preprocessing data to ensure quality and consistency for analysis.

3. Algorithm Development: Implementing the Yule model to detect, analyze, and predict information spread, including text analysis, trend detection, and network analysis.

4. System Development: Integrating all components into a cohesive system with a user-friendly interface for end-users.

Data Collection Module:

1. Social Media Integration: Integrate with APIs of major social networks such as Twitter, Facebook, Instagram, and Reddit.

2. Data Types: Collect various types of data, including text posts, images, videos, comments, likes, shares, and timestamps.

3. Historical Data: Ability to retrieve historical data within the limits set by each platform.

4. Data Storage: Store collected data in a scalable, efficient database.

Data Processing Module

1. Data Cleaning: Implement processes to remove duplicates, correct errors, and standardize

data formats.

2. Preprocessing: Tokenization, stop-word removal, stemming/lemmatization for text data.

3. Data Transformation: Transform raw data into structured formats suitable for analysis, such as JSON or CSV.

Visualization Tools

1. Dashboard: Create an interactive dashboard for users to visualize trends, predictions, and network dynamics.

2. Graphs and Charts: Provide various types of visualizations such as line graphs, bar charts, heat maps, and network graphs.

3. Customizable Views: Allow users to customize views based on parameters like time range, data source, and specific topics.

Required Knowledge and skills:

Programming Languages

1. Python: For data processing and backend development.

2. JavaScript: For frontend development.

3. SQL/NoSQL: For database management and querying.

Adaptability and Collaboration

1. Learning New Technologies: Willingness to learn and adapt to new technologies and methodologies as the project evolves.

2. Team Collaboration: Ability to work effectively within a multidisciplinary team.

Expected outcomes/deliverables:

source code, user guide, and documentation.

Supervision:

Jiaojiao Jiang

Additional resources: