

УДК 631.1,62-5

5.2.2. Математические, статистические и инструментальные методы экономики (физико-математические науки, экономические науки)

### **РАЗРАБОТКА ВЕБ-ПОРТАЛА С АДАптиРОВАННОЙ ИНФОРМАЦИЕЙ ДЛЯ ПОИСКОВЫХ СИСТЕМ, ИССЛЕДОВАНИЕ И РЕАЛИЗАЦИЯ АЛГОРИТМА ПОДБОРА НОВОСТЕЙ**

Мурлин Алексей Георгиевич  
Кандидат технических наук, доцент кафедры информационных систем и программирования  
РИНЦ-SCIENCE INDEX SPIN-код: 4991-8507  
murlinag@mail.ru  
*ФГБОУ ВО “Кубанский государственный технологический университет”, Краснодар, Россия*

Александров Эдуард Павлович  
студент кафедры информационных систем и программирования  
fess-2009@yandex.ru  
*ФГБОУ ВО “Кубанский государственный технологический университет”, 350072, улица Московская, 2, Краснодар, Россия*

Цель исследования заключается в разработке веб-портала с адаптированной информацией для поисковых систем, а также в создании алгоритма подбора новостей, который будет формировать рекомендации для пользователей. В рамках портала будут реализованы следующие ключевые функции: предоставление новостного контента для пользователей на основе анализа их предпочтений и истории просмотров, отображение статистики взаимодействия пользователей с контентом. Важным элементом системы станет алгоритм подбора новостей, который с помощью методов машинного обучения будет адаптироваться к индивидуальным интересам пользователей, повышая уровень вовлеченности и время пребывания на портале. Это решение позволит не только оптимизировать процесс поиска и получения информации, но и создаст эффективную систему взаимодействия пользователя с контентом

Ключевые слова: ВЕБ-ПОРТАЛ, РЕКОМЕНДАЦИИ, SEO, АЛГОРИТМЫ, МАШИННОЕ ОБУЧЕНИЕ, ПЕРСОНАЛИЗАЦИЯ, PHP, MYSQL, КРОССПЛАТФОРМЕННОСТЬ, FRONTEND, BACKEND, КЛИЕНТ-СЕРВЕРНОЕ ПРИЛОЖЕНИЕ

<http://dx.doi.org/10.21515/1990-4665-204-040>

UDC 631.1,62-5

5.2.2. Mathematical, statistical and instrumental methods of economics (physical and mathematical sciences, economic sciences)

### **DEVELOPMENT OF A WEB PORTAL WITH ADAPTED INFORMATION FOR SEARCH ENGINES, RESEARCH AND IMPLEMENTATION OF A NEWS SELECTION ALGORITHM**

Murlin Alexey Georgievich  
Candidate of Technical Sciences, Associate Professor of the Department of Information Systems and Programming  
RSCI-SCIENCE INDEX SPIN code: 4991-8507  
murlinag@mail.ru  
*Kuban State Technological University, Krasnodar, Russia*

Aleksandrov Eduard Pavlovich  
student of the Department of Information Systems and Programming  
fess-2009@yandex.ru  
*Kuban State Technological University, 350072, Moskovskaya 2, Krasnodar, Russia*

The aim of the study is to develop a web portal with adapted information for search engines, as well as to create a news selection algorithm that will generate recommendations for users. The portal will implement the following key functions: providing news content to users based on an analysis of their preferences and viewing history, displaying statistics on user interaction with the content. An important element of the system will be the news selection algorithm, which, using machine learning methods, will adapt to the individual interests of users, increasing the level of engagement and time spent on the portal. This solution will not only optimize the process of searching and obtaining information, but will also create an effective system of user interaction with the content

Keywords: WEB PORTAL, RECOMMENDATIONS, SEO, ALGORITHMS, MACHINE LEARNING, PERSONALIZATION, PHP, MYSQL, CROSS-PLATFORM, FRONTEND, BACKEND, CLIENT-SERVER APPLICATION

**Введение.** В современном мире количество информации, доступной

<http://ej.kubagro.ru/2024/10/pdf/40.pdf>

в интернете, продолжает расти с невероятной скоростью, что создаёт новые вызовы для поисковых систем и пользователей. Люди хотят получать наиболее релевантные данные быстро и эффективно, а компании и разработчики, в свою очередь, стремятся сделать свои веб-ресурсы максимально заметными для поисковых систем. Именно здесь ключевую роль играет адаптация информации для SEO (search engine optimization - оптимизация для поисковых систем), которая помогает повысить видимость сайта и привлечь целевую аудиторию.

