Tajziya.org

Software Requirement Specification (SRS)

(https://github.com/marxbilal/FYP---Tajziya-.git)

BY

Team Lead: Bilawal Ali/17278

Member(s): Syed Muneeb Ur Rehman/17928

Hamraj Singh/18013

# PROJECTS COMMITTEE (PC)

SUPERVISOR: Ma’am Zarmeen Naseem

CO-SUPERVISOR: NAME

Member(s): NAME(S)

**SUBMITTED TO**

pROJECT Manager – FYP

**ON**

DATE (21/11/2021)

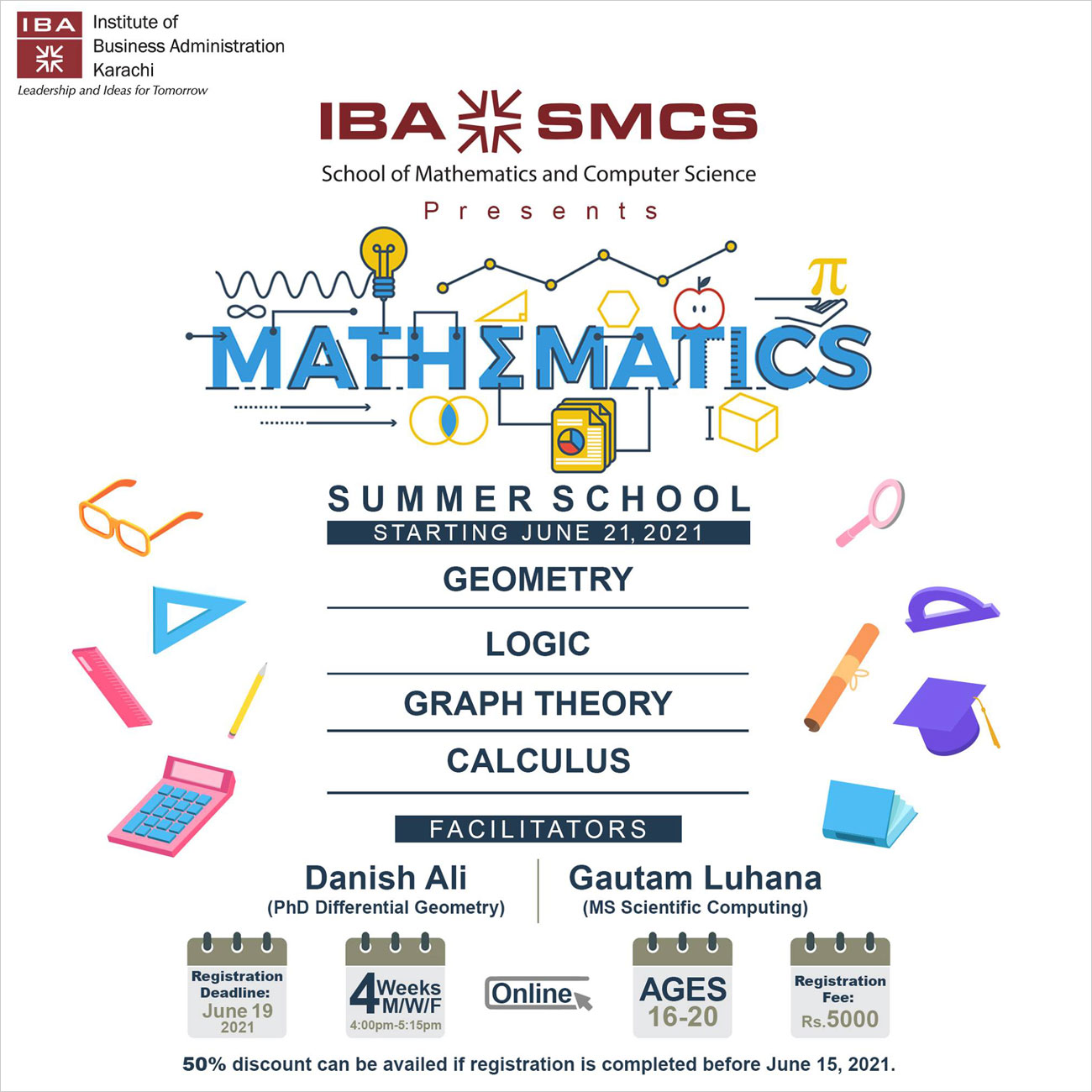


Table of Contents

[**PROJECTS COMMITTEE (PC)** 1](#_Toc88383327)

[1. Problem Statement 3](#_Toc88383328)

[2. System Requirement 3](#_Toc88383329)

[a. List of Functional Requirements 3](#_Toc88383330)

[b. List of Non-functional Requirements 3](#_Toc88383331)

[c. User Interface Requirements 4](#_Toc88383332)

[3. Functional Requirements Specification 4](#_Toc88383333)

[a. Stake Holders 4](#_Toc88383334)

[b. Actors and Goals 4](#_Toc88383335)

[c. Use Cases 5](#_Toc88383336)

[Use case 1: Searching for news article 5](#_Toc88383337)

[Use case 2: Exploring news under a topic 5](#_Toc88383338)

[Use case 3: Viewing Clusters 5](#_Toc88383339)

[Use case 4: Viewing Tag Cloud 6](#_Toc88383340)

[Use case 5: Viewing Sentiment 6](#_Toc88383341)

[Use case 6: Personalizing My Feed 6](#_Toc88383342)

[Use case 7: Viewing My Feed 7](#_Toc88383343)

[Use case 8: Exploring trending topics 7](#_Toc88383344)

[4. User Interface Specification 8](#_Toc88383345)

[5. Domain Analysis 13](#_Toc88383346)

[a. Pre-processing - TF-IDF 13](#_Toc88383347)

[b. Clustering: K-means for document clustering 13](#_Toc88383348)

[c. Topic Modeling 14](#_Toc88383349)

[d. Stanza 14](#_Toc88383350)

Software Requirement Specification (SRS)

# Problem Statement

We are concentrating on Urdu news media analysis in order to uncover patterns in news publishing. As a result, in this project, we propose a method for clustering news from mass media such as (Newspapers, social platforms- twitter, blogs, etc.). We then study the parameters of the clusters after they have been formed. The primary purpose of our project is to create a novel clustering technique for mass media and obtain the corresponding topic distribution with the help of NLP.

In today's news environment, it has become very easy to retrieve information from almost any story. However, for users, when it is not necessary to search for a specific news story, but rather to select one that matches the user's preferences, information overload makes it impossible to locate the desired article. That is a big problem that can lead to information blindness, as the user is limited to the narrow viewpoints of the feeds he or she accepts to follow and misses the broad panorama that exists outside.

Hence the major goal of this project is to extract some kind of meaning from news stories in order to connect them, understanding about the subject they're discussing and how close articles are based on the words they use in a programmatic fashion. Our proposed software will cater for the above problems and for that we are using Natural Language Processing techniques such as Tag Cloud: User will be able to see just in a glance all the major stories that happened in a day, Trending Topics: what news is trending today, Sentiment: Whether specific news is positive or negative and a personalized user feed in which user will only see subscribed category news.

