

## Разработка эргономичного приложения-помощника на основе нейронной сети

Д.С. Савенков<sup>1</sup>, С.Ю. Поздняков<sup>1</sup>, А.А. Кузьменко<sup>1</sup>,  
Р.А. Филиппов<sup>1</sup>, Л.Б. Филиппова<sup>1</sup> ✉

<sup>1</sup> Брянский государственный технический университет  
бул. 50 лет Октября, д. 7, г. Брянск 241035, Российская Федерация

✉ e-mail: libv88@mai.ru

### Резюме

**Целью работы** является разработка структурированной модели для обучения нейронной сети на основе электронной энциклопедии.

**Методы.** В данной работе были использованы методы для анализа текстовых статей: 1) TF-IDF – это статистическая мера, используемая для оценки важности термов в документе относительно корпуса документов. Она вычисляет относительную важность термов, учитывая частоту их появления в документе или всем корпусе документов. TF-IDF позволяет выделить ключевые термины, которые характеризуют содержание каждого документа. 2) Кластеризация методом *k*-средних++ – это метод машинного обучения, применяемый для разделения объектов на группы (кластеры) на основе их сходства. Этот метод позволил создать новые категории статей на основе их содержания. 3) Метод *t*-SNE – это метод визуализации сложных многомерных данных в двух или трех измерениях.

**Результаты.** Анализ текстовых статей с использованием методов обработки естественного языка, таких как TF-IDF, и их кластеризация методом *k*-средних++ позволяют создать новые категории статей на основе их содержания. В данной работе был получен результирующий график кластеризации, который показал хорошо разделенные кластеры с высокой точностью.

**Заключение.** Для анализа уникальности триграмм в статьях были использованы значения TF-IDF. Триграммы были классифицированы на различные группы уникальности, что позволяет выявить области, где статьи содержат высокую уникальность триграмм. Комбинации различных групп уникальности могут указывать на разные тематические аспекты или контексты в текстах. Дополнительно, с помощью метода *t*-SNE был получен график кластеризации, который визуально выделяет новые категории статей. Этот график помогает наглядно представить структуру кластеров и их взаимное расположение в двух измерениях. Таким образом, исследование позволяет более глубоко понять и систематизировать содержание статей и выделить связи между ними.

**Ключевые слова:** кластеризация; *k*-средних++; метрика силуэта; обработка естественного языка; метод *t*-SNE.

**Конфликт интересов:** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Для цитирования:** Разработка эргономичного приложения-помощника на основе нейронной сети / Д.С. Савенков, С.Ю. Поздняков, А.А. Кузьменко, Р.А. Филиппов, Л.Б. Филиппова // Известия Юго-Западного государственного университета. 2023; 27(3): 52-65. <https://doi.org/10.21869/2223-1560-2023-27-3-52-65>.

Поступила в редакцию 09.08.2023

Подписана в печать 03.10.2023

Опубликована 28.10.2023

© Савенков Д.С., Поздняков С.Ю., Кузьменко А.А., Филиппов Р.А., Филиппова Л.Б., 2023

## Development of an Ergonomic Assistant Application Based on a Neural Network

Danila S. Savenkov <sup>1</sup>, Semin Yu. Pozdnyakov <sup>1</sup>, Alexander A. Kuzmenko <sup>1</sup>,  
Rodion A. Filippov <sup>1</sup>, Lyudmila B. Filippova <sup>1</sup> ✉

<sup>1</sup> Bryansk State Technical University  
50 Let Oktyabrya Blvd. 7, Bryansk 241035, Russian Federation

✉ e-mail: libv88@mai.ru

### Abstract

**Purpose of research** is to develop a structured model for training a neural network based on an electronic encyclopedia.

**Methods.** In this work, methods were used to analyze text articles: 1) TF-IDF is a statistical measure used to assess the importance of terms in a document relative to the corpus of documents. It calculates the relative importance of terms, taking into account the frequency of their appearance in a document or the entire corpus of documents. TF-IDF allows you to highlight the key terms that characterize the content of each document. 2) Clustering by k-means++ is a machine learning method used to divide objects into groups (clusters) based on their similarity. This method made it possible to create new categories of articles based on their content. 3) The t-SNE method is a method of visualizing complex multidimensional data in two or three dimensions.

**Result.** The analysis of text articles using natural language processing methods, such as TF-IDF, and their clustering by the k-means++ method allow you to create new categories of articles based on their content. In this paper, the resulting clustering graph was obtained, which showed well-separated clusters with high accuracy.

**Conclusion.** TF-IDF values were used to analyze the uniqueness of trigrams in the articles. The trigrams were classified into various uniqueness groups, which allows us to identify areas where articles contain high uniqueness of trigrams. Combinations of different uniqueness groups may indicate different thematic aspects or contexts in the texts. Additionally, using the t-SNE method, a clustering graph was obtained that visually highlights new categories of articles. This graph helps to visualize the structure of clusters and their mutual location in two dimensions. Thus, the study allows a deeper understanding and systematization of the content of articles and highlight the links between them.