Важность разработки веб-порталов с эффективной SEO-оптимизацией заключается не только в улучшении индексации сайта, но и в том, что правильная оптимизация позволяет предоставить пользователям именно тот контент, который соответствует их запросам. Сложность задачи заключается в том, что поисковые системы, такие как Google и Яндекс, используют сложные алгоритмы ранжирования, и для достижения высокого рейтинга сайт должен соответствовать множеству критериев, таких как скорость загрузки страниц, адаптивность, использование микроразметки и корректная структура данных. В связи с этим, задача адаптации веб-портала для поисковых систем требует глубокого анализа и комплексного подхода.

Отдельного внимания заслуживает вопрос автоматизации подбора новостей. В эпоху цифровых технологий объём новостного контента настолько велик, что пользователям бывает сложно выбрать информацию, которая действительно интересует их. Для решения этой проблемы разрабатываются специальные алгоритмы, которые анализируют предпочтения пользователей и предлагают новости, соответствующие их интересам.

Такие алгоритмы активно используются крупными новостными платформами и социальными сетями, и их успешная реализация значительно повышает вовлечённость пользователей и удовлетворенность

от использования сервиса. Основная задача состоит в создании системы, которая не только будет оптимизирована для поисковых систем, но и обеспечит высокую релевантность новостного контента для каждого отдельного пользователя.

**Проблема исследования** заключается в отсутствии какой-либо универсальной системы, способной решить все поставленные задачи без создания, обучения и внедрения алгоритмов машинного обучения.

**Целью данного исследования** является повышение эффективности новостных веб-порталов и улучшение пользовательского опыта за счёт применения современных технологий анализа данных и персонализации контента.

**Методы исследования.** Методы исследования в контексте разработки веб-портала с адаптированной информацией для поисковых систем и реализации алгоритма подбора новостей охватывают несколько ключевых аспектов.

В первую очередь, необходимо определить основные технологии и инструменты, которые были использованы на разных этапах разработки проекта. Это включает в себя выбор языков программирования, фреймворков, баз данных и программного обеспечения для серверов, а также инструментов для тестирования и оптимизации. В данной работе был выбран стек технологий, состоящий из PHP для серверной части, JavaScript для клиентской части, а также MariaDB для хранения данных.