# System Requirement

## List of Functional Requirements

1. Filtering news with respect to categories
2. Overviewing Trending Topics
3. Timeline of all news-covered under a specific topic
4. Visualizing sentiment of news topics and keywords
5. Observing high-frequency words, topics, or categories with respect to their negative/positive context
6. News-clustering (Big clusters for widely covered topics.)
7. Setting preferences for a customized newsfeed.

## List of Non-functional Requirements

1. Consistent design (fonts, colors, icons, etc.)
2. Comprehensive features
3. Thorough Navigation
4. Color coding (red for negative sentiments and green for positive)
5. Effortless zooming/filtering in sentiment and cluster visualizations

## User Interface Requirements

User Interface is a browser-based window that will enable access to this application. The interface needs to be helpful for users in achieving the following tasks:

1. **Browsing News Category:** A menu of broad-categories such as sports, politics, and technology etc., is required to let user drill-down through the desired news-category.
2. **Trending News:** A window section needs to be added to the homepage to give a quick glimpse of hot topics.
3. **Tag Cloud:** A visualization representing the most frequently used words, with respect to their contextual sentiment.
4. **Sentiment Analysis:** A separate page enabling users to overview the sentiment news about recent topics. And filters enabling users to drill-down in a specific category or topic or keyword to run sentiment analysis.
5. **News Cluster:** A bubble chart visualization representing clusters of news stories. Bigger the bubble, higher the coverage of story.
6. **News Feed**: A page where the user’s subscribed topics and categories will be visible. The user will be able to change their preferences and the page will update in real time with live stories.

# Functional Requirements Specification

Provide use cases based on 1 and 2 above.

## Stake Holders

**Project Manager:** Ms. Zarmeen

**Project Team:** Bilawal Ali, Hamraj Singh, Syed Muneeb Ur Rehman

**External:** National News Outlets (Urdu): Jang, Dawn, Bol News, DailyKhabrain, etc.

International Outlets (English): BBC, Tribune, etc.

## Actors and Goals

**Actors:**

1. Users
2. News Outlets
3. Urdu News journalists

**Goals:**

* To view the trending topics
* To view emerging topics under a news category
* To read articles/news stories under a specific topic
* To search news articles
* To run the sentiment analysis on news headlines
* To run sentiment analysis on a keyword
* To observe the most-frequently used keywords and their contextual sentiment
* To cluster all the news under categories
* To personalize the news feed
* To save news articles to read later

## Use Cases

### Use case 1: Searching for news article

**Overview:** A user wants to search for a specific news article using certain keywords.

**Normal Flow:**

1. User selects search bar and types their search keyword
2. Page redirects to search page with articles
3. Users click on articles they find relevant
4. Article page opens in new tab

**Alternative Flow:**

3a. User Selects Category and Topic filter

3b. Page refreshes to show filtered content

**Post Condition:** User can view the news article

### Use case 2: Exploring news under a topic

**Overview:** User can select from a list of auto-generated topics to view news articles based on that topic

**Normal Flow:**

1. User selects the Categories from the secondary navigation bar
2. Page redirects to the category page where different topics are displayed in forms of buttons
3. User selects a topic
4. Page redirects to topic page where articles are shown of that topic
5. User clicks on article they find relevant
6. Article page opens in new tab

**Post Condition:** User can view the news article for the topic

### Use case 3: Viewing Clusters

**Overview:** User can view the broad topics/categories of news article and also view the size of each topic/category and view the news headline name and the overall topic.

**Normal Flow:**

1. User selects the Cluster tab from the navigation bar
2. Page redirects to the cluster page where at first cluster for all trending topics is shown
3. User hover over cluster name
4. The hovered cluster gets a bold outline

**Alternative Flow:**

3a. i. User selects the filter option by category

ii. User selects the filter option by topic

3b. i. Clusters updated and show cluster of topics for that category

3b. ii. Clusters updated show cluster of articles for that topic

**Post Condition:** User can view different clusters and see their names assigned

### Use case 4: Viewing Tag Cloud

**Overview:** user will be able to view tag cloud of similar news keywords

**Normal Flow:**

1. User selects the Tag Cloud tab form the top navigation bar
2. Page refreshes to load Tag Cloud page

**Post Condition:** User is able to view and interact with the tag cloud.

### Use case 5: Viewing Sentiment

**Overview:** To visualize the sentiments of headlines of all the captured news.

**Normal Flow:**

1. User clicks the sentiment button
2. A graph maps all the recent headlines on a x-scale of sentiment and y-scale of trending news.

**Alternative Flow**

2a. User applies filter on the news categories.

2b. User applies filter on the news-topics.

2c. User searches a specific keyword under selected categories and news-topics.

**Post Condition:**

A visualization showing sentiments of filtered (by default: recent news> appears on the page.

### Use case 6: Personalizing My Feed

**Overview:** Users Signs up to customize the preferred news categories.

**Normal Flow:**

1. User clicks on Login button
2. Login page appears
3. User enters login id and password and clicks sign in

**Alternate Flow:**

3a. Password doesn’t match and user is redirected to the password reset page.

3b. User clicks forgot password button and is redirected to the password reset page.

3c. User clicks the register button

* 1. User enters the profile info
  2. User selects the topics and news categories of interest.
  3. User clicks create profile button

**Post Condition:** User is redirected to the My Feed page.

### Use case 7: Viewing My Feed

**Overview:** User views the personalized news according to his interest.

**Normal Flow:**

1. User scrolls down through the recent headlines.
2. User observes the trending topics
3. User observes the rank and immediate trend(inclined/declined/stable) of a topic.

**Alternate Flow:**

1a. User clicks on a headline to read the full coverage.

2a. User clicks on a trending topic

* + 1. User is redirected to the topic page
    2. User browses headlines under the chosen topics.
    3. User clicks on a headline to read the full coverage

**Post Condition:** My Feed page is available to the user.

### Use case 8: Exploring trending topics

**Overview:** User sees gets to see and click the trending topics to read headlines under them.

**Normal Flow:**

1. User clicks on a trending topic

* 1. User is redirected to the topic page
  2. User browses headlines under the chosen topics.
     1. User clicks on a headline to read the full coverage

