THE EFFECTIVENESS OF COMBINING TEXT TOPIC DETERMINATION ALGORITHMS WITH ARTIFICIAL INTELLIGENCE

Vlokh Serhii, Maria Bilova
National Technical University «Kharkiv Polytechnic Institute», Kharkiv

Intelligent Text Analysis Systems are designed to extract meaningful information from large volumes of textual data to support various decision-making processes. One of the fundamental tasks in such systems is topic detection – the process of determining the main themes or subjects discussed in a given text.

A widely used method for this is TF-IDF (Term Frequency–Inverse Document Frequency), which evaluates the importance of words in documents relative to a corpus. While TF-IDF is efficient and interpretable, it has limitations in understanding deeper semantic relations in language.

To enhance the quality of topic detection, the system being developed aims to combine TF-IDF with artificial intelligence techniques, such as supervised machine learning algorithms, which can learn patterns in data and classify texts into predefined topics. This hybrid approach is expected to improve accuracy and adaptability in dynamic content environments.

To begin the process, textual data is collected from various sources and preprocessed through standard NLP techniques, such as tokenization, stop-word removal, and lemmatization. TF-IDF values are then computed to extract key features from each document.

These features are subsequently fed into a machine learning model (e.g., Naive Bayes, SVM, or neural networks), which has been trained on labeled data to associate term patterns with specific topics. The system outputs predicted topic labels, which can be used to personalize recommendations or organize large corpora of content.

To evaluate the effectiveness of this combined approach, an experimental module was developed to allow comparative testing between TF-IDF alone and the TF-IDF + AI integration, using metrics such as precision, recall, and F1-score across several datasets.

The results of this analysis are used to refine the model and assess its suitability for use in real-world intelligent systems.