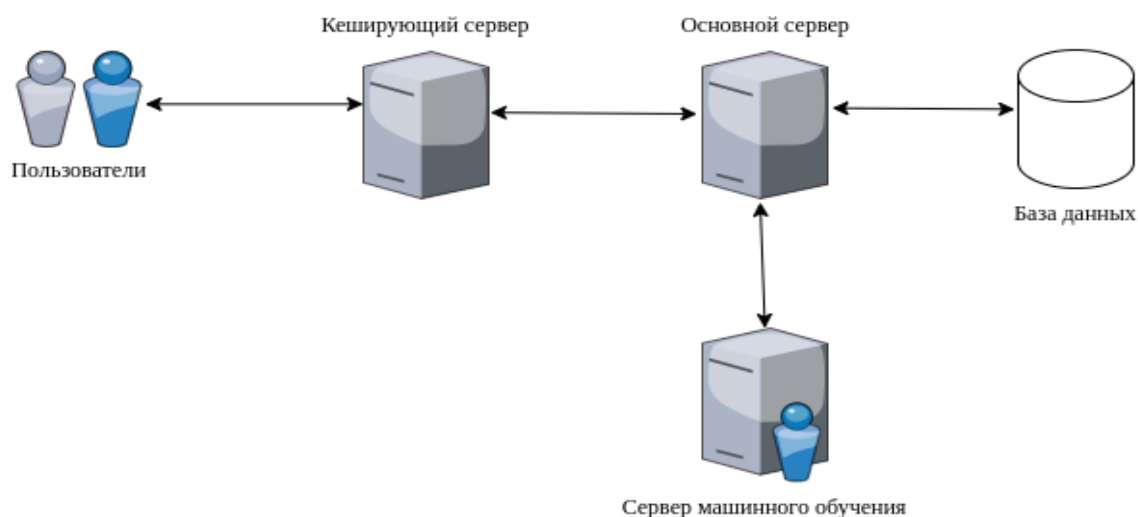


Рисунок 1 – Схема серверной части приложения

Архитектура веб-портала построена на принципах MVC (Model-View-Controller), что позволило разделить логику приложения, представление и управление данными, обеспечивая более лёгкое сопровождение и масштабируемость. В качестве фреймворка для разработки использован Laravel, что обеспечило быструю интеграцию необходимых функциональных возможностей и безопасность приложения. Для клиентской части использовался JavaScript, что обеспечивало интерактивность и удобство работы с интерфейсом и безопасность пользователей на стороне всех поддерживаемых браузеров.

Проектирование базы данных сыграло важную роль в обеспечении функциональности веб-портала. Структура базы данных разработана с учётом потребностей проекта, включая таблицы для хранения новостей, пользователей и тегов. Каждый из компонентов имел чётко определённые поля и связи, что обеспечивало целостность данных и возможность выполнения сложных запросов.

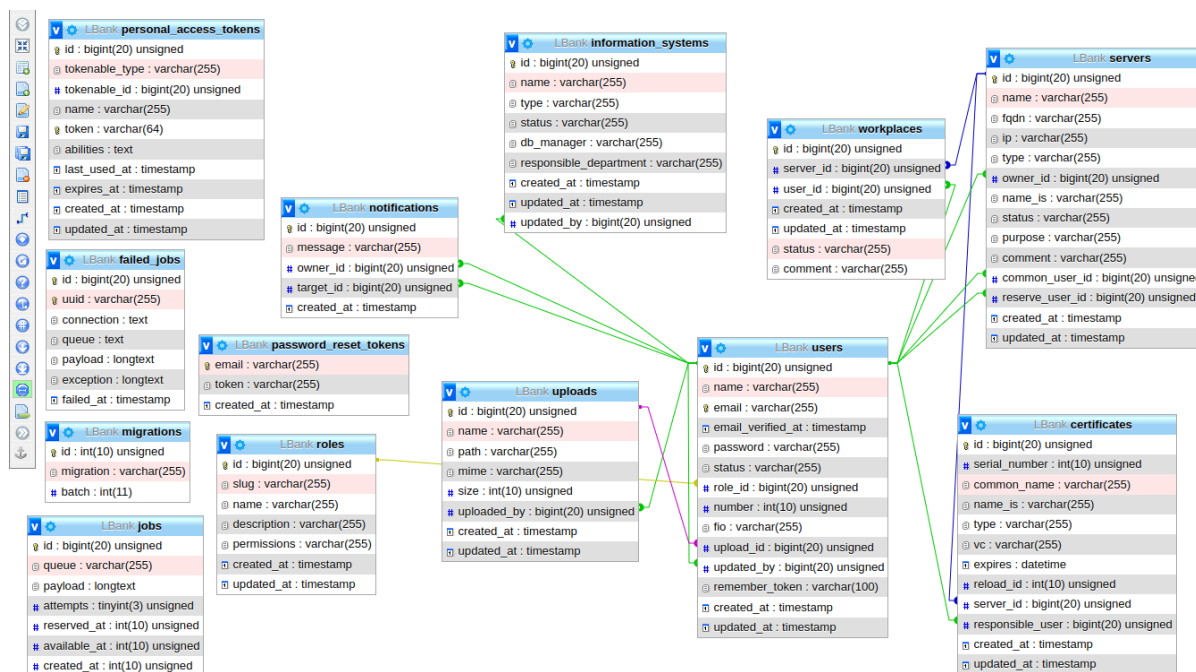


Рисунок 2 – Схема базы данных

Для реализации алгоритма подбора новостей использовались методы машинного обучения и рекомендательные системы. Основная идея заключалась в анализе пользовательских предпочтений и истории взаимодействия с контентом. Были применены алгоритмы, основанные на коллаборативной фильтрации и контентной фильтрации, что позволяло предлагать пользователям наиболее релевантные новости. Алгоритм анализировал данные о том, какие новости пользователи читали ранее, и на основании этого формировал рекомендации.

Тестирование системы проводилось на нескольких этапах. Сначала было выполнено функциональное тестирование, которое охватывало все ключевые функции веб-портала, такие как регистрация пользователей, поиск новостей и подача новостей. В процессе тестирования проверялась корректность работы интерфейса, взаимодействие с базой данных и надёжность алгоритма подбора новостей. Затем проводилось нагрузочное тестирование, направленное на оценку производительности системы под

разными условиями, включая увеличение числа одновременных пользователей и объёмов данных.