**Keywords:** clustering; k-means++; silhouette metrics; natural language processing; t-SNE method.

**Conflict of interest.** The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

**For citation:** Savenkov D. S., Pozdnyakov S. Yu., Kuzmenko A. A., Filippov R. A., Filippova L. B. Development of an Ergonomic Assistant Application Based on a Neural Network. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2023; 27(3): 52-65 (In Russ.). <https://doi.org/10.21869/2223-1560-2023-27-3-52-65>.

Received 09.08.2023

Accepted 03.10.2023

Published 28.10.2023

## Введение

Актуальность разработки структуры модели по электронной энциклопедии выражается в ряде аспектов:

Электронная энциклопедия – один из самых крупных источников данных, содержащий много статей по разным направлениям. Используя эти данные, обогатит модель большим объемом информации, что способствует повышению ее точности и эффективности.

В электронной энциклопедии существует структурированная организация статей, в которой содержатся связи между разными темами и категориями. Эту структуру можно использовать для того, чтобы создать семантическую модель, позволяющую устанавливать взаимосвязи между понятиями и понять их контекст. Эта модель может пригодиться для обучения нейронных сетей, в частности, в области классификации, поисковых систем и синтаксического анализа текста.

Применение электронной энциклопедии для создания структуры модели способствует доступности и понятности результатов обучения нейронных сетей, поскольку электронная энциклопедия написана на языке понятном большинству пользователей, ее результаты можно представить в виде информации, которую люди легко интерпретируют и используют в различных областях, таких как принятие решений, изучение

новых тем и получение новой информации.

Исходя из вышеперечисленных аспектов можно сделать вывод, что создание структурированной модели для нейронных сетей на основе электронной энциклопедии актуально, в связи с тем, что позволяет применять обширную информацию, структурированную связь между терминами и обеспечивать доступность и понятность результатов для пользователей.

В заключении научной работы Ермакова П. Д. «Исследование методов машинного обучения в задаче автоматического определения тональности текстов на естественном языке», говорится, что удаление стоп-слов и стемминг уменьшают точность работы методов машинного обучения, а использование биграмм и триграмм не сказываются на ней. Исходя из этого, использование n-грамм для модели будет предпочтительней использования униграмм [1].

Шокина М. О в научной статье «Применение алгоритма k-means++ для кластеризации последовательностей с неизвестным количеством кластеров» пришла к выводу о том, что использование алгоритма k-средние++ является быстрым и точным методом решения задачи кластеризации, с неизвестным количеством кластеров [2].

В работе Крижановского А. А «Преобразование структуры словарной статьи словаря электронной библиотеки в таб-

лицы и отношения реляционной базы данных» сделан вывод, что статьи электронной энциклопедии имеют жёсткую структуру, что позволяет легко извлекать информацию из этого ресурса [3].

На основе результатов, полученных из статьи Частиковой В.А. «Применение методов кластеризации для повышения точности работы нейронных сетей», можно сделать вывод, что оптимальнее использовать кластеризованную нейронную сеть с использованием алгоритма *k-means++*, потому что: повышается точность работы нейронной сети, а также возможно реализация параллельных вычислений [4].

В своей работе Соболева Е. Д. «Визуализация многомерных наборов данных при помощи алгоритмов снижения пространства признаков *rsa* и *t-sne*», говорит о том, что в зависимости от задачи необходимо выбирать метод снижения размерности данных по условиям задачи. В этой работе будет использован метод *t-sne* [5,6].

На основе рассмотренных научных работ, было принято решение, разбить очищенный дамп электронной энциклопедии на биграммы и триграммы, а также провести кластеризацию по ним с использованием алгоритма *k-средних++* и метода снижения размерности данных *t-SNE*.

Достаточно большое число работ рассматривает проблемы моделирования удобных интерфейсных приложений и у рациональных алгоритмов [7, 10, 11, 12, 15]. Однако разработка интеллектуальных помощников с приме-

нением современных методов эргономического анализа, оценивающих удобство трудовой деятельности операторов, проводится не так часто [8, 13, 14]. В рамках данной статьи нами рассматриваются подходы к разработке дата сета для интеллектуальной системы. Основные направления анализа данных [9, 16, 17, 18, 19, 20], с которыми может в дальнейшем работать система, будут учитываться при проектировании итоговой архитектуры.

## Материалы и методы

На начальном этапе был обработан набор данных по статьям русской электронной энциклопедии, представленный в виде XML файла с разметкой тегов, которые представлены в табл. 1.

Следующим шагом является формирование очищенного набора данных от разметки. В результате получается новый XML файл со структурой, представленной в табл. 2.

