

Научная статья
УДК 004.021
<https://doi.org/10.24143/2072-9502-2023-4-49-60>
EDN WFBVHV

Поиск технологических возможностей на основе анализа патентного массива

Дмитрий Михайлович Коробкин[✉],
Всеволод Сергеевич Горкин, Сергей Алексеевич Фоменков

*Волгоградский государственный технический университет,
Волгоград, Россия, dkorobkin80@mail.ru*[✉]

Аннотация. В настоящее время компании по-прежнему в значительной степени полагаются на экспертные знания при поиске технологических возможностей и выборе партнеров. Рассматривается автоматизация поиска патентообладателей, которые могут быть партнерами предприятий Волгоградской области. В качестве партнеров рассматриваются компании не только из РФ, но и из Китая, Индии, других дружественных стран. Процесс выявления партнеров строится на основе сходства решаемых технологических проблем, извлеченных из патентных документов, патентообладателями которых являются данные предприятия. Посредством анализа деревьев зависимостей, извлеченных из полей патента «Область применения изобретения» и «Название изобретения», формируются структуры «Проблема-Решение». На основе анализа патентных документов потенциальных партнеров выявляются «Проблемы-Решения», которые не запатентованы в России, чем обусловлен «технологический вакуум» для предприятий РФ. Разработан метод выявления технологических возможностей на основе анализа мирового патентного массива. Новизна метода, обеспечивающего поиск потенциальных вакантных технологических областей для ключевых предприятий Волгоградской области, заключается в использовании технологий глубокого обучения применительно к анализу естественно-языковых данных мирового патентного массива. Разработан программный модуль поиска технологических возможностей для предприятий Волгоградской области на основе схожести решаемых технологических проблем. Модуль программно реализован на языке Python, для семантического анализа текстовых полей патентов использовалась библиотека Yargy из проекта Natasha, для перевода патентных документов на русский язык – Deep-translator, для построения деревьев зависимости – библиотека Stanza, для подбора гиперонимов (для учета синонимичных слов) – библиотека RuWordNet. Для выявления технологических возможностей для предприятий Волгоградской области было проанализировано 6 785 патентов.

Ключевые слова: патент, парсинг, семантический анализ, технологический вакуум, технологические возможности, предприятие, Python

Благодарности: исследование выполнено за счет гранта Российского научного фонда № 23-21-00464, <https://rscf.ru/project/23-21-00464/>.

Для цитирования: Коробкин Д. М., Горкин В. С., Фоменков С. А. Поиск технологических возможностей на основе анализа патентного массива // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2023. № 4. С. 49–60. <https://doi.org/10.24143/2072-9502-2023-4-49-60>. EDN WFBVHV.

Original article

Search for technological opportunities based on the patent array analysis

Dmitry M. Korobkin[✉], *Vsevolod S. Gorkin, Sergey A. Fomenkov*

*Volgograd State Technical University,
Volgograd, Russia, dkorobkin80@mail.ru*[✉]

Abstract. Currently companies still rely heavily on expert knowledge when searching for technological opportunities and choosing partners. Automation of search of patent holders who can be partners of the enterprises of the Volgograd

region is considered. Companies not only from Russia, but also from China, India, and other friendly countries are considered as partners. The process of identifying partners is based on the similarity of solved technological problems extracted from patent documents, the patent holders of which are these enterprises. By analyzing the dependency trees extracted from the patent fields “Scope of the invention” and “Name of the invention” the “Problem-Solution” structures are formed. Based on the analysis of patent documents of potential partners “Problems-Solutions” that are not patented in Russia are identified. It causes a “technological vacuum” for enterprises of the Russian Federation. A method for identifying technological capabilities based on the analysis of the world patent array has been developed. The novelty of the method which provides a search for potential innovative technological areas for key enterprises of the Volgograd region lies in the use of deep learning technologies in relation to the analysis of natural language data of the world patent array. A software module has been developed to search for technological opportunities for enterprises of the Volgograd region based on the similarity of technological problems. The module is programmatically implemented in Python, the Yargy library from the Natasha project was used for semantic analysis of text fields of patents, DeepTranslator was used for translating patent documents into Russian, Stanza library was used for building dependency trees, and RuWordNet library was used for selecting hyperonyms (to account for synonymous words). In order to identify technological opportunities for enterprises of the Volgograd region, 6 785 patents were analyzed.

Keywords: patent, parsing, semantic, technological vacuum, technological capabilities, enterprise, Python

Acknowledgment: the study was supported by the grant of the Russian Science Foundation No. 23-21-00464, <https://rscf.ru/en/project/23-21-00464/>.

For citation: Korobkin D. M., Gorokin V. S., Fomenkov S. A. Search for technological opportunities based on the patent array analysis. *Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics*. 2023;4:49-60. (In Russ.). <https://doi.org/10.24143/2072-9502-2023-4-49-60>. EDN WFBVHV.

