Project: TrendCatcher - Identifying Trending Topics for Engaging Social Media Posts

1. Project Overview

TrendCatcher is a software tool that identifies trending topics related to innovation on social media platforms (Thread, Twitter, and Pinterest) to help users create engaging social media content. It leverages automation, web scraping, and machine learning to analyze large amounts of data and deliver prioritized insights.

2. Technology Stack

- Social Media Platforms: Thread, Twitter, Pinterest
- Automation: Selenium
- Web Scraping: Beautiful Soup
- Text Preprocessing and Tokenization: BERT Model
- Topic Clustering and Relevance: K-means Clustering
- Frontend: Gradio
- Data Storage: MongoDB

3. System Architecture

1. Data Acquisition:

o Selenium automates browser interaction to navigate and search for relevant content on each social media platform.

o Beautiful Soup extracts text and relevant metadata from the scraped content.

2. Data Preprocessing and Tokenization:

o BERT Model cleans the text data by removing noise, correcting errors, and normalizing text.

o Tokenization breaks down the text into individual words or phrases for further analysis.

3. Topic Clustering and Relevance:

o K-means clustering groups similar posts based on their content and metadata.

o The algorithm measures the relevance of each cluster to the "Young Innovator" keyword using a combination of TF-IDF score and semantic similarity.

4. Data Storage and Analysis:

o Scraped data and analysis results are stored in MongoDB for efficient access and future analysis.

o Additional analytics can be performed on the stored data to identify trends, audience demographics, and other insights.

5. Frontend:

o Gradio provides a user-friendly interface for interacting with TrendCatcher.

o Users can input the "Young Innovator" keyword and apply filters based on platform, region, and country.

o The interface displays the most relevant trending topics and provides access to detailed information about each topic.

4. Implementation Details

• Selenium: Specific libraries like Selenium-Twitter or Selenium-Pinterest can be used for platform-specific automation.

• Beautiful Soup: Parsing techniques should be tailored to each platform's HTML structure.

• BERT Model: Pre-trained models like BERT Base or DistilBERT can be fine-tuned for specific domain terms related to innovation.

- K-means Clustering: Experimentation with different clustering parameters will be needed to optimize topic identification.
- Gradio: The interface can be customized with visualizations and interactive features to enhance user experience.
- MongoDB: Schema design should allow efficient storage and retrieval of different types of data.

4. Conclusion

TrendCatcher has the potential to become a valuable tool for social media marketers and anyone interested in staying up-to-date on the latest innovations. By combining automation, machine learning, and user-friendly interface, TrendCatcher offers a powerful solution for identifying engaging social media content opportunities within the world of innovation.