Для проведения кластерного анализа текстовой информации необходимо преобразовать тексты в числовые векторы [7]. Для решения поставленной задачи будет использоваться статистическая мера TF-IDF на основе триграмм. Выбор триграмм обусловлен несколькими причинами:

- Контекстуальная информация. Триграммы позволяют учитывать взаимосвязь и контекст слов.
- Расширенный словарь. Использование триграмм позволяет расширять словарь и увеличить количество уникальных комбинаций слов.

- Учет контекстных зависимостей. Использование триграмм при анализе текста позволяет находить контекстные зависимости между словами.
- Устранение неоднозначности. Триграммы помогают устранить неоднозначность некоторых слов, так как контекст может дать ясное понимание их значения.

**Таблица 1.** Структура статьи набора данных русской электронной энциклопедии**Table 1.** The structure of the article of the data set of the Russian electronic encyclopedia

Наименование тега / Tag name	Вложенный тег / Nested tag	Значение / Value
title	title	Содержит заголовок страницы и указывает на название статьи в электронной энциклопедии
ns	ns	Указывает на пространство имён, к которому принадлежит страница
revision	revision	Содержит информацию о последней версии статьи
	id	Содержит уникальный идентификатор ревизии статьи
	parentid	Содержит идентификатор ревизии, являющейся предшественником текущей ревизии
	timestamp	Указывает на временную метку, когда была создана или отредактирована ревизия
contributor	contributor	Содержит информацию о вкладе пользователя в ревизию
	username	Содержит имя пользователя, внесшего вклад
	id	Содержит уникальный идентификатор пользователя
comment		Содержит комментарий, связанный с ревизией статьи
model		Указывает на модель контента, используемую для представления текста статьи
format		Указывает на формат файла, в котором хранится содержимое страницы
text		Содержит текст статьи

**Таблица 2.** Структура статьи очищенного дампа русской электронной энциклопедии**Table 2.** The structure of the article of the cleaned dump of the Russian electronic encyclopedia

Наименование тега / Tag name	Значение / Value
title	Содержит заголовок страницы и указывает на название статьи в электронной энциклопедии
text	Содержит очищенный текст статьи

TF-IDF – это статистическая мера, используемая в области обработки естественного языка для оценки важности термов в документе относительно корпуса документов.

TF (частота термина) отражает, насколько часто терм (слово или фраза) встречается внутри конкретного документа. TF рассчитывается путем подсчета количества вхождений термина в документ и деления этого числа на общее количество термов в документе:

$$tf(t,d) = \frac{n_t}{\sum_k n_k}, \quad (1)$$

где  $n_t$  – число вхождений слова  $t$  в документ;  $\sum_k n_k$  – общее число слов в данном документе.

IDF (обратная частота документа) отражает важность термина в контексте всего корпуса документов [8]. IDF рассчитывается путем деления общего количества документов в корпусе на количество документов, в которых встречается данный терм, а затем применяется логарифмическая функция к полученному результату:

$$idf(t,D) = \log \frac{|D|}{|\{d_i \in D \mid t \in d_i\}|}, \quad (2)$$

где  $|D|$  – число документов в коллекции,  $|\{d_i \in D \mid t \in d_i\}|$  – число документов из коллекции  $D$ , в которых встречается  $t$  (когда  $n_t \neq 0$ ).

TF-IDF комбинирует TF и IDF, умножая их значения. В результате получается числовая оценка, которая указывает на важность термина в конкретном документе относительно всего корпуса

$$td-idf(t,D) = tf(t,d) \times idf(t,D). \quad (3)$$

Термы с высоким значением TF-IDF считаются более репрезентативными и информативными для данного документа.

Следующим этапом была кластеризация методом  $k$ -средних++.

Метод  $k$ -средних++ — это модификация метода  $k$ -средних, он применяется для эффективного и стабильного инициирования начальных кластерных центров. Он помогает избавиться от проблемы неудачной инициализации, которая приводит к сходимости с нежелательным локальным оптимумом [9].

Процесс инициализации в методе  $k$ -средних++ отличается от стандартного метода  $k$ -средних. Рассмотрим алгоритм метода:

1. Первый центроид случайно выбирается из всех точек данных.
2. Для каждой точки данных квадратное евклидово расстояние вычисляется до ближайшего выбранного центроида.
3. Выбирается новый центроид с вероятностью, равной квадрату расстояния до ближайшего центроида.
4. Шаги 2 и 3 повторяются, пока не найдено ранее заданное количество центроидов.

После инициализации начальных кластерных центров метод  $k$ -средних++ продолжает работать как обычный метод  $k$ -средних, в котором точки данных приравниваются к ближайшим кластерам, а потом центроиды пересчитываются, процесс повторяется до тех пор, пока не будет достигнута сходимость.

В проекте кластеризация нужна для того, чтобы построить новые категории статей электронной энциклопедии, это позволит выявить ключевые особенности статей для лучшего понимания тематики электронной энциклопедии. Также это даёт возможность организовать индексированный поиск для быстрого доступа к информации о статьях в категориях.