Показатели времени отклика системы и производительность алгоритма подбора новостей были оценены с использованием специальных инструментов мониторинга. Эти данные стали основой для дальнейшей оптимизации и настройки параметров системы, чтобы обеспечить высокую скорость работы и релевантность рекомендаций.

Таким образом, методы исследования в данной работе объединяют современные подходы к разработке веб-приложений, применение алгоритмов машинного обучения и тщательное тестирование, что в совокупности позволяет достигнуть поставленных целей по созданию эффективного и удобного веб-портала.

### **Описание концепции работы систем подбора**

Системы автоматического подбора новостей основываются на концепции персонализации контента, что подразумевает использование данных о поведении пользователей для создания индивидуальных рекомендаций. Основная цель такой системы заключается в том, чтобы повысить удовлетворенность пользователей, увеличив вероятность того, что они будут взаимодействовать с представленными новостями. Теория автоматического подбора контента включает в себя несколько ключевых подходов: коллаборативную фильтрацию, контентную фильтрацию и гибридные методы.

Коллаборативная фильтрация предполагает анализ данных о взаимодействиях пользователей с новостями. Например, если несколько пользователей читают одни и те же статьи, система может предположить, что новости, которые понравились одному пользователю, будут интересны и другим. Это позволяет системе рекомендовать новостные статьи на основе схожести между пользователями.

Контентная фильтрация, с другой стороны, основывается на анализе самих новостей. Система может использовать метаданные статей, такие как заголовки, ключевые слова и категории, чтобы понять, какие новости могут быть интересны конкретному пользователю на основе его предыдущих взаимодействий. Таким образом, если пользователь часто читает статьи о политике, система будет рекомендовать аналогичный контент.

Гибридные методы сочетают оба подхода, что позволяет создавать более точные и персонализированные рекомендации. Система может использовать как данные о предпочтениях пользователей, так и характеристики самих новостей, что увеличивает шансы на успешное предсказание интересов пользователей.

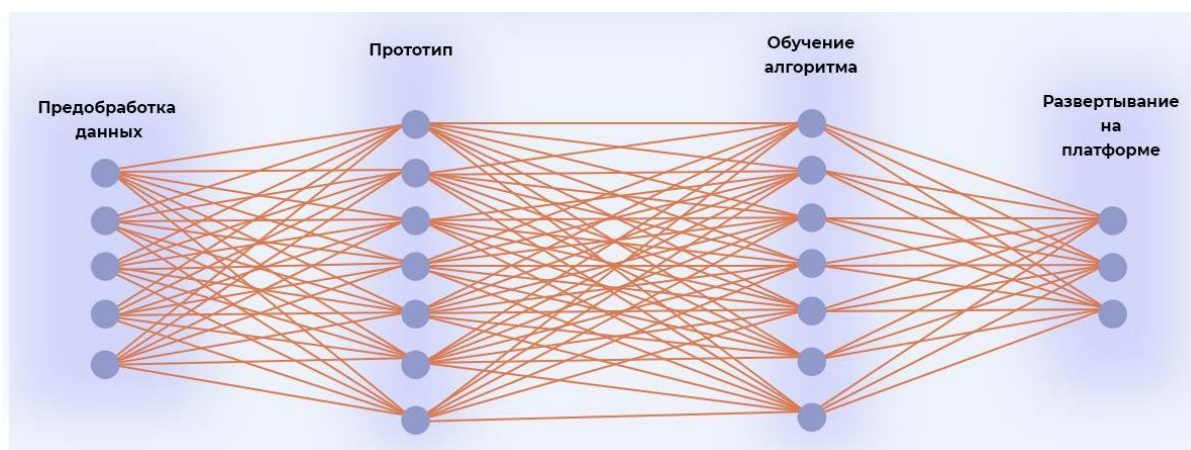


Рисунок 3 – Схема обучения нейронов в ИИ

### Техническое описание

С технической точки зрения, система автоматического подбора новостей включает несколько ключевых компонентов. На первом этапе происходит сбор и хранение данных о пользователях и новостях. Данные могут включать историю просмотров, лайки, комментарии и другие взаимодействия. Для хранения этой информации обычно используется реляционная база данных, такая как MySQL, или NoSQL-системы, которые

обеспечивают большую гибкость в работе с неструктурированными данными.