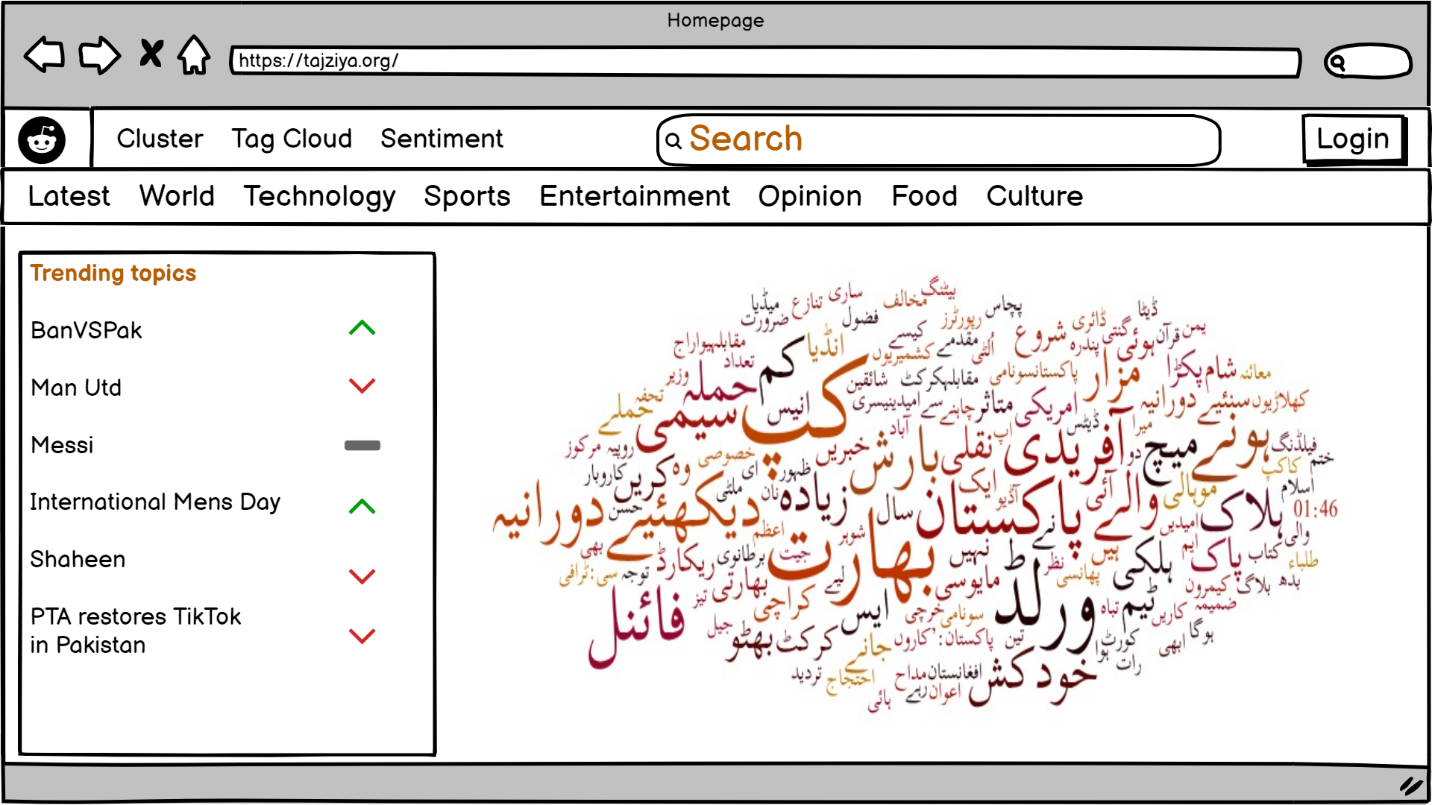
**Alternative Flow:**

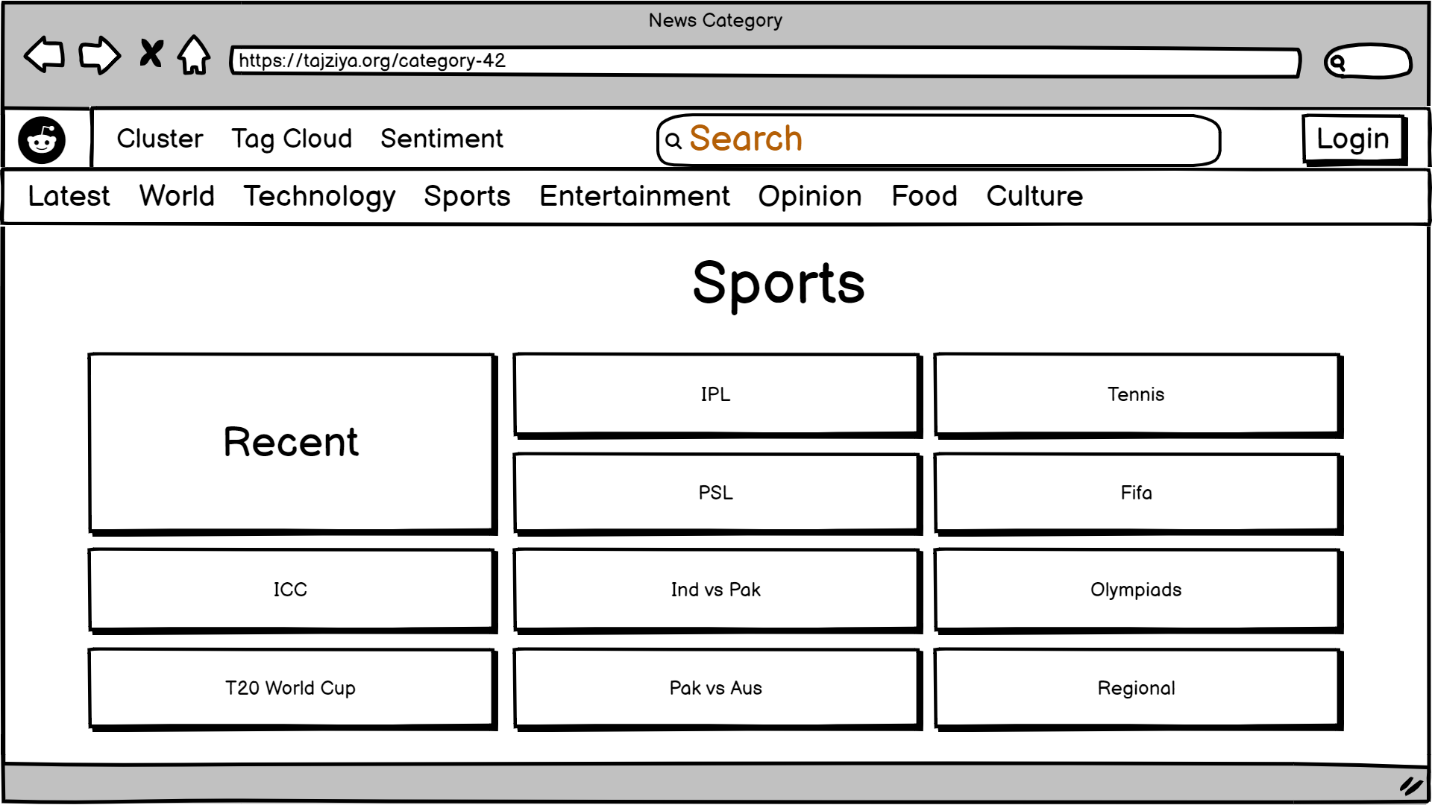
1a. User goes back to the My feed/Home page