Для более наглядного отображения использовался метод t-SNE.

Цель t-SNE – визуализация сложных многомерных данных в двух или трех измерениях, сохраняя локальные и глобальные сходства между объектами. Данный метод является нелинейным методом снижения размерности и широко используется для обнаружения скрытых структур и кластеров в данных.

Математическое описание t-SNE – имеется набор исходных точек  $X = \{x_1, x_2, \dots, x_n\}$ , где каждая точка  $x_i$  принадлежит пространству высокой размерности.

Вычисление вероятности схожести  $p_{ij}$  между точками  $x_i$  и  $x_j$  в исходном пространстве

$$p_{ij} = \frac{p_{ji} + p_{ij}}{2n}, \quad (4)$$

где  $p_{ij} = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma_i^2}\right)$ ,  $\sigma_i$  – ширина гауссовского ядра.

Вычисление вероятности схожести  $q_{ij}$  между соответствующими точками  $y_i$  и  $y_j$  в пространстве низкой размерности

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}}. \quad (5)$$

Минимизация дивергенции Кульбака-Лейблера между  $p_{ij}$  и  $q_{ij}$  при помощи градиентного спуска

$$y_i(\text{new}) = y_i(\text{old}) + \eta \sum_{j=1}^n (p_{ij} - q_{ij}) \cdot (1 + \|y_i - y_j\|^2)^{-1} \cdot (y_i - y_j), \quad (6)$$

где  $\eta$  – скорость обучения.

### Результаты и их обсуждение

Результатом кластеризации методом k-средних++ является график, изображённый на рис. 1.

На графике ось Unique\_percent отвечает за процент содержания уникальных триграмм в статье, где значения варьируются от 0-100. Ось Usual\_percent отвечает за процент содержания в статье часто распространённых триграмм в статье, где значения варьируются от 0-100. На данном графике располагается 98 кластеров. Точность кластеризации определена метрикой силуэта и равна 96.6%, что указывает на хорошо разделённые кластеры.

Уникальность триграмм определяется на основе значений TF-IDF, представленных в табл. 3.

Смотря на график, нельзя однозначно выделить кластеры. Для решения данной проблемы был использован метод снижения размерности данных t-SNE, его результат представлен на рис. 2.

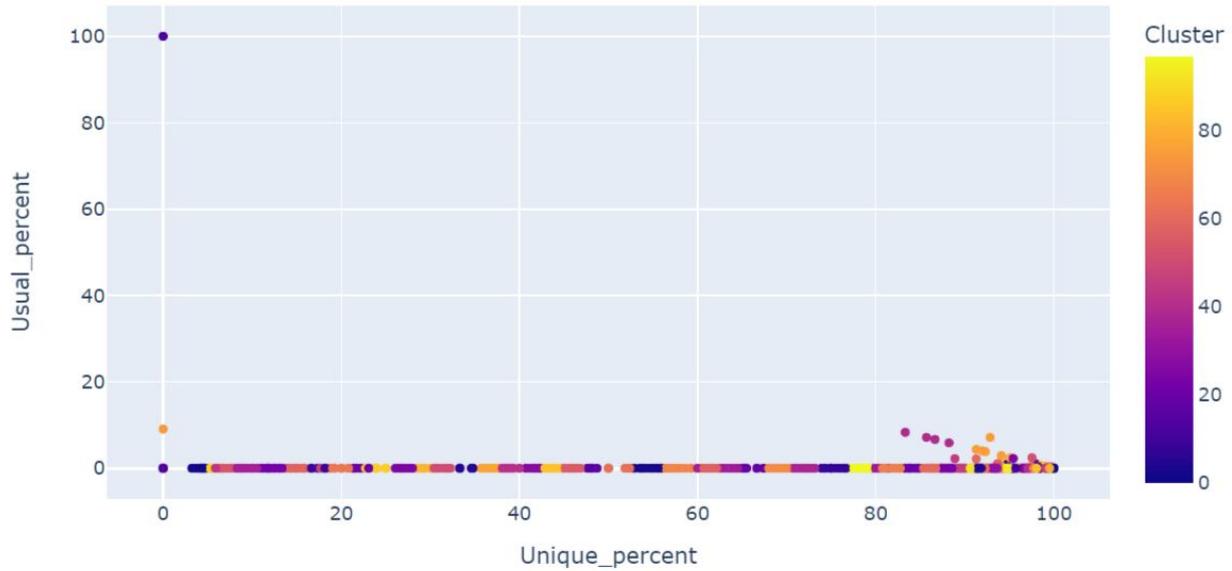


Рис. 1. Результирующий график кластеризации

Fig. 1. The resulting clustering graph

Таблица 3. Группы частотности

Table 3. Frequency groups

Наименование / Name	Расшифровка / Decryption	Диапазон / Range
unique	Уникальные триграммы	от 0.8 до 1
less unique	Менее уникальные триграммы	от 0.6 до 0.8
usually unique	Часто уникальные триграммы	от 0.4 до 0.6
less usual	Менее часто распространённые триграммы	от 0.2 до 0.4
usual	Часто распространённые триграммы	от 0 до 0.2

