

УДК 004.891.2

М. С. Амлин, Е. В. Решетникова

M. S. Amlin, E. V. Reshetnikova

Амлин Максим Станиславович, студент 2 курса магистратуры, ФИМЭ, КГПИ ФГБОУ ВО «КемГУ», г. Новокузнецк, Россия.

Решетникова Елена Васильевна, к. т. н., доцент, КГПИ ФГБОУ ВО «КемГУ», г. Новокузнецк, Россия.

Amlin Maxim Stanuslavovich, 2-year master's student, Kuzbass Humanitarian Padagogiczl Institute of Kemerovo State University, Novokuznetsk, Russia.

Reshetnikova Elena Vasilievla, candidate of technical Sciences, associate Professor, Kuzbass Humanitarian Padagogiczl Institute of Kemerovo State University, Novokuznetsk, Russia.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПОДБОРА ИНФОРМАЦИИ ДЛЯ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ

INTELLIGENT INFORMATION SELECTION SYSTEM FOR RESEARCH ACTIVITIES

Аннотация. *Статья посвящена проблеме поиска научных материалов при изучении какого-либо раздела науки. В работе представлен проект системы, с помощью которой обучающийся сможет искать научные материалы, которые соответствуют его научным интересам.*

Annotation. *The article is devoted to the problem of searching for scientific materials in the study of any aspect of science. The paper presents a project of a system by which the student will be able to search for scientific materials that correspond to his scientific interests.*

Ключевые слова: *интеллектуальная система, научно-исследовательская работа обучающихся, автоматизация научно-исследовательского поиска, нейронная сеть.*

Keywords: *intelligent system, scientific work of students, automation of scientific search, neural network.*

Научно-исследовательская деятельность это наиважнейшая часть в образовании. Обучающемуся, не важно на каком уровне обучения он находится: будь это школьник, студент колледжа или ВУЗа, не всегда легко удается войти в ритм научно-исследовательской деятельности, а именно, перед ним возникают вопросы: с чего начать поиск, где искать, что искать, в каком направлении идти?

Затруднения при поиске информации [1] отнимают много времени у любого человека, тем более у того, кто только начинает заниматься такими исследованиями, не хватает опыта и навыков в поиске информации по интересующим направлениям.

В связи с этим представляется актуальной разработка интеллектуальной системы [2] для автоматизации подбора информации для научно-исследовательской деятельности.

Структура данной системы, представленная на рисунке 1, имеет следующие особенности. Пользователь может перейти на форму выбора научных интересов, после чего запускается интеллектуальный анализ и появляется форма со списком ключевых слов. Пользователю предоставляется возможность редактирования этого списка, после чего ключевые слова заносятся в базу. Далее запускается поисковый алгоритм и пользователю выдается форма со списком найденных материалов.

Также система предоставляет возможность выбирать и добавлять источники, для чего служит соответствующая форма. Для активации этой функции необходимо авторизироваться и после указания ссылки на новый источник, он также будет включен в поиск. По умолчанию в качестве поискового источника используется «Google Scholar».