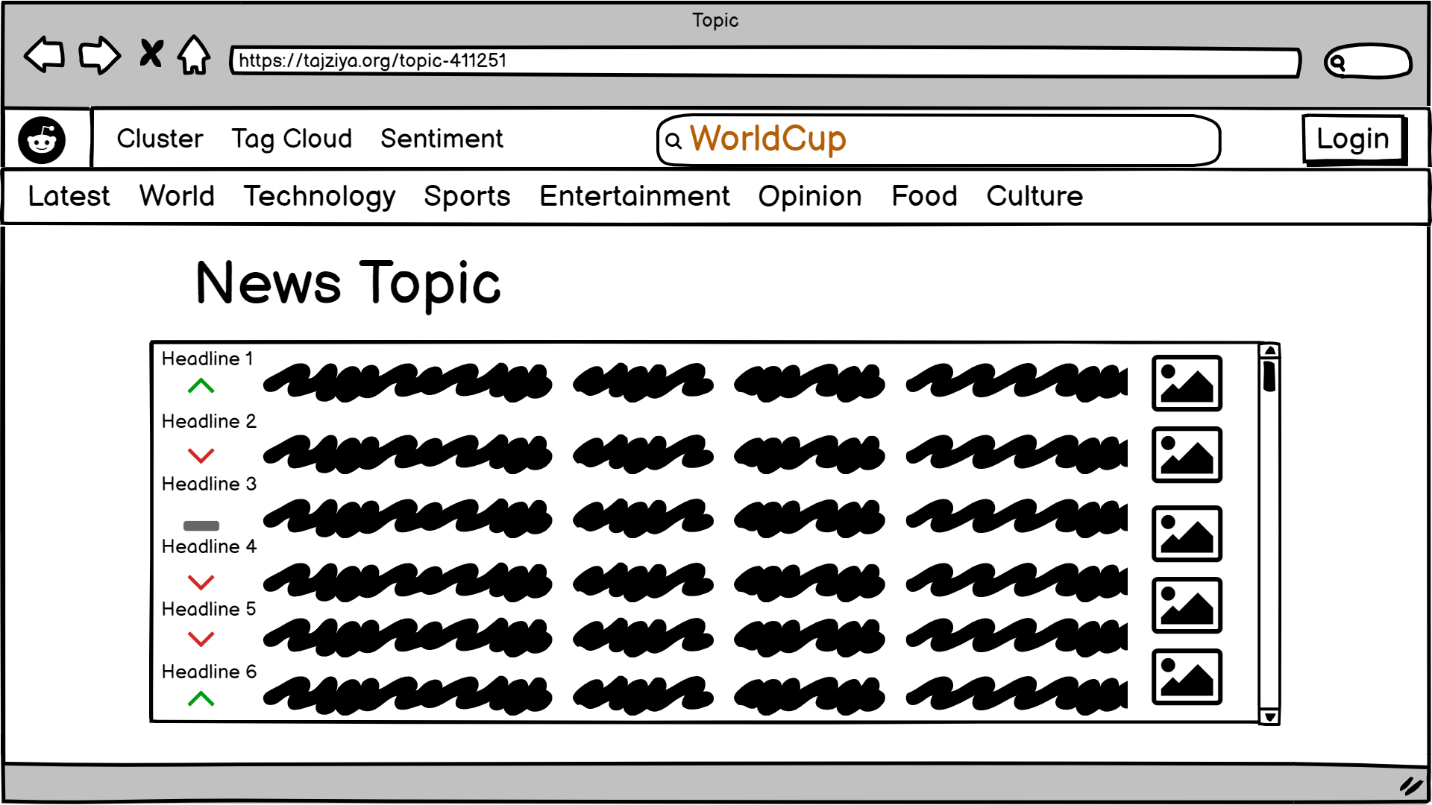
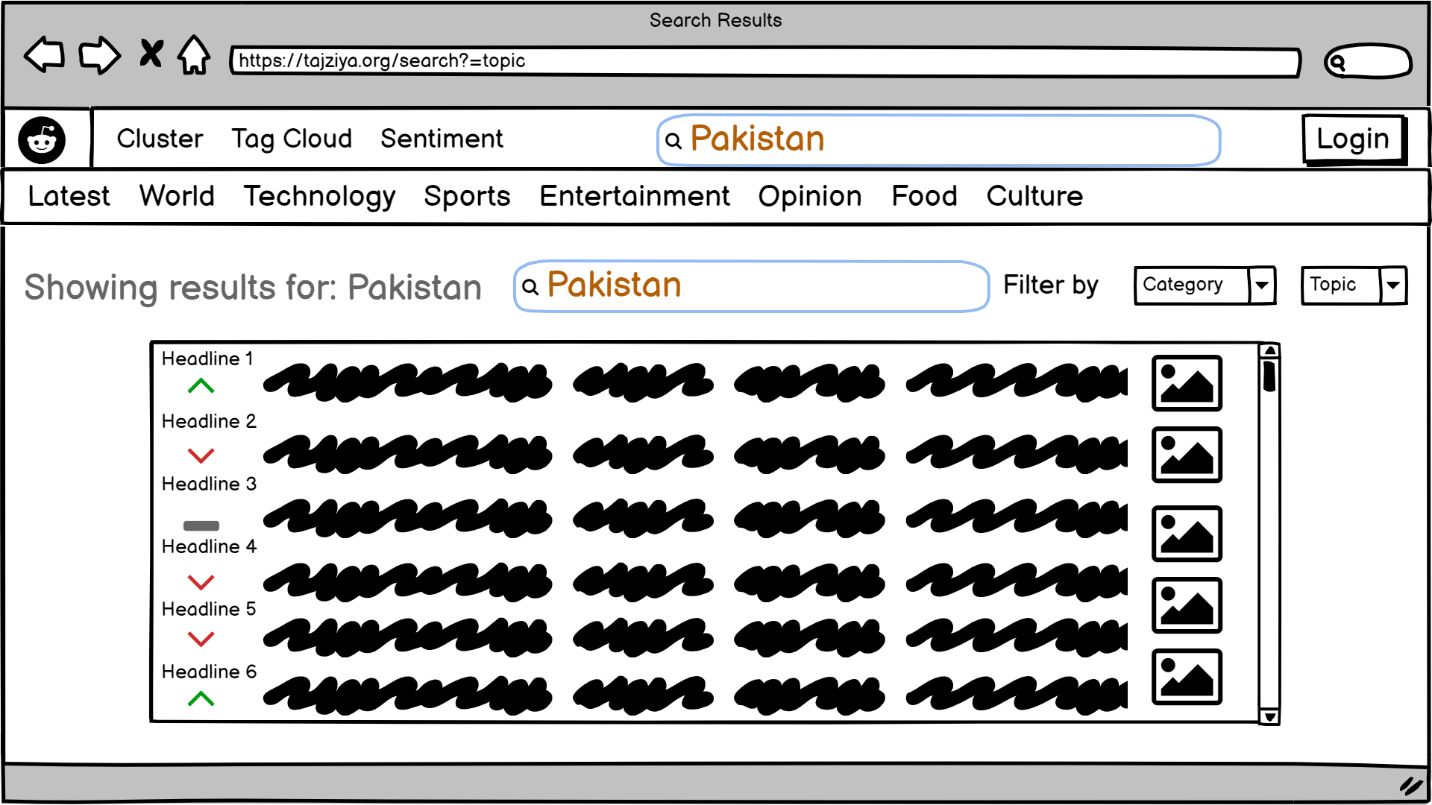
**Post Condition:** A full coverage of story under the topics is accessible to the user.