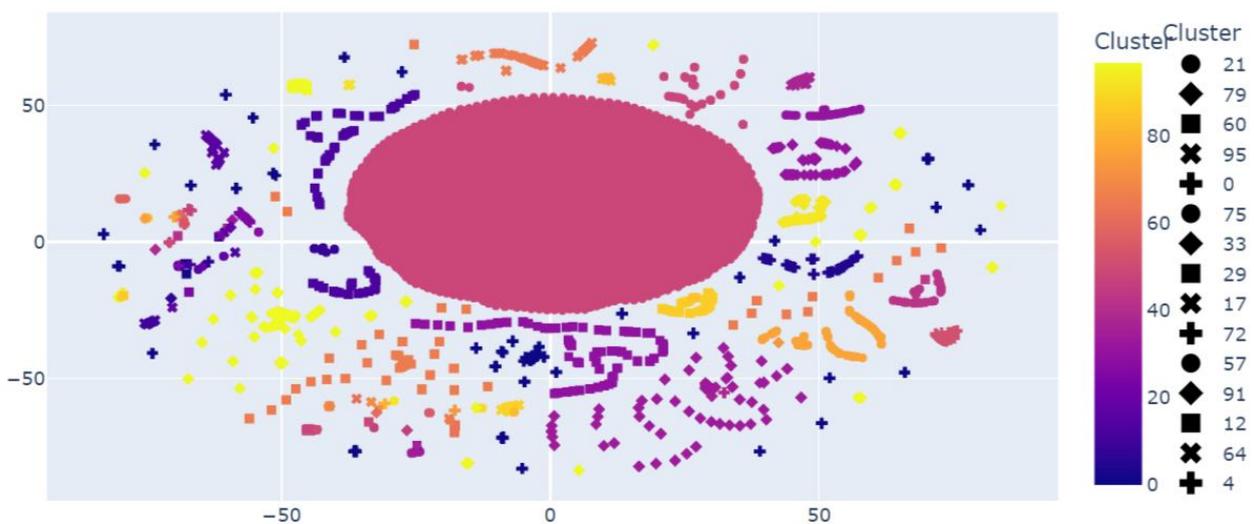


Рис. 2. Результирующий график кластеризации с применением t-SNE

Fig. 2. The resulting clustering graph using t-SNEP

Из этого графика выделяется большая область розовых кластеров, они обозначают статьи, у которых сто процентов уникальных триграмм. Фиолетовым цветом обозначаются статьи, у которых большая часть содержания статьи составляют уникальные триграммы, меньшую часть – менее уникальные триграммы. Оранжевым цветом выделены статьи, в которых большая часть – уникальные триграммы, остальное – менее уникальные триграммы и обычно уникальные триграммы. Остальные кластера обозначают сочетания содержания статей уникальности триграмм, где всё равно большую часть составляют уникальные триграммы в сочетании с другими группами уникальности триграмм.

### Выводы

В данной работе был проведен анализ текстовых статей с использованием методов обработки естественного языка. Применялась статистическая мера TF-IDF для оценки важности термов в документе относительно корпуса документов. Далее была выполнена кла-

стеризация методом k-средних++ для создания новых категорий статей на основе их содержания. Результаты кластеризации были визуализированы с помощью метода t-SNE, позволяющего отобразить сложные многомерные данные в двух измерениях.

Результирующий график кластеризации показал хорошо разделенные кластеры с высокой точностью, оцененной метрикой силуэта в 96.6%. Уникальность триграмм была определена на основе значений TF-IDF, и триграммы были классифицированы на различные группы уникальности.

Дополнительно, с применением метода t-SNE был получен график кластеризации, который помог визуально выделить новые категории статей. На графике были выявлены области, где статьи содержат высокую уникальность триграмм, а также комбинации различных групп уникальности.

В ходе данной работы была подготовлена модель для построения эргономичной системы помощника на основе данных открытой библиотеки электронной энциклопедии.

### Список литературы

1. Ермаков П. Д., Федянин Р.В. Исследование методов машинного обучения в задаче автоматического определения тональности текстов на естественном языке // Новые информационные технологии в автоматизированных системах. 2015. №. 18. С. 600-615.
2. Шокина М. О. Применение алгоритма k-means++ для кластеризации последовательностей с неизвестным количеством кластеров // Новые информационные технологии в автоматизированных системах. 2017. №. 20. С. 160-163.

3. Крижановский А. А. Преобразование структуры словарной статьи Викисловаря в таблицы и отношения реляционной базы данных. URL: <http://scipeople.com/publication/1000231>.

