Добрий день. Мене звати Федорко Андрій.

[Слайд 6 Демонстрація програми]

Давайте перейдемо до демонстрації програми.

Програма складається з двох модулів.

Перший модуль створює для кожної категорії список характерних для неї слів і дозволяє прослідкувати як це працює на кожному етапі.

Другий модуль

Визначає до якої категорії належить введений текст.

Давайте спробуємо ввести якийсь текст з категорії Економіка.

Перехід на сайт новин і копіювання інформації з блоку Економіка.

Натискаємо Аналізувати і програма в зручному форматі показує до якої категорії ймовірніше всього відноситься текст. Два центральних стовпчики показують к-сть співпадінь і відсоток. Останній параметр був виведений мною і він працює з більшою точністю. По ньому і варто орієнтуватись.

Оскільки я перед цим вже сформував необхідні дані, зараз ми можемо подивитись на частоту слів для будь-якої категорії.

Якщо ми натискаємо не туди то спрацьовує підказка, яка допомагає розібратися в властивостях програми.

Для того щоб переглядати дані, потрібно вибрати категорію і натиснути на відповідну кнопку.

*Показати що відбувається коли натиск на кноп*

Якщо ми натиснемо TF то програма покаже які слова найчастіше зустрічаються в тексті.

Тепер перейдемо до першого модуля і розберемося як з ним працювати.  
Програма інтуїтивно зрозуміла, так як кнопки розташовані в такому порядку в якому їх потрібно натискати. Також спрацьовують підказки якщо користувач робить неправильні дії.

Якщо натиснути Save програма збереже необхідні дані у файл, який можна буде потім легко прочитати.

Щоб змінити основний текст для категорії просто напишіть її назву і введіть оновлений текст і натисніть на ці кнопки щоб перерахувати дані.

Tf визначає наскільки часто слово зустрічається в документі.

Idf визначає наскільки слово є спецефічним

TFIDF це показник який показує наскільки слово є ключовим

Більшу вагу TF-IDF отримають слова з високою частотою появи в межах документа та низькою частотою вживання в інших документах колекції.