Введение

Развитие предприятий Волгоградской области, особенно в условиях санкций, когда ограничен доступ к зарубежным технологическим возможностям и оборудованию, требует самостоятельной разработки инновационных технологий. Внедрение созданных технологий в производственный процесс ключевых предприятий Волгоградской области сможет способствовать созданию ими конкурентноспособной как на российском, так и на мировом рынках продукции, повысить их экономическую привлекательность для инвесторов, расширить рынки сбыта и, соответственно, увеличить налоговые отчисления в бюджет области и создать новые рабочие места.

Раннее выявление технологических возможностей поможет предприятиям Волгоградской области разработать инновационные технологии для получения конкурентных преимуществ в будущем. Выявление новых технологических возможностей – набора перспективных технологических достижений, – согласно большинству исследований, происходит в результате исследования области «технологического вакуума». В настоящее время компании по-прежнему полагаются на экспертные мнения при поиске технологических возможностей для своих производств, поэтому проблема автоматизации процесса поиска технологических возможностей остается актуальной.

Основным инструментом поиска технологических возможностей для предприятий является процесс анализа запатентованных технологий. Патентный анализ является исключительным источ-

ником конкурентной технической информации, его ценность выражается в большой вариативности стратегических применений [1]. Решение таких задач патентного анализа, как оценка технического уровня, изучение тенденций развития объекта техники, требует анализа больших объемов неструктурированных данных, содержащихся в названии, формуле и описании патента [2].

Процесс поиска технологических возможностей без автоматизации является весьма трудозатратным делом. Например, для формирования списка предприятий, решающих схожие технологические задачи, человеку вручную придется:

- получить информацию о каждом патенте, зарегистрированном предприятием;
- анализировать текстовые поля патента, выделив технологическую проблему и ее решение;
- организовать поиск патентов со схожим классом международной патентной классификации (МПК);
- анализировать решаемые проблемы, описанные в патентах предприятий, на предмет сходства;
- формировать список возможных технологических партнеров среди предприятий на основе найденных сходств в решаемых проблемах.

Определение «технологического вакуума» для предприятий РФ посредством поиска патентов, содержащих технологические проблемы, которые не запатентованы в России

Диаграмма существующего процесса (AS-IS) приведена на рис. 1.

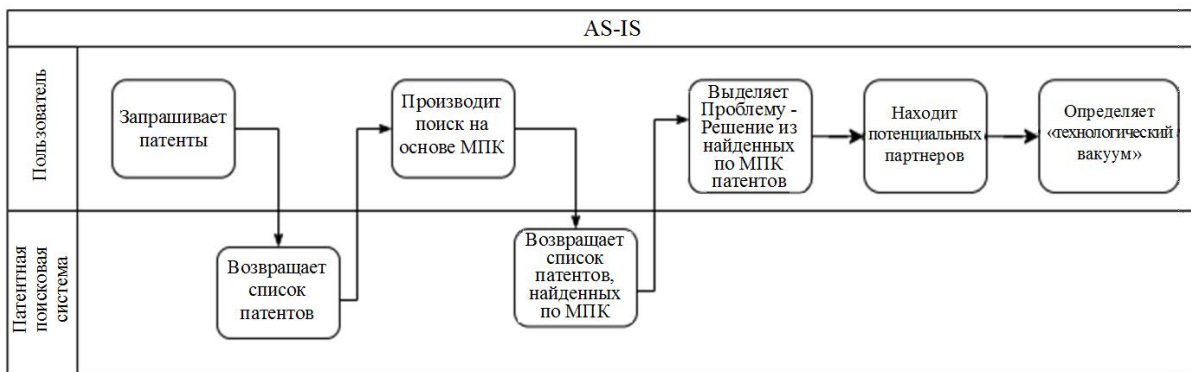


Рис. 1. Диаграмма AS-IS

Fig. 1. AS-IS diagram

Проектируемый процесс поиска технологических возможностей призван автоматизировать все вышеуказанные процессы [3]. Пользователю необходимо лишь проанализировать выходные данные

программы и принять на их основе определенные решения. Диаграмма автоматизированного процесса (TO-BE) представлена на рис. 2.