4. Частикова В.А., Остапов Д.С. Применение методов кластеризации для повышения точности работы нейронных сетей // Современные проблемы науки и образования. 2015. № 1-1.

5. Соболева Е. Д., Попова И. А., Попова А. А. Визуализация многомерных наборов данных при помощи алгоритмов снижения пространства признаков рса и t-sne // StudNet. 2020. Т. 3. №. 11. С. 982-1004.

6. Прогнозирование живучести мультисервисной корпоративной сети связи / Ф.Ю. Лозбинев, А.С. Сазонова, А.А. Тищенко, Ю.А. Леонов // Вестник Брянского государственного технического университета. 2017. № 4 (57). С. 144-150.

7. Методология и технология проектирования информационных систем / Ю.М. Казаков, А.А. Тищенко, А.А. Кузьменко, Ю.А. Леонов, Е.А. Леонов. М.: Общество с ограниченной ответственностью ФЛИНТА, 2018. 136 с.

8. Лагерев А.В., Сазонова А.С., Филиппов Р.А. Модель оценки социально-демографического потенциала и его влияние на структуру высшего профессионального и послевузовского образования в регионе // Информационные системы и технологии. 2012. № 3 (71). С. 72-77.

9. Поиск оптимальных технологических процессов с использованием алгоритмов эвристического поиска / Ю.А. Леонов, Е.А. Леонов, А.С. Зуева, А.С. Сазонова // Вестник Брянского государственного технического университета. 2017. № 4 (57). С. 122-127.

10. Модель размещения данных во внутренней памяти вычислителя, реализующего схему кодирования данных в режиме сцепления блоков / М.О. Таныгин, А.А. Ахмад, О.В. Казакова, Д.А. Голубов // Известия Юго-Западного государственного университета. 2023; 27(1): 73-91. <https://doi.org/10.21869/2223-1560-2023-27-1-73-91>.

11. Экономическая безопасность в региональных социально-экономических системах / Е.Э. Аверченкова, А.В. Аверченков, Ю.А. Леонов, Д.В. Кравцов, Л.Б. Филиппова, Е.А. Леонов. М.: Общество с ограниченной ответственностью "ФЛИНТА", 2019. 157 с.

12. Разработка математической модели информационной системы для инвентаризации и мониторинга программного и аппаратного обеспечения на основе методов нечеткой логики / Р.А. Филиппов, Р.А. Филиппов, Л.Б. Филиппова, А.В. Аверченков, А.С. Сазонова, С.А. Шептунов // Качество. Инновации. Образование. 2018. № 7 (158). С. 105-112.

13. Аверченков В.И., Спасенников В.В., Филиппов Р.А. Исследование точности позиционирования объектов при оптической микроскопии с управлением через Интернет // Вестник Брянского государственного технического университета. 2012. № 1 (33). С. 125-130.

14. Сазонова А.С., Филиппова Л.Б., Филиппов Р.А. Оценка инновационного потенциала региона // Вестник Воронежского государственного университета инженерных технологий. 2017. Т. 79. № 2 (72). С. 273-279.

15. Жадаев Д.С., Кузьменко А.А., Спасенников В.В. Особенности нейросетевого анализа уровня подготовки студентов в процессе адаптивного тестирования их профессиональных компетенций // Вестник Брянского государственного технического университета. 2019. № 2 (75). С. 90-98

16. Кондратенко С.В., Кузьменко А.А., Спасенников В.В. Методология оценки деятельности операторов в человеко-машинных системах // Вестник Брянского государственного технического университета. 2017. № 1 (54). С. 261-270.

17. KMeansTrainer Класс. URL: <https://docs.microsoft.com/ru-ru/dotnet/api/microsoft.ml.trainers.kmeanstrainer?view=ml-dotnet>, свободный.

18. Реализация агентно-базированных метакомпьютерных систем и приложений / Н.С. Карамышева, Д.С. Свищев, К.В. Попов, С.А. Зинкин // Известия Юго-Западного государственного университета. 2022; 26(1): 148-171. <https://doi.org/10.21869/2223-1560-2022-26-1-148-171>.

19. Милостная Н. А. Исследование устойчивости нейро-нечёткой системы вывода, основанной на методе отношения площадей // Известия Юго-Западного государственного университета. 2021; 25(3): 70-85. <https://doi.org/10.21869/2223-1560-2021-25-3-70-85>.

20. Мартышкин А. И., Кирюткин И. А., Мереняшева Е. А. Автотестирование встраиваемой реконфигурируемой вычислительной системы // Известия Юго-Западного государственного университета. 2023; 27(1): 140-152. <https://doi.org/10.21869/2223-1560-2023-27-1-140-152>.

## References

1. Ermakov P. D. Issledovanie metodov mashinnogo obucheniya v zadache avtomaticheskogo opredeleniya tonal'nosti tekstov na estestvennom yazyke [Research of machine learning methods in the problem of automatic determination of the tonality of texts in natural language]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh = New Information Technologies in Automated Systems*, 2015, no. 18, pp. 600-615.