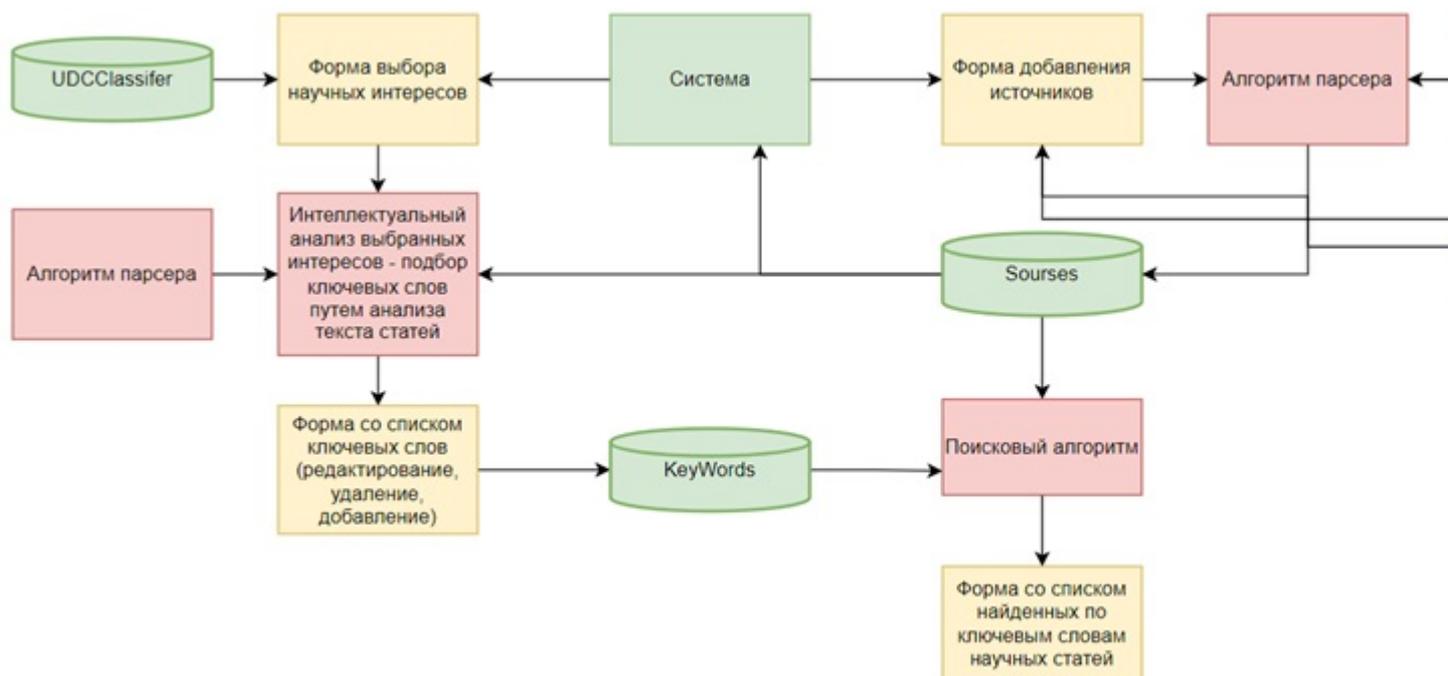


Рисунок 1. Структура интеллектуальной системы

Блок-схема алгоритма интеллектуального подбора ключевых слов для поиска материала [3], представлена на рисунке 2. Согласно этому алгоритму:

- на первом этапе загружаются все выбранные направления из базы данных,
- на втором этапе с помощью парсера [4] производится поиск существующих статей по выбранным направлениям,
- на третьем этапе в каждой из отобранных статей из текста извлекаются ключевые слова по таким триггерам, как:
 1. ключевые слова;
 2. кл. слова;
 3. key words.

Таким образом, по завершении работы данный алгоритм возвращает список с найденными ключевыми словами. Этот список в дальнейшем может быть скорректирован пользователем для сужения области поиска.

