**TYPE THE TITLE OF YOUR PROJECT HERE**

STUDENT NAMES and ID Numbers:

Olimjonov Diyorjon 12225177

* **Title of the Project**

**Customizable Word Cloud Visualization of Google Search Trends in Central Asian Countries**

* **Introduction and Project Overview**

In today’s digital age, understanding public interest and trends through search behavior is crucial. The goal of this project is to visualize the most frequently searched terms from Central Asian countries, including Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan, using a customizable word cloud. By analyzing these search trends, we can gain valuable insights into the interests and concerns of the population.

The word cloud will display words according to their frequency. Central Asian populations are diverse in the languages they use, including Uzbek, Kazakh, Turkmen, Kyrgyz, Tajik, Uyghur, Russian, and other less commonly used languages. All words from Google Trends will be translated into English to standardize the analysis.

* **Background, Significance, and Objectives of the Research**

The internet has become a major medium for people to seek information, and understanding search behavior is critical for identifying societal trends, preferences, and needs. In Central Asia, where multiple languages and scripts are used, search trends are influenced by diverse linguistic, cultural, and socio-economic factors. This project aims to provide a clear picture of the most commonly searched terms in the region by translating multilingual search queries into English and visualizing them using a word cloud.

* **Significance:**

- Cultural Insight: Understanding what people are searching for in different languages provides insight into the concerns, interests, and popular trends across diverse populations.

- Market Analysis: This research can help businesses, government entities, and organizations better understand the region’s digital behavior, aiding in targeted marketing or public awareness campaigns.

- Linguistic Trends: By analyzing multilingual search terms, this study can reveal which languages dominate online searches and whether language use differs by country.

* **Objectives:**

- Collect Google search data for each of the Central Asian countries.

- Implement a language detection algorithm to identify the language of each search query.

- Use Google Translate to translate non-English search terms into English.

- Create an interactive, customizable word cloud that displays search trends over time (monthly or yearly).

- Provide a comprehensive understanding of public interests and digital trends in Central Asia.

* **Analytical Method**

1. Data Collection

- Tool: Use the `pytrends` Python library to scrape Google Trends search data for selected keywords across Central Asian countries.

- Focus: Collect data on the most popular search terms across different time periods (monthly, yearly).

- Filtering: Filter out irrelevant or redundant search terms to focus on key trends.

2. Language Detection and Translation

- Tool: Use the `langdetect` library to detect the language of each search term.

- Translation: Use the Google Translate API to translate all non-English terms into English. This ensures consistency and enables cross-country comparison.

3. Data Visualization

- Tool: Use the `WordCloud` Python library to create word clouds where the size of each word reflects its frequency in the dataset.

- Customization: Allow users to select specific time periods for viewing trends.

* **Data**

- Source: Google Trends data for search queries in Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan.

- Variables: Language, country, frequency of search terms, time period (month, year).

- Data Range: Data will cover the most recent years available, with the possibility to extend into historical trends if the data is accessible.

- Preprocessing:

- Convert search queries into lowercase.

- Remove duplicate or redundant terms.

- Detect and translate non-English terms into English.

- Remove stop words or common filler terms that don't provide significant insight (e.g., “the,” “and”).

* **Data Preprocessing**

Data preprocessing will involve the following steps:

- Lowercasing: Convert all text to lowercase for uniformity.

- Duplicate Removal: Remove duplicate search terms.

- Stop Words: Eliminate commonly used words with little analytical value.

- Language Detection: Use `langdetect` to identify the language of each query.

- Translation: Translate all non-English terms using Google Translate API.

* **Project Constraints**

Several constraints may impact the project:

1. Data Availability: Historical data might not be fully accessible for all countries or time periods. The project will focus on available and recent data but may have limitations in the depth of historical analysis.

2. Translation Accuracy: Automated translation may lead to inaccuracies or loss of meaning in some terms. Manual review might be necessary for critical terms.

3. Handling of Multilingual Data: Complexities in handling diverse scripts and language structures, especially with languages like Uzbek and Kazakh written in both Latin and Cyrillic scripts.

* Project Plan



* **Expected Results of This Research**

- An interactive word cloud that displays the most popular search terms from Central Asian countries, customized by time period (monthly or yearly).

- Insights into regional digital trends and public interests, across different languages.

- Technical skills I will learn and improve, such as working with plotly and dash API, python libraries for data analysis.

* **References**

1. GeeksforGeeks - Word Cloud in Python:  
   GeeksforGeeks. (n.d.). Generating word cloud in Python. GeeksforGeeks. <https://www.geeksforgeeks.org/generating-word-cloud-python/>
2. Google Translate API:  
   Google. (n.d.). Google Cloud Translation API. Google Cloud. <https://cloud.google.com/translate>
3. pytrends:  
   Salvatier, J., & McKinney, W. (n.d.). pytrends: Unofficial API for Google Trends. GitHub. <https://github.com/GeneralMills/pytrends>
4. langdetect:  
   Nakatani, N. (n.d.). langdetect: Language Detection Library in Python. GitHub. <https://github.com/Mimino666/langdetect>
5. WordCloud Library:

Mueller, A. (n.d.). WordCloud for Python. GitHub. <https://github.com/amueller/word_cloud>