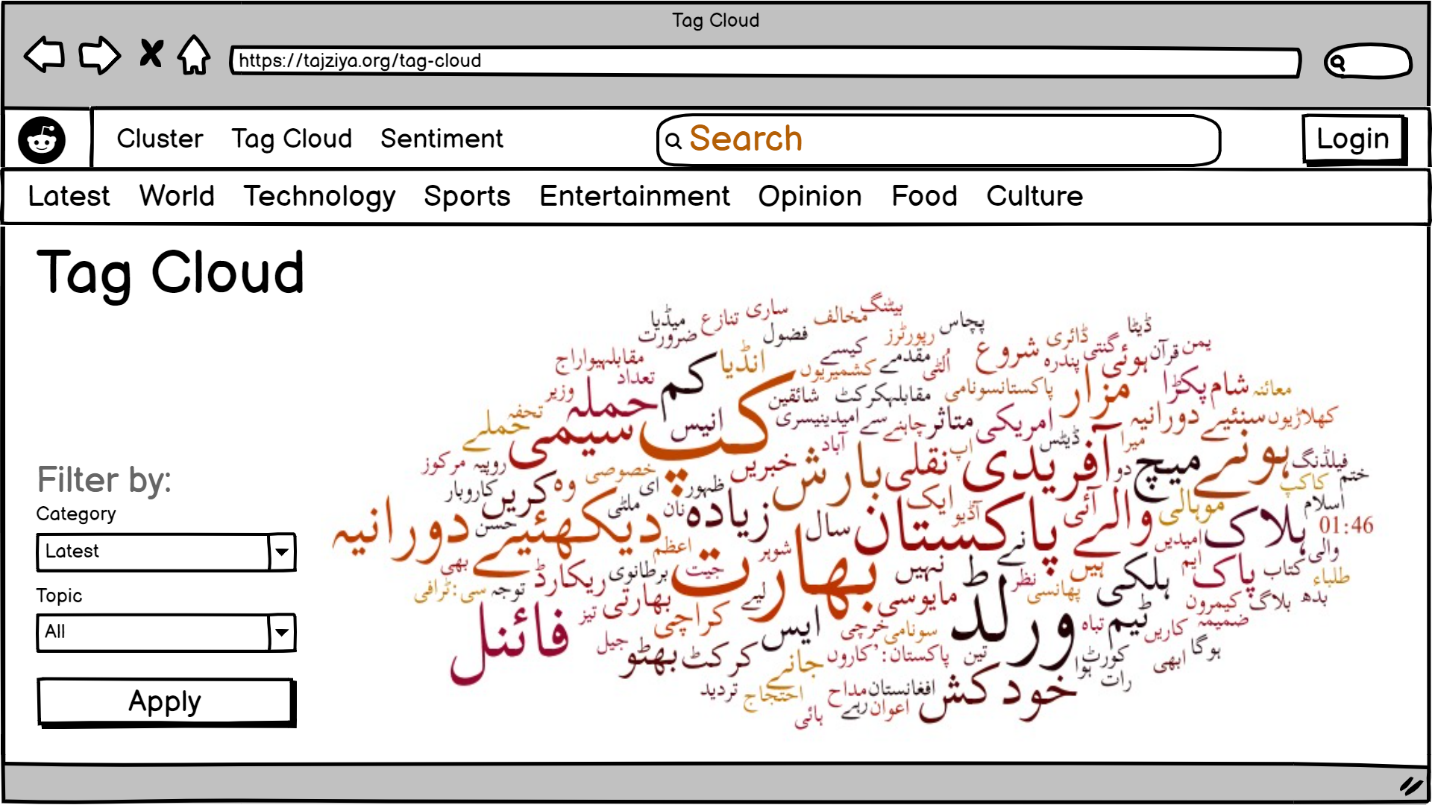
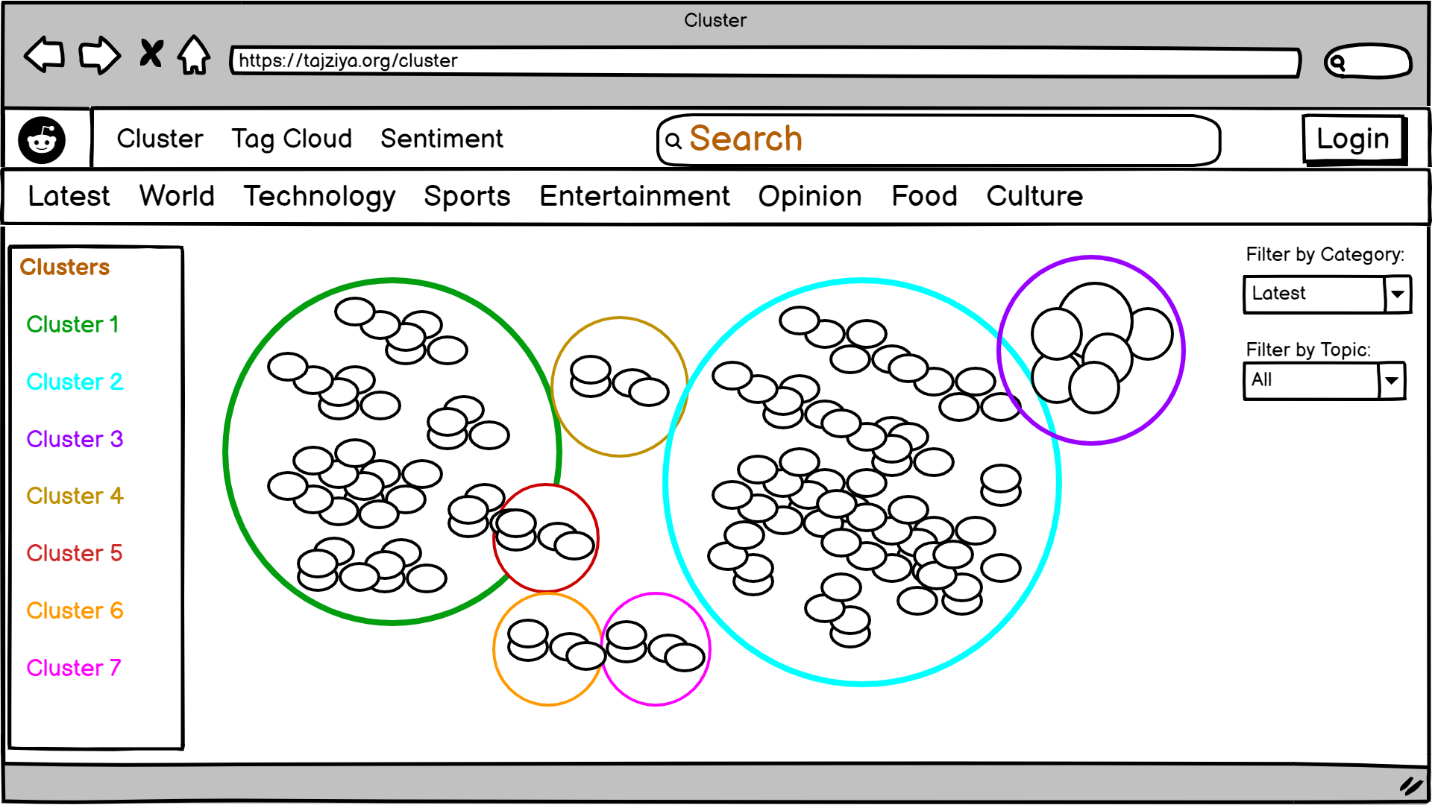
# User Interface Specification