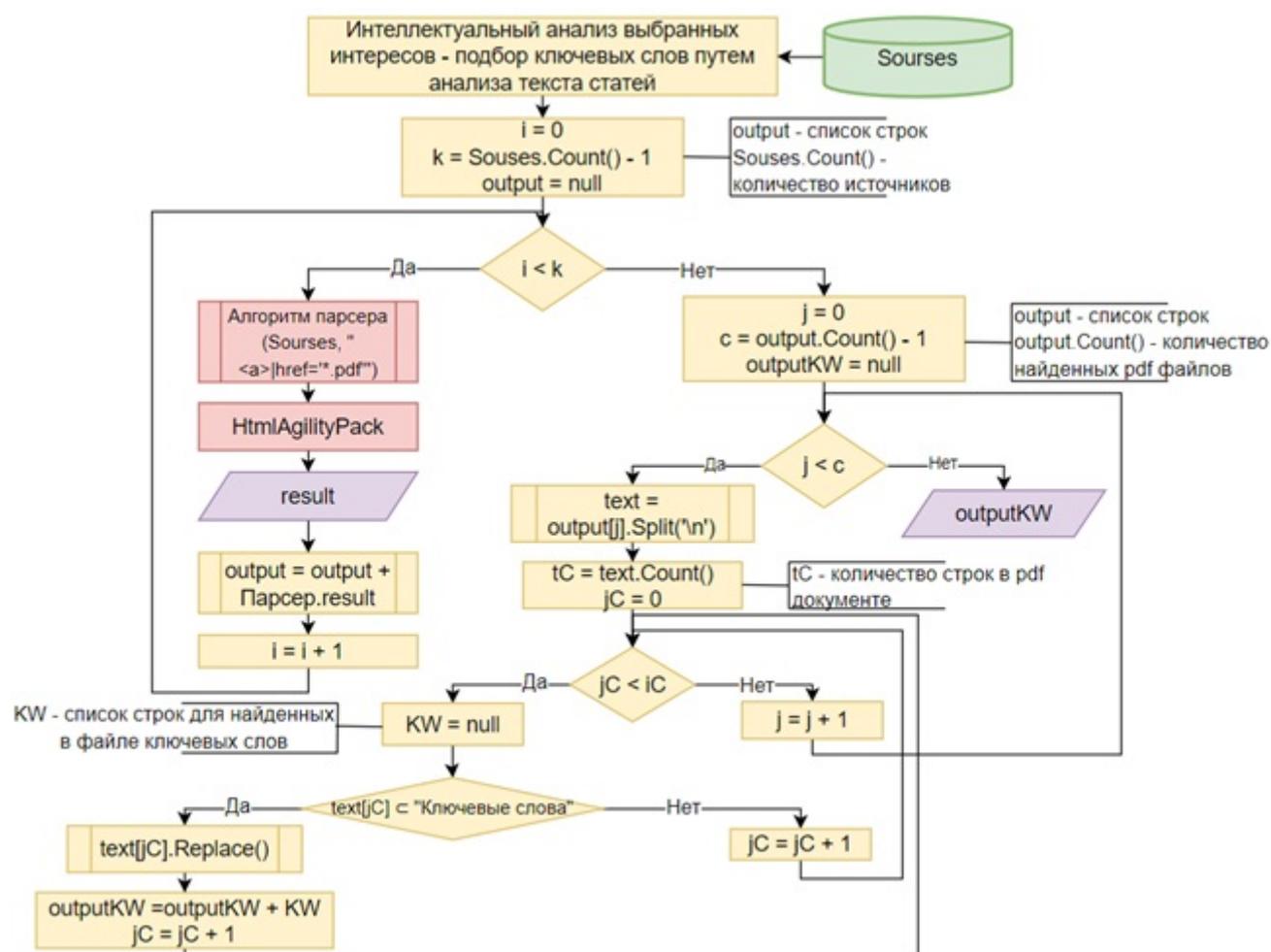


Рисунок 2. Интеллектуальный алгоритм поиска ключевых слов

В качестве поискового алгоритма исследовались три наиболее эффективных алгоритма.

1. **Двухслойная рекуррентная нейронная сеть LSTM.** Принцип работы алгоритма: первый LSTM слой проходит вдоль текста слева направо. Второй LSTM слой проходит в обратном направлении по

тексту справа налево. Выходы обоих LSTM слоёв объединяются и входят в полно связанный слой.

- 2. Метод N-Грамм.** Принцип работы алгоритма: «если слово А совпадает со словом Б с учетом нескольких ошибок, то с большой долей вероятности у них будет общая часть длины N». Во время индексации слово разбивается на N-граммы, а затем это слово попадает в списки для каждой из этих N-грамм. Во время поиска запрос также разбивается на N-граммы, и для каждой из N-грамм выдается список соответствующих ключевых слов.
- 3. Хеширование по сигнатуре.** Принцип работы алгоритма: слово представляется в виде бинарного вектора (хеш-сигнатуры) и записывается в хэш таблице. При индексации хэш вычисляется для каждого слова, и из хэш-таблицы выбираются ключевые слова.

По результатам тестирования быстродействия рассматриваемых алгоритмов, выявилось явное преимущество по времени работы двухслойной рекуррентной нейронной сети LSTM. На рисунке 3 с помощью гистограмм изображено сравнение времени работы над одинаковым потоком входных данных различных алгоритмов.

