

## BOOK SELECTION SERVICE BASED ON MACHINE LEARNING METHODS

**Arina S. Petrova** (ITMO University, Saint-Petersburg), **Yuliia A. Burlakova** (ITMO University, Saint-Petersburg), **Daria A. Loza** (ITMO University, Saint-Petersburg)

**Supervisor – Kseniya I. Buraya**  
(ITMO University, Saint-Petersburg)

Nowadays, there is growing user dissatisfaction with the recommendation services, as these services use personal data. Our service offers to focus artificial intelligence on the analysis of the text of books, and not on user actions. The work is dedicated to development of web service with the purpose of merging common filters with machine learning approaches.

**Introduction.** Nowadays, book selection services are quite common. An interested reader can freely use such resources, and choose a book to his or her liking. However, having analyzed the currently available services, our team came to the conclusion that the filters that are used to select books are very similar to each other in all services, and are based on rather primitive filters (selection by genre, author, year of writing).

Our team wants to present the idea of such book selection services from a new angle. We decided to add filters, which have not been used anywhere before, and in the same way to bring more modern and technological algorithms of machine learning into the filtering method.

In addition, we would also like to mention the possibility of working together with a cultural institution - the Mayakovsky library. For us this is first of all a great opportunity to expand the search of possible books for the service and together with the curators of the library choose the most preferred area of literature, for which this service will be most interesting and useful. It is also a great opportunity to conduct surveys among library visitors and readers, to understand the pains and needs of the target audience and, thus, to analyze the further development and direction of our service.

**Main part.** Our project consists of three main parts: backend, frontend and text analysis using natural language processing methods.

The main goal of the backend part is to develop the server of the application, for this part the Node.js Express application framework was chosen. PostgreSQL was chosen as the database, since this version of the database seemed to us more user-friendly for people who do not often use the JSON format.

The goal of the NLP filters part is to create and test the use of three filters (for the first version of the service): frequency analysis of colors in books, frequency analysis of words and collocations related to the weather, and sentiment analysis of the text.

The first two filters that we calculated for the 10 books we have selected for the database are a frequency analysis of the color and weather words in the texts of the books. We decided to use the classical methods and, first of all, began to work with the text and bring each text into a form acceptable for further processing (we cleaned the text from unnecessary characters, did tokenization and lemmatization). Then, based on the processed text, we applied various algorithms to count the number of words and collocations associated with colors and weather.

In the future, we plan to apply more advanced techniques and machine learning algorithms (for example, gpt3) to train the neural network to find even subtle words related to these topics in order to expand the words for which the frequency is calculated.

The third filter is the sentiment analysis of the text. The texts were also analyzed using machine learning algorithms, as well as libraries such as pandas, numpy and sklearn (the linear regression method was chosen). In the future, it is also planned to improve the existing algorithm.

The frontend part consists of prototyping and specifically developing the website. Firstly, we prepared a website prototype. When creating a prototype, the best practices for implementing book and movie selection services were reviewed. At the moment, we managed to locally connect our site to the database, then we plan to design the pages that the user goes to after selecting a specific filter.

**Conclusion.** At the moment, the site is under development, now we are working on options for introducing our site to the information resources of the Mayakovsky Library. In addition, we are preparing the mvp of the product in order to test the hypothesis about the demand for our idea among the library residents. To do this, we plan to prepare an announcement and post it on the library's social networks. In the future, we would like to continue cooperation with the Mayakovsky Library and develop ideas for other projects.