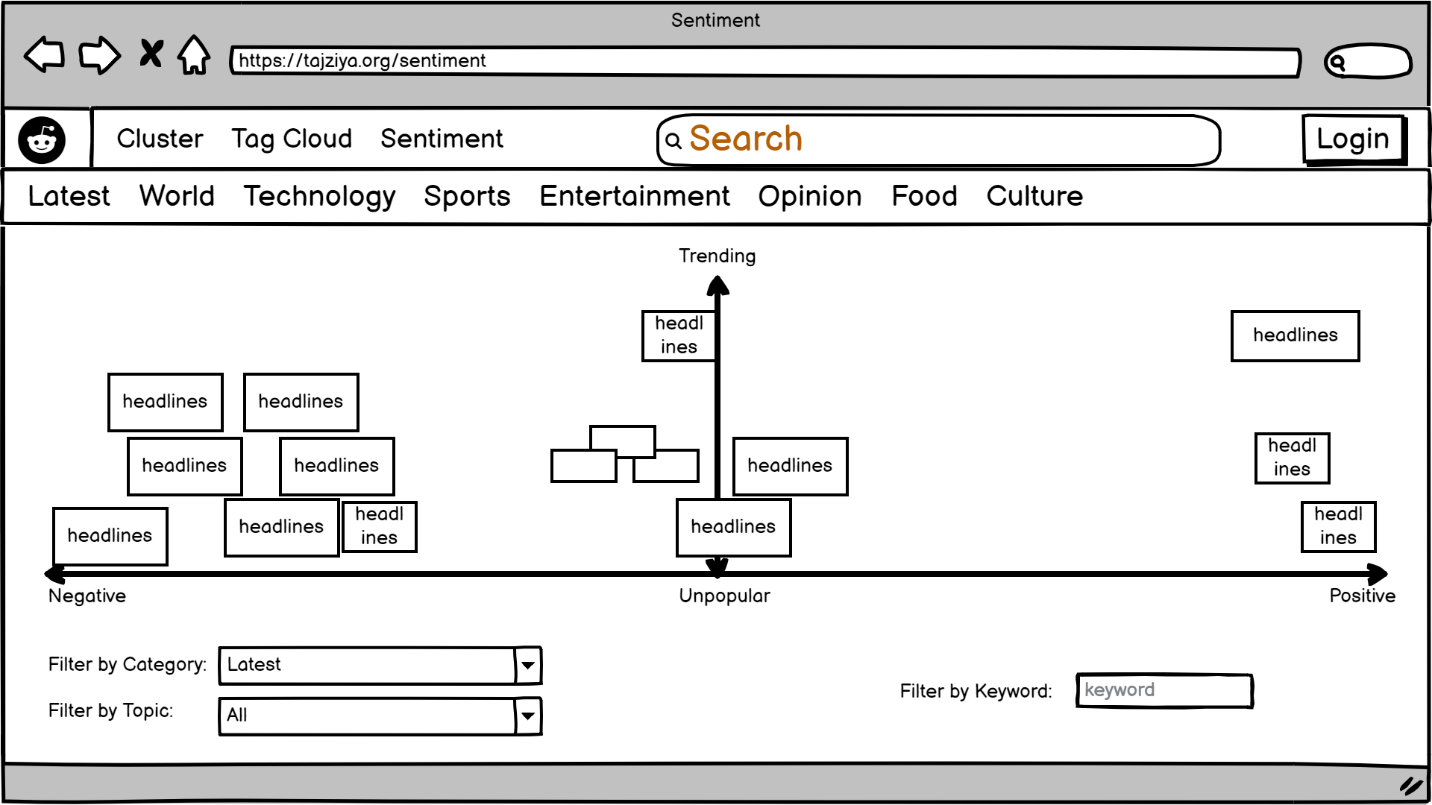
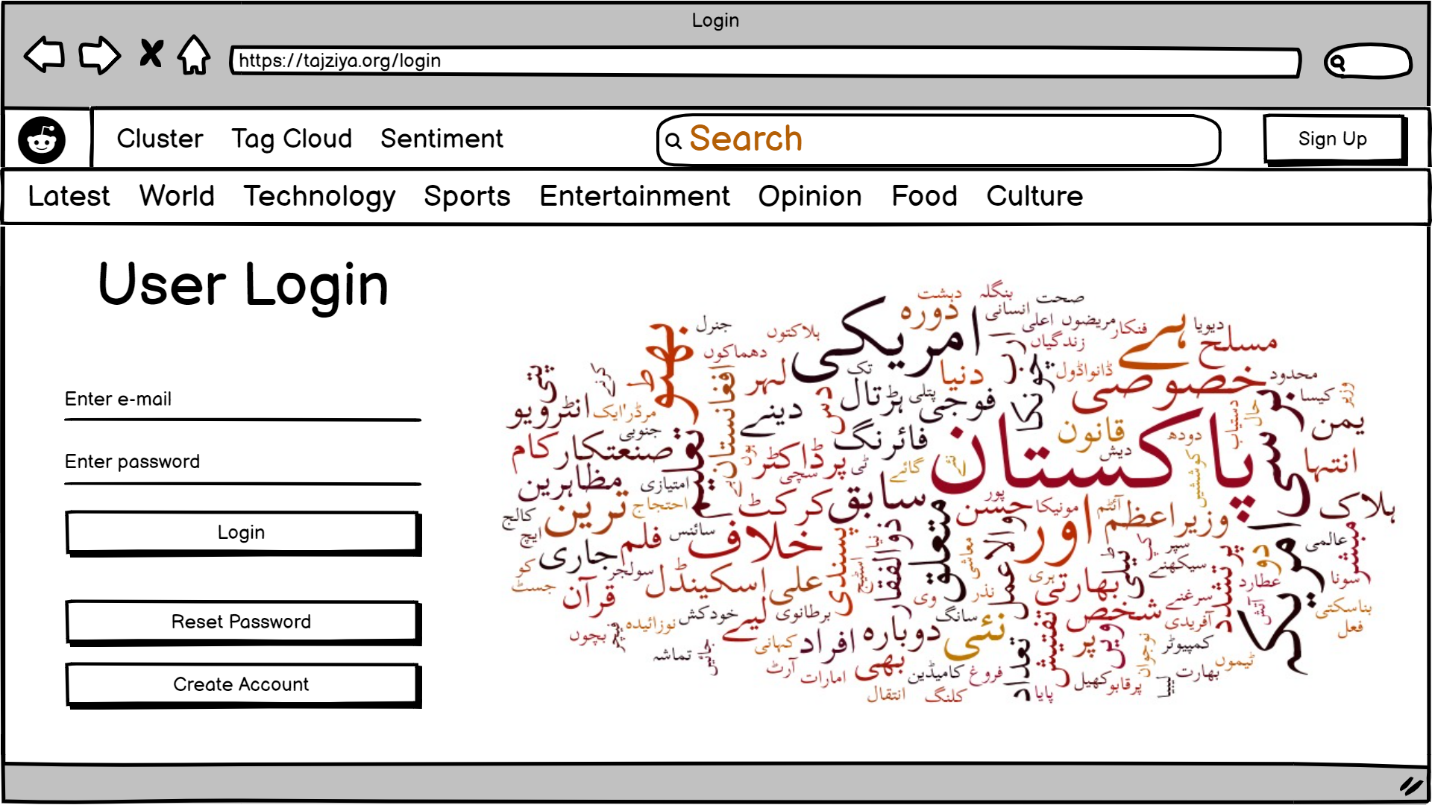
The final UI of our app is as follows:

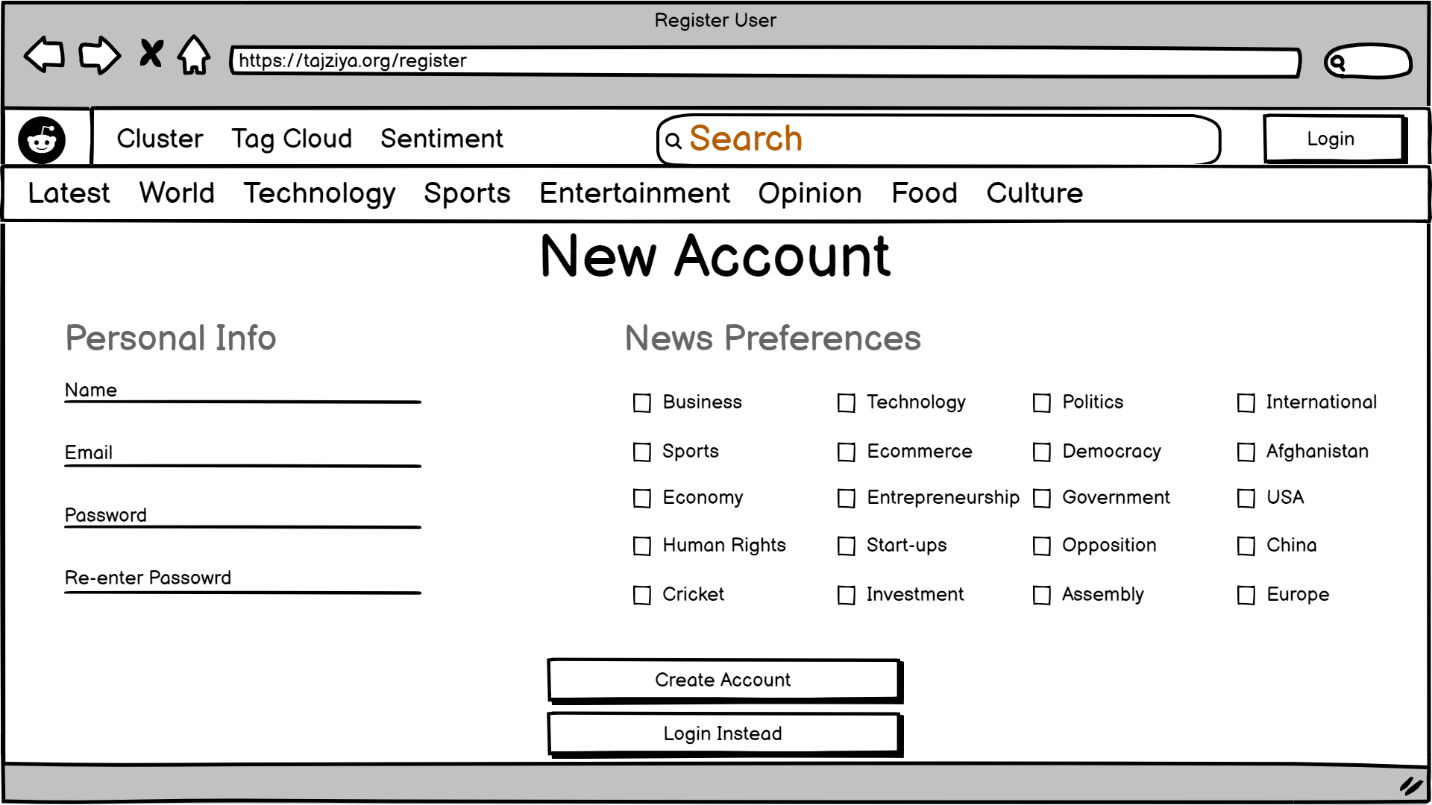
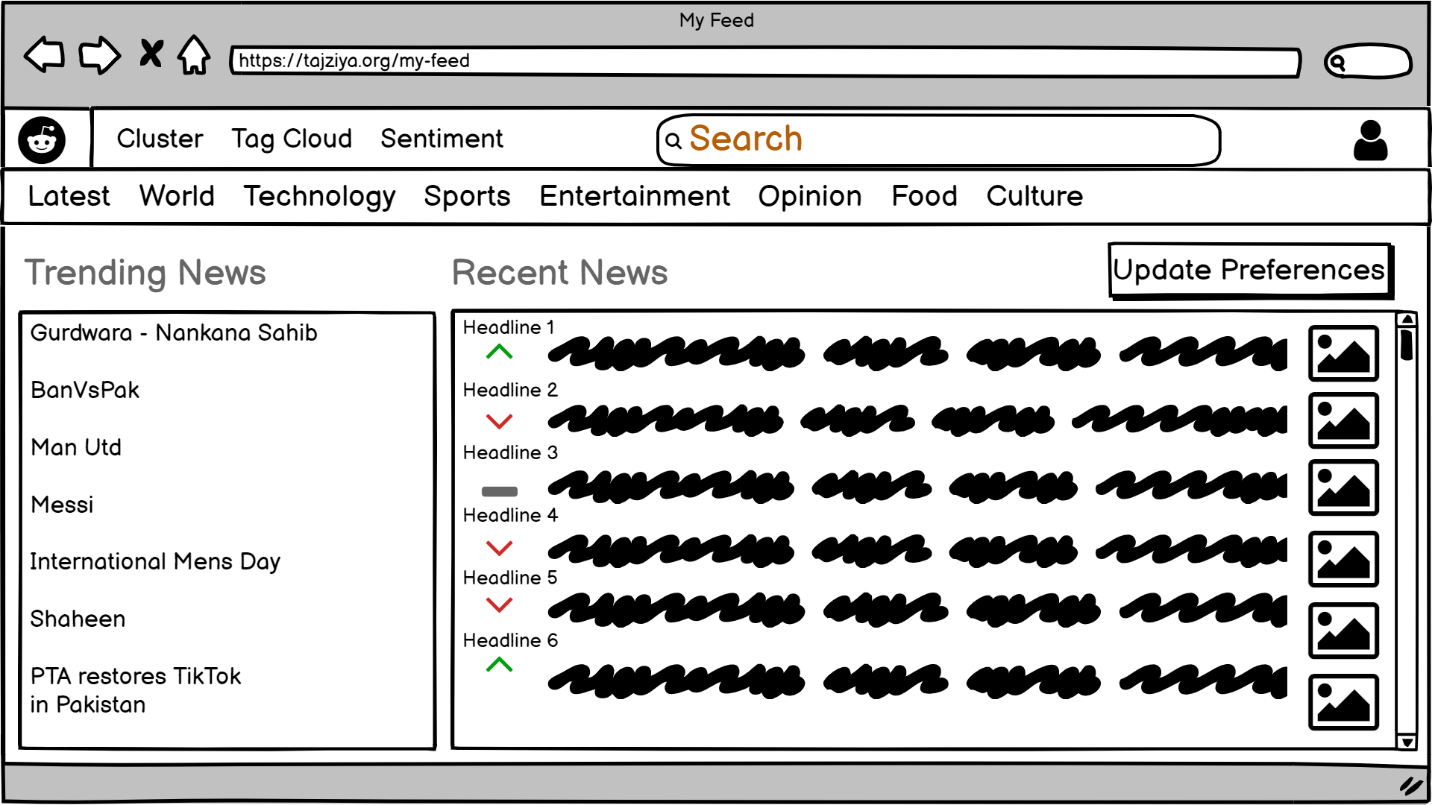