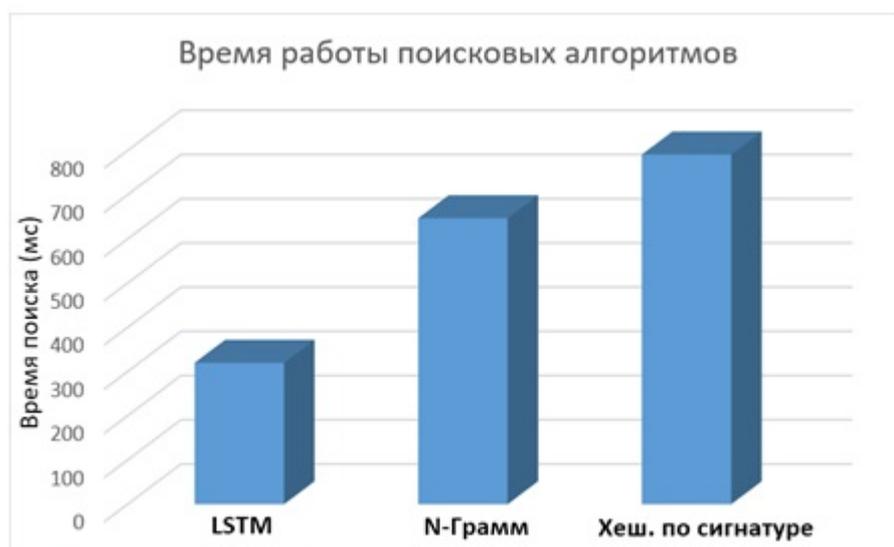
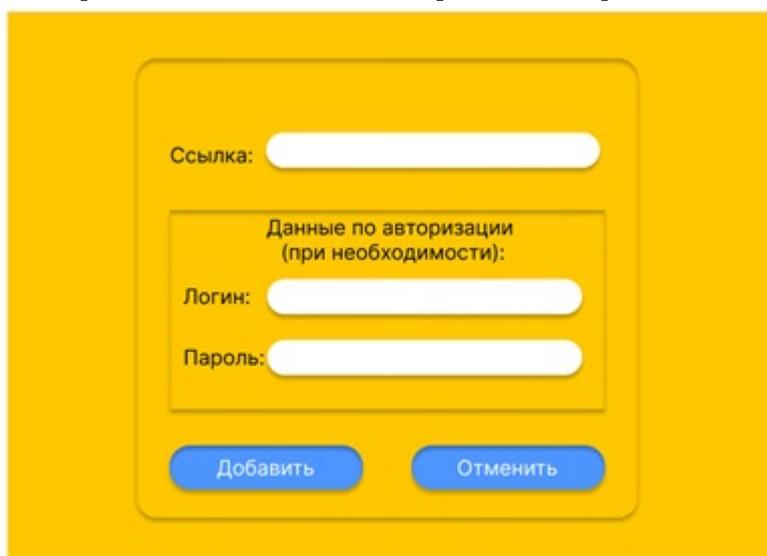


Рисунок 3. Результат быстродействия рассматриваемых алгоритмов

Пользовательский интерфейс системы будет содержать четыре основных формы.

Интерфейс добавления источника представлен на рисунке 4. Данная форма содержит поле для ввода ссылки на источник, а также поля для ввода логина и пароля, если в данном источнике необходима авторизация.



Ссылка:

Данные по авторизации
(при необходимости):

Логин:

Пароль:

Добавить Отменить

Рисунок 4. Форма добавления источника

На рисунке 5 представлена форма, для добавления материала в базу данных. Добавление материала может происходить либо непосредственным вписыванием своего варианта, либо посредством выбора из классификатора УДК.



Составление списка научных интересов:

Свой вариант:

Выбор по УДК:

- Наука в целом
- Философия. Психология
- Религия. Теология
- Теория и методы общественных наук

Выбранные направления:

- Качество систем и программ;
- Архитектура вычислительных машин;
- Программные средства;
- Искусственный интеллект;

Поиск Отменить

Рисунок 5. Форма добавления научных интересов

На рисунке 6 представлена форма, в которой по результату анализа выбранных направлений выводятся найденные ключевые слова. В этом диалоговом окне существует возможность редактирования ключевых слов или добавления дополнительных ключевых слов в специальном поле.