2. Shokina M. O. Primenenie algoritma k-means++ dlya klasterizatsii posledovatel'nostei s neizvestnym kolichestvom klasterov [Application of the k-means++ algorithm for clustering sequences with an unknown number of clusters]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh = New Information Technologies in Automated Systems*, 2017, no. 2, pp. 160-163.

3. Krizhanovsky A. A. *Preobrazovanie struktury slovarnoi stat'i Vikislovary v tablitsy i otnosheniya relyatsionnoi bazy dannykh* [Transformation of the structure of a Wiktionary

dictionary entry into tables and relational database relationships], Available at: <http://scipeople.com/publication/1000231>.

4. Chastikova V.A., Ostapov D.S. *Primenenie metodov klasterizatsii dlya povysheniya tochnosti raboty neironnykh sete* [Application of clustering methods to improve the accuracy of neural networks]. *Sovremennye problemy nauki i obrazovaniya = Modern Problems of Science and Education*, 2015, no. 1-1.

5. Soboleva E. D., Popova I. A., Popova A. A. *Vizualizatsiya mnogomernykh naborov dannykh pri pomoshchi algoritmov snizheniya prostranstva priznakov pca i t-sne* [Visualization of multidimensional data sets using algorithms for reducing the feature space of pca and t-sne]. *StudNet Publ.*, 2020, vol. 3, no. 11, pp. 982-1004.

6. Lozbinev F.Yu., Sazonova A.S., Tishchenko A.A., Leonov Yu.A. *Prognozirovanie zhivuchesti mul'tiservisnoi korporativnoi seti svyazi* [Forecasting the survivability of a multi-service corporate communication network]. *Vestnik Bryanskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Bryansk State Technical University*, 2017, no. 4 (57), pp. 144-150.

7. Kazakov Yu.M., Tishchenko A.A., Kuzmenko A.A., Leonov Yu.A., Leonov E.A. *Metodologiya i tekhnologiya proektirovaniya informatsionnykh sistem* [Methodology and technology of information systems design]. Moscow, FLINT Publ., 2018, 136 p.

8. Lagerev A.V., Sazonova A.S., Filippov R.A. *Model' otsenki sotsial'no-demograficheskogo potentsiala i ego vliyanie na strukturu vysshego professional'nogo i poslevuzovskogo obrazovaniya v regione* [A model for assessing socio-demographic potential and its impact on the structure of higher professional and postgraduate education in the region]. *Informatsionnye sistemy i tekhnologii = Information Systems and Technologies*, 2012, no. 3 (71), pp. 72-77.

9. Leonov Yu.A., Leonov E.A., Zueva A.S., Sazonova A.S. *Poisk optimal'nykh tekhnologicheskikh protsessov s ispol'zovaniem algoritmov evristicheskogo poiska* [Search for optimal technological processes using heuristic search algorithms]. *Vestnik Bryanskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Bryansk State Technical University*, 2017, no. 4 (57), pp. 122-127.

10. Tanygin M. O., Ahmad A.A., Kazakova O. V., Golubov D. A. *Recursive Algorithm for Forming Structured Sets of Information Blocks that Increase the Speed of Performing Procedures for Determining Their Source*. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2023; 27(1): 73-91 (In Russ.). <https://doi.org/10.21869/2223-1560-2023-27-1-73-91>.

11. Averchenkova E.E., Averchenkov A.V., Leonov YU.A., Kravcov D.V., Filippova L.B., Leonov E.A. *Ekonomicheskaya bezopasnost' v regional'nyh social'no-ekonomicheskikh sistemah* [Economic security in regional socio-economic systems]. Moscow, FLINTA Publ., 2019, 157 p.

12. Filippov R.A., Filippov R.A., Filippova L.B., Averchenkov A.V., Sazonova A.S., Sheptunov S.A. Razrabotka matematicheskoy modeli informacionnoj sistemy dlya inventarizatsii i monitoringa programmnoy i apparatnoy obespecheniya na osnove metodov nechetkoj logiki [Development of a mathematical model of an information system for inventorying and monitoring software and hardware based on fuzzy logic methods]. *Kachestvo. Innovatsii. Obrazovanie = Quality. Innovation. Education*, 2018, no. 7 (158), pp. 105-112.

13. Averchenkov V.I., Spasennikov V.V., Filippov R.A. Issledovanie tochnosti pozicionirovaniya ob"ektov pri opticheskoy mikroskopii s upravleniem cherez Internet [Investigation of the accuracy of object positioning in optical microscopy with control via an Internet]. *Vestnik Bryanskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Bryansk State Technical University*, 2012, no. 1 (33), pp. 125-130.