На втором этапе осуществляется анализ собранных данных. Для этого применяются алгоритмы машинного обучения и статистические методы. Например, алгоритмы, основанные на машинном обучении, могут использоваться для создания моделей, которые предсказывают предпочтения пользователей на основе их истории взаимодействий и профилей. Для реализации этих алгоритмов часто используются библиотеки TensorFlow, PyTorch или scikit-learn, позволяющие эффективно обрабатывать большие объёмы данных и обучать модели.

Третий этап заключается в реализации механизма рекомендаций. Здесь система использует предварительно обученные модели, чтобы генерировать персонализированные списки новостей для каждого пользователя. Это может быть реализовано через API, который получает данные о пользователе и возвращает список рекомендованных статей.

Для улучшения производительности системы и ускорения времени отклика часто применяются кэширование и оптимизация запросов к базе данных. Например, популярные статьи могут кэшироваться, чтобы сократить время обработки запросов. Кеширующим сервером может выступать различное программное обеспечение вроде Memcached или Redis.

Не менее важным аспектом является мониторинг и оценка эффективности работы системы рекомендаций. Это может включать A/B тестирование, где сравниваются результаты разных подходов к подбору контента, а также использование метрик, таких как Click-Through Rate (CTR) и Time On Site (TOS), для анализа того, насколько успешно пользователи взаимодействуют с рекомендованными новостями.



## Используемые алгоритмы

В системах автоматического подбора новостей для пользователей часто используются различные алгоритмы машинного обучения и глубокого обучения. Эти алгоритмы помогают анализировать данные о взаимодействиях пользователей и характеристиках новостей, чтобы делать точные рекомендации.

**Коллаборативная фильтрация** является одним из основных методов, применяемых в рекомендационных системах, — это коллаборативная фильтрация. Она может быть реализована с помощью различных подходов, включая матричную факторизацию. В этой модели пользователи и элементы (в данном случае новости) представляются в виде векторов, и система пытается предсказать рейтинг или предпочтение пользователя к элементу на основе их взаимодействий с другими элементами.

В TensorFlow можно использовать алгоритмы, такие как SVD (сингулярное разложение), для уменьшения размерности данных и выявления скрытых факторов, влияющих на предпочтения пользователей. С помощью этого подхода можно предсказывать рейтинги новостей для пользователей, основываясь на их предыдущих взаимодействиях.



Рисунок 4 – Описание алгоритма коллаборативной фильтрации

**Контентная фильтрация** использует информацию о самом контенте для рекомендаций. Здесь можно применить алгоритмы, основанные на

векторизации текста. Например, TF-IDF (терминная частота — обратная документная частота) может быть использован для представления новостей в виде векторов, где каждое слово имеет вес в зависимости от его частоты в документе и обратной частоты в корпусе.

После векторизации можно использовать такие алгоритмы, как KNN (метод k ближайших соседей) для поиска похожих новостей на основе их векторного представления. В TensorFlow это можно реализовать с помощью нейронных сетей, которые обучаются на характеристиках контента.

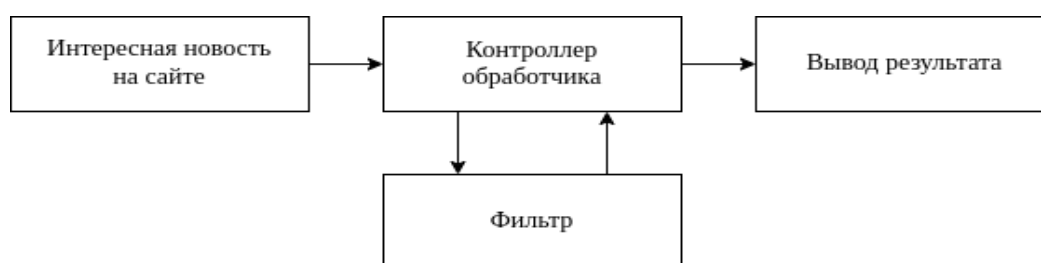


Рисунок 5 – Описание алгоритма контентной фильтрации

**Глубокие нейронные сети** используются для более сложных задач можно использовать глубокие нейронные сети, которые способны извлекать сложные зависимости из данных. Одним из популярных подходов является использование многослойных перцептронов (MLP) или рекуррентных нейронных сетей (RNN) для предсказания предпочтений пользователей.