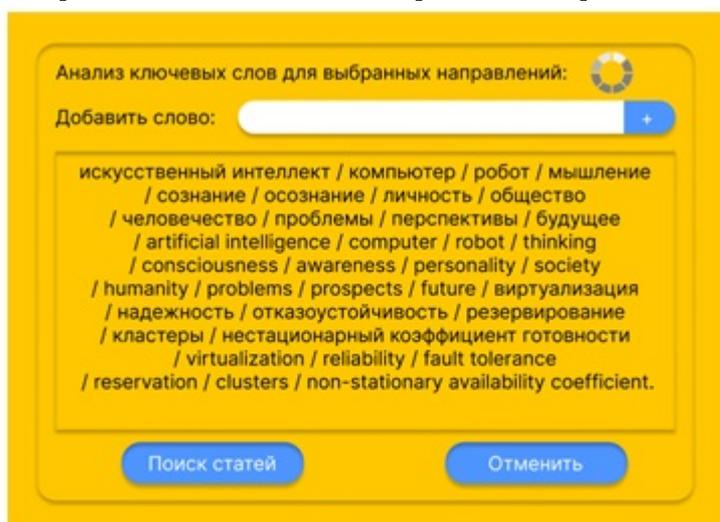


Рисунок 6. Просмотр материала по ключевым словам

На рисунке 7 представлена форма, на которую выводится текст статей, найденных по ключевым словам. Статьи можно читать и сохранять.

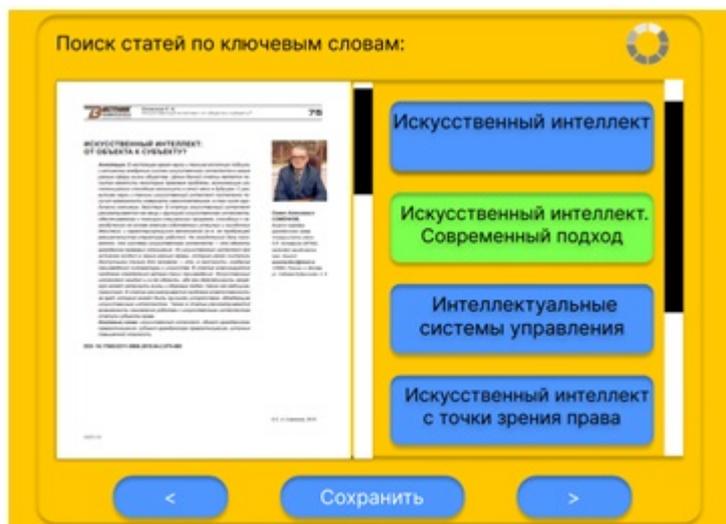


Рисунок 7. Список отображенных материалов

В результате работы спроектирована общая структура функционирования интеллектуальной системы подбора информации для научно-исследовательской деятельности, которая включает в себя: поисковый алгоритм, алгоритм по парсингу страниц, интеллектуальный алгоритм подбора материала по ключевым словам. Разработан алгоритм интеллектуального подбора материала. В качестве поискового алгоритма выбрана двухслойная рекуррентная нейронная сеть LSTM. Спроектирован пользовательский интерфейс системы.

Список литературы

1. Ясницкий, Л. Н. Интеллектуальные системы: учебник [Текст]. / Л. Н. Ясницкий. – М. : Лаборатория знаний, 2016. – Т. 21.
2. Алешева, Л. Н. Интеллектуальные обучающие системы [Текст]. / Л. Н. Алешева. // Вестник университета, 2018. – №. 1.

Материалы I Всероссийской научно-практической конференции «Искусственный интеллект в образовании. Современные достижения и перспективы применения», 2022, №5 (80).

3. Смирнов, И. В. Результаты и перспективы поискового алгоритма Exactus [Текст]. / И. В. Смирнов и др. // Труды российского семинара по оценке методов информационного поиска РОМИП, 2007. – Т. 2008. – С. 66-76.
4. Писарев, И. А. С# парсер для извлечения структуры криптографических протоколов из исходного кода [Текст]. / И. А. Писарев, Л. К. Бабенко. // Труды Института системного программирования РАН, 2019. – Т. 31. – № 3. – С. 191-202.