14. Sazonova A.S., Filippova L.B., Filippov R.A. Ocenka innovacionnogo potentsiala regiona [Assessment of the innovative potential of the region]. *Vestnik Voronezhskogo gosudarstvennogo universiteta inzhenernykh tekhnologiy = Bulletin of the Voronezh State University of Engineering Technologies*, 2017, vol. 79, no. 2 (72), pp. 273-279.

15. ZHadaev D.S., Kuz'menko A.A., Spasennikov V.V. Osobennosti nejrosetevogo analiza urovnya podgotovki studentov v processe adaptivnogo testirovaniya ih professional'nykh kompetentsiy [Features of neural network analysis of the level of training of students in the process of adaptive testing of their professional competencies]. *Vestnik Bryanskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Bryansk State Technical University*, 2019, no.2 (75), pp. 90-98.

16. Kondratenko S.V., Kuz'menko A.A., Spasennikov V.V. Metodologiya ocenki deyatel'nosti operatorov v cheloveko-mashinnykh sistemakh [Methodology for evaluating operator activity in human-machine systems]. *Vestnik Bryanskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Bryansk State Technical University*, 2017, no. 1 (54), pp. 261-270.

17. *KMeansTrainer Klass*. Available at: <https://docs.microsoft.com/ru-ru/dotnet/api/microsoft.ml.trainers.kmeanstrainer?view=ml-dotnet>, svobodnyj.

18. Karamysheva N. S., Svishchev D. S., Popov K. V., Zinkin S. A. Implementation of Agent-Based Metacomputersystems and Applications. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2022; 26(1): 148-171 (In Russ.). <https://doi.org/10.21869/2223-1560-2022-26-1-148-171>

19. Milostnaya N. A. Stability Study of a Neuro-Fuzzy Output System Based on Ratio Area Method. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2021; 25(3): 70-85 (In Russ.). <https://doi.org/10.21869/2223-1560-2021-25-3-70-85>.

20. Martyshkin A. I., Kiryutkin I. A., Merenyasheva E. A. Autotesting an Embedded Reconfigurable Computing System. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2023; 27(1): 140-152 (In Russ.). <https://doi.org/10.21869/2223-1560-2023-27-1-140-152>.

---

### Информация об авторах / Information about the Authors

**Савенков Данила Сергеевич**, студент,  
Брянский государственный  
технический университет,  
г. Брянск, Российская Федерация,  
e-mail: [savenkov-lokin2013@yandex.ru](mailto:savenkov-lokin2013@yandex.ru)

**Danila S. Savenkov**, Student, Bryansk State  
Technical University, Bryansk, Russian Federation,  
e-mail: [savenkov-lokin2013@yandex.ru](mailto:savenkov-lokin2013@yandex.ru)

**Поздняков Семён Юрьевич**, студент,  
Брянский государственный  
технический университет,  
г. Брянск, Российская Федерация,  
e-mail: [santapolusa@gmail.com](mailto:santapolusa@gmail.com)

**Semyon Yu. Pozdnyakov**, Student, Bryansk State  
Technical University, Bryansk, Russian Federation,  
e-mail: [santapolusa@gmail.com](mailto:santapolusa@gmail.com)

**Кузьменко Александр Анатольевич**,  
кандидат биологических наук, доцент  
кафедры «Компьютерные технологии  
и системы», Брянский государственный  
технический университет,  
г. Брянск, Российская Федерация,  
e-mail: [alex-rf-32@yandex.ru](mailto:alex-rf-32@yandex.ru)

**Alexander A. Kuzmenko**, Cand. of Sci.  
(Biological), Associate Professor of Computer  
Technologies and Systems Department,  
Bryansk State Technical University,  
Bryansk, Russian Federation,  
e-mail: [alex-rf-32@yandex.ru](mailto:alex-rf-32@yandex.ru)

**Филиппов Родион Алексеевич**, кандидат  
технических наук, доцент, доцент кафедры  
«Компьютерные технологии и системы»,  
Брянский государственный  
технический университет,  
г. Брянск, Российская Федерация,  
e-mail: [redfil@mail.ru](mailto:redfil@mail.ru)

**Rodion A. Filippov**, Cand. of Sci. (Engineering),  
Associate Professor, Associate Professor of  
Computer Technologies and Systems Department,  
Bryansk State Technical University,  
Bryansk, Russian Federation,  
e-mail: [redfil@mail.ru](mailto:redfil@mail.ru)

**Филиппова Людмила Борисовна**, кандидат  
технических наук, доцент, доцент кафедры  
«Компьютерные технологии и системы»,  
Брянский государственный  
технический университет,  
г. Брянск, Российская Федерация,  
e-mail: [libv88@mail.ru](mailto:libv88@mail.ru)

**Lyudmila B. Filippova**, Cand. of Sci. (Engineering),  
Associate Professor, Associate Professor of  
Computer Technologies and Systems Department,  
Bryansk State Technical University,  
Bryansk, Russian Federation,  
e-mail: [liv88@mail.ru](mailto:liv88@mail.ru)