CS 6109 – COMPILER DESIGN

MINI PROJECT PROPOSAL

NEWS ARTICLE RECOMMENDATION

MITHUN RAAM M – 2018103562
NAREN T P – 2018103568
SANJAY CHINNI KARTHICK V – 2018103586

NEWS ARTICLE RECOMMENDTATION

PROBLEM STATEMENT:

With roughly a thousand news articles published by journalists at BBC News every day, it is a challenge to bring the right articles to the right readers at the right time. So to provide right articles to the readers, we recommend similar articles to them based on their interests.

OVERVIEW:

A news article recommendation is an application used to provide similar news articles for users based on their history. This is achieved through lexical analysis; it recommends articles on various authors and categories.

ABSTRACT:

Steps involved:

1. Input data
2. Creation of the initial dataset
3. Lexical analysis phase
4. Vectorization
5. Recommendation

A person generally reads a news article, the application will analyse the headline of that article. Before creating any feature from the raw text, we must perform a cleaning process to ensure no distortions are introduced to the model. We have followed these steps:

- **Special character cleaning**
- **Up case/down case**
- **Punctuation signs**
- **Possessive pronouns**
- **Lemmatization**
- **Stop words**

The idea for the above steps is to tokenize the sentence, and proceed further. During the vectorization process, it will compare the given articles with the dataset provided to the system. Eventually, the articles are recommended based on user's history.

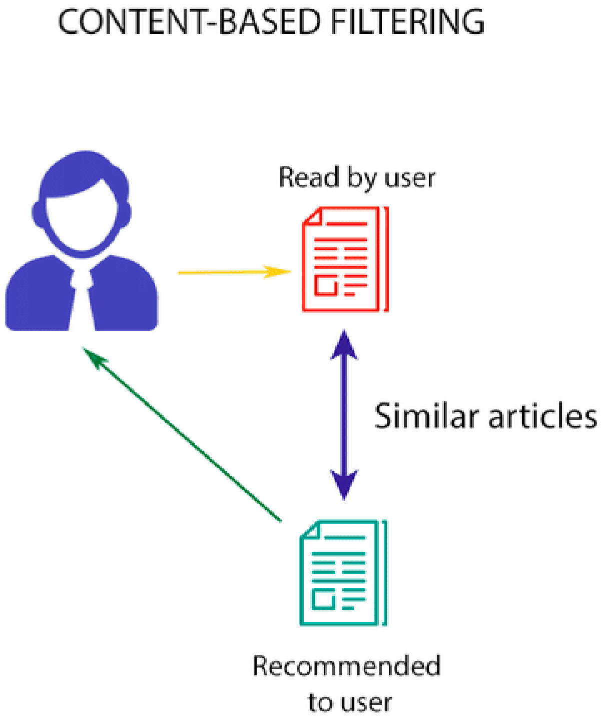
EUCLIDEAN SIMILARITY

We have used Euclidean similarity to compare two lists of numbers (i.e. vectors), and compute a single number which evaluates their similarity. Most measures were developed in the context of comparing pairs of variables across cases. In other words, the objective is to determine to what extent two variables co-vary, which is to say, the same values for the same cases have.

A news article dataset, originating from BBC News has been used.
The dataset has:

- Categories: 26
- Articles: 1, 80,543

APPROACH:



REFERENCE:

Personalized News Recommendation Based on Collaborative Filtering. Florent Garcin, Kai Zhou, Boi Faltings, Vincent Schickel.