В этом случае входные данные могут включать векторизованные новости и информацию о пользователе. Сеть обучается на исторических данных, чтобы минимизировать функцию потерь, которая может отражать разницу между предсказанными и фактическими предпочтениями.

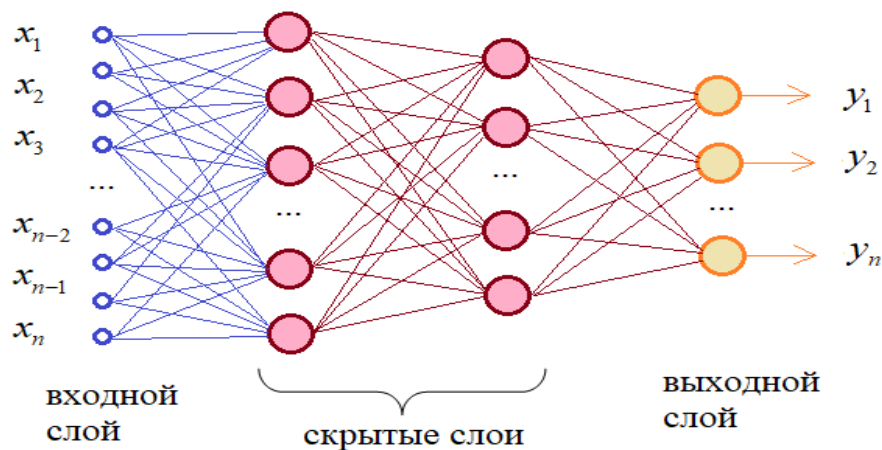


Рисунок 6 – Описание алгоритма нейронных сетей

**Гибридные модели** объединяют методы коллаборативной и контентной фильтрации. Это может быть реализовано с помощью ансамблевого подхода, где разные модели комбинируются для создания более точных рекомендаций. Например, можно использовать предварительно обученную модель коллаборативной фильтрации и дополнить её результатами, полученными с помощью контентной фильтрации.

В TensorFlow это можно реализовать с помощью функционального API, который позволяет строить более сложные модели, состоящие из нескольких компонентов, каждый из которых отвечает за определённый аспект анализа данных.

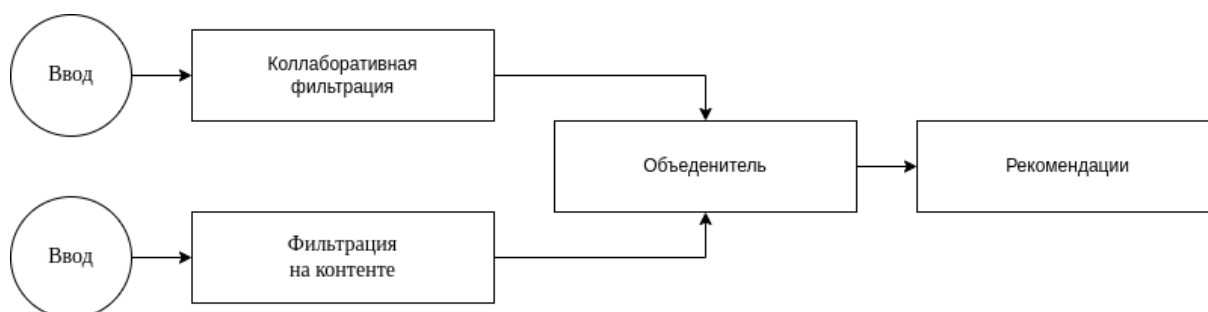


Рисунок 7 – Описание гибридного алгоритма

**Алгоритмы обучения с подкреплением** стали использоваться совсем недавно в системах рекомендаций. Эти алгоритмы учатся на основе взаимодействия с пользователем, оптимизируя свою стратегию рекомендаций в реальном времени. Например, модель может оценивать, как пользователи реагируют на предлагаемые новости, и на основании этого адаптировать свои рекомендации. В TensorFlow для реализации таких моделей можно использовать библиотеки, такие как TF-Agents, которые обеспечивают удобный интерфейс для создания и обучения агентов.