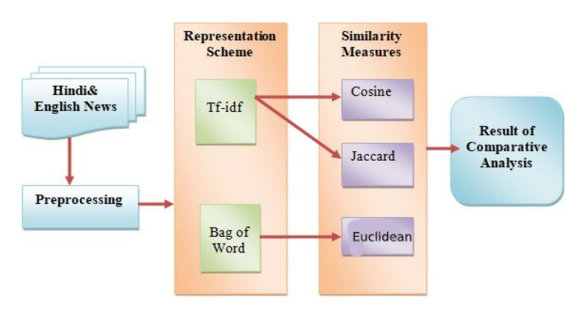


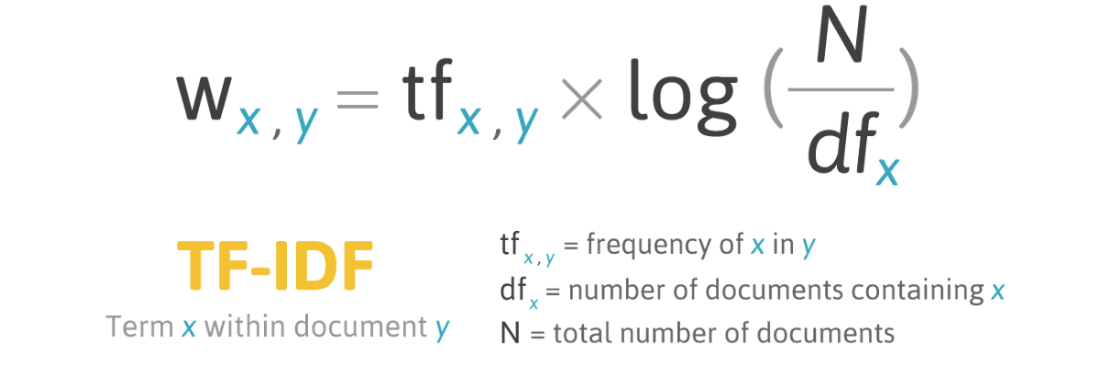


# Domain Analysis

## Pre-processing - TF-IDF

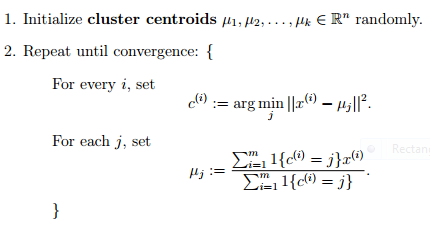
Before being presented in a more structural style, the textual news data is pre-processed. TF-IDF and Bag of Words are the two representation approaches for generating features from text that are explored in this study.





## Clustering: K-means for document clustering

Clustering is an analysis technique to identify information about structure of the data. K-means is an algorithm that partitions data into predefined k number of distinct non-overlapping clusters.

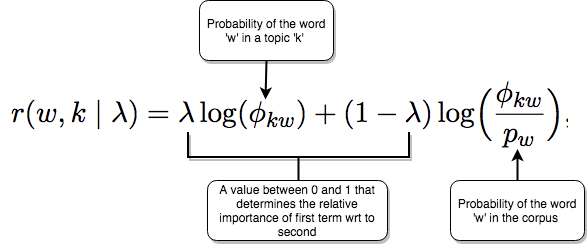
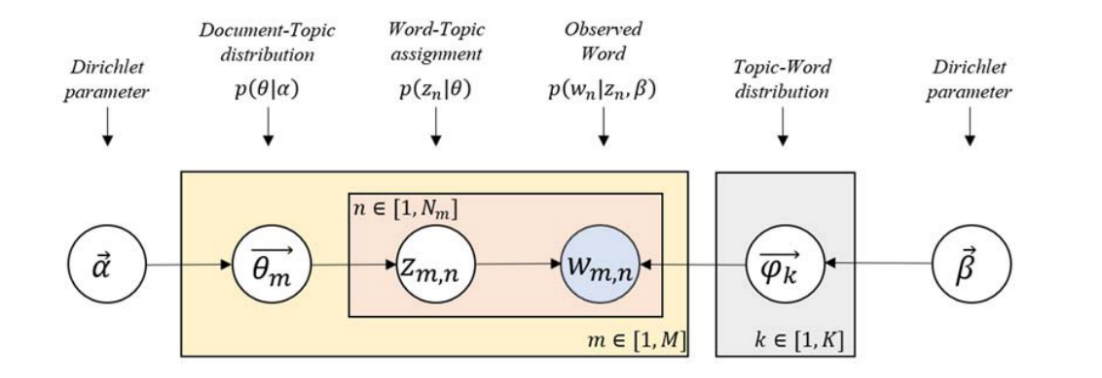


## Topic Modeling

Topic modelling is an unsupervised document categorization method, which finds natural groups of objects (topics) even when we don't know what we're looking for.

Diagram

Description automatically generated



## Stanza

Stanza is the python library for natural language processing in multiple languages including Urdu. There are multiple tools available in the library to convert string into lists of sentences and words and process them into their base word forms, parts of speech tags and other structural features and more.