Рисунок 8 – Описание алгоритма обучения с подкреплением

**Выводы.** Исследованы современные подходы к SEO-оптимизации, что позволило внедрить методы, повышающие качество взаимодействия с поисковыми роботами. Это включало в себя создание структурированных данных, микроразметки, а также корректное формирование метатегов, HTML-структуры страниц и файла robots.txt.

Разработан и реализован алгоритм подбора новостей, который основывается на анализе пользовательских предпочтений. Алгоритм учитывает различные параметры, такие как история просмотров, категории

интересующих новостей и общая активность пользователя.

Это позволило разработать веб-портал, способный адаптироваться индивидуально под каждого пользователя, предлагая ему наиболее релевантный контент. Проведённое тестирование показало, что предложенный алгоритм улучшает персонализацию новостной ленты и обеспечивает более точное ранжирование новостей по сравнению с традиционными методами.

### ЛИТЕРАТУРА

1. Фаулер Мартин, Садаладж Прамодкумар Дж. NoSQL. Методология разработки нереляционных баз данных. М.: ИГ «Диалектика—Вильямс», 2020. 292 с.
2. Тарасов С.В. СУБД для программиста. Базы данных изнутри. М.: Солон-Пресс, 2015. 320 с.
3. Остроух А.В., Суркова Н.Е. Проектирование информационных систем. СПб.: Лань, 2019. 364 с.
4. Гарькавый К. А., Бегдай С. Н. Анализ эффективности микроклимата помещения // Энергосбережение и водоподготовка. 2006. No 6(44). С. 76-77.
5. Оцоков Ш. А., Чельшев Э. А., Шибитов Д. В., Раскатова М. В. Разработка программно-аппаратного комплекса сбора и хранения данных термометрии // Инженерный вестник Дона, 2022, No 2. URL: ivdon.ru/ru/magazine/archive/n2y2022/7436.
6. Добробаба, Ю. П. Анализ переходных характеристик системы шестого порядка с кратными корнями характеристического уравнения / Ю. П. Добробаба, А. Г. Мурлин, А. Д. Серкин // Наука. Техника. Технологии (политехнический вестник). – 2019. – № 1. – С. 430-437. – EDN WAMJVI.
7. Добробаба, Ю. П. Анализ переходных характеристик системы второго порядка с кратными корнями характеристического уравнения / Ю. П. Добробаба, А. Г. Мурлин, А. Д. Серкин // Наука. Техника. Технологии (политехнический вестник). – 2019. – № 1. – С. 405-410. – EDN QFPFBX.

### REFERENCES

1. Fauler Martin, Sadaladzh Pramodkumar Dzh. NoSQL. Metodologija razrabotki nereljacionnyh baz dannyh. M.: IG «Dialektika—Vil'jams», 2020. 292 s.
2. Tarasov S.V. SUBD dlja programmista. Bazy dannyh iznutri. M.: Solon-Press, 2015. 320 s.
3. Ostrouh A.V., Surkova N.E. Proektirovanie informacionnyh sistem. SPb.: Lan', 2019. 364 s.
4. Gar'kavyj K. A., Begdaj S. N. Analiz jeffektivnosti mikroklimate pomeshhenija // Jenergoberezenie i vodopodgotovka. 2006. No 6(44). S. 76-77.
5. Ocokov Sh. A., Chelyshev Je. A., Shibitov D. V., Raskatova M. V. Razrabotka programmno-apparatnogo kompleksa sbora i hranenija dannyh termometrii // Inzhenernyj vestnik Dona, 2022, No 2. URL: ivdon.ru/ru/magazine/archive/n2y2022/7436.

6. Dobrobaba, Ju. P. Analiz perehodnyh harakteristik sistemy shestogo porjadka s kratnymi kornjami harakteristicheskogo uravnenija / Ju. P. Dobrobaba, A. G. Murlin, A. D. Serkin // Nauka. Tehnika. Tehnologii (politehnicheskij vestnik). – 2019. – № 1. – S. 430-437. – EDN WAMJVI.

7. Dobrobaba, Ju. P. Analiz perehodnyh harakteristik sistemy vtorogo porjadka s kratnymi kornjami harakteristicheskogo uravnenija / Ju. P. Dobrobaba, A. G. Murlin, A. D. Serkin // Nauka. Tehnika. Tehnologii (politehnicheskij vestnik). – 2019. – № 1. – S. 405-410. – EDN QFPFBX.