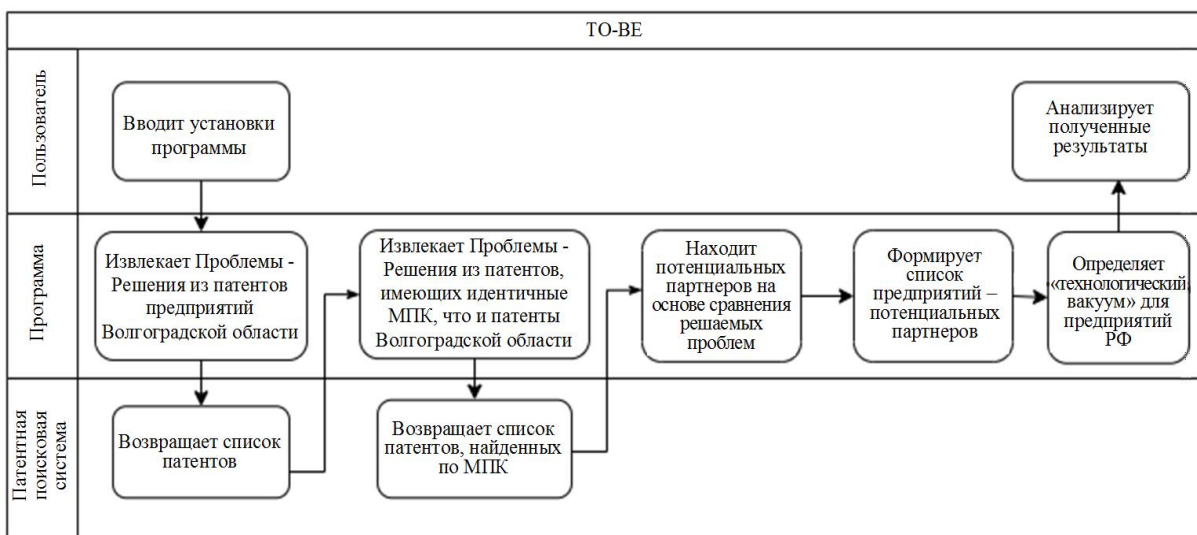


Рис. 2. Диаграмма TO-BE

Fig. 2. TO-BE diagram

Для достижения поставленной задачи автоматизации поиска технологических возможностей необходимо пройти два этапа работы с патентными документами: патентный поиск, а также систематизация и анализ полученных данных.

Для успешного проведения патентного исследования [4] изначально необходимо осуществить патентный поиск. Поиск может быть выполнен вручную или с помощью специальных программных средств в зависимости от требуемой глубины поиска [5]. В нашем случае поиск данных будет выполнен с использованием разработанного парсера и передачи ему необходимых фильтров для поиска нужных патентов.

После проведения патентного поиска необходима обработка полученной выборки данных, чтобы сформировать необходимые семантические структуры. Из текстовых полей патентов необходимо выделить проблему, на решение которой направлено изобретение. Далее анализ патентной коллекции позволяет сделать выводы и предоставить рекомендации, которые могут использоваться в практических задачах, в данном случае для поиска партнеров и выявления технологических возможностей.

Следует предъявить следующие требования:

1. Парсинг патентных документов должен осуществляться из систем Яндекс.Патенты и Google Patents.

2. Формирование списка классов МПК, соответствующих сферам интересов предприятий Волгоградской области.

3. При извлечении структур «Проблема-Решение» из полей патента «Название изобретения» и «Область применения изобретения» задействуются технологии анализа естественно-языковых данных на основе глубокого обучения.

4. Использование англо-русскоязычного перевода терминов, описывающих технологические проблемы и их решения.

5. Определение «технологического вакуума» для предприятий РФ на основе анализа патентных документов потенциальных партнеров посредством поиска структур «Проблемы-Решения», которые не запатентованы в российских патентах.

Материалы и методы

Парсинг патентной информации происходит на основе анализа html документа и последующего поиска тегов, содержащих информацию, необходимую для дальнейшей реализации модуля. Полученные данные заносятся в таблицу БД Clickhouse [6].

Чтобы из большого по объему текста патента извлечь описание решаемой проблемы, необходимо сегментировать его на предложения, а также составить грамматику для библиотеки Yargy [7, 8] на основе разработанных словарей ключевых слов. Всего составляется 4 словаря с различными наборами слов:

– первый словарь – главное слово («цель», «способ», «изобретение», «задача», «модель», «использование»);

– второй словарь – вспомогательное слово, которое с некоторой вероятностью может встретиться при описании проблемы («предлагаемого», «настоящего», «данного», «решаемая»);

– третий словарь – второе слово, составляющее типичную конструкцию описания решаемой патентом проблемы («изобретения», «предназначено», «решения», «устройства»);

– четвертый словарь – другой набор вторых слов в конструкции, используется в случае, если слово из третьего словаря в предложении не найдено («позволяет», «достигается», «результат», «решает», «направлено», «модели», «способа»). Поиск искомого предложения – цели изобретения – происходит в два этапа, для этого составляется два правила на основе вышеописанных словарей: первое правило состоит из первого, второго и третьего словаря, второе – из первого, второго и четвертого. Если в тексте отсутствует предложение, содержащее конструкцию слов по первому правилу, то поиск происходит по второму.

В качестве решения проблемы будем понимать название изобретения, его можно получить путем запроса в базу данных. Алгоритм выделения структуры «Проблема-Решение» [9] представлен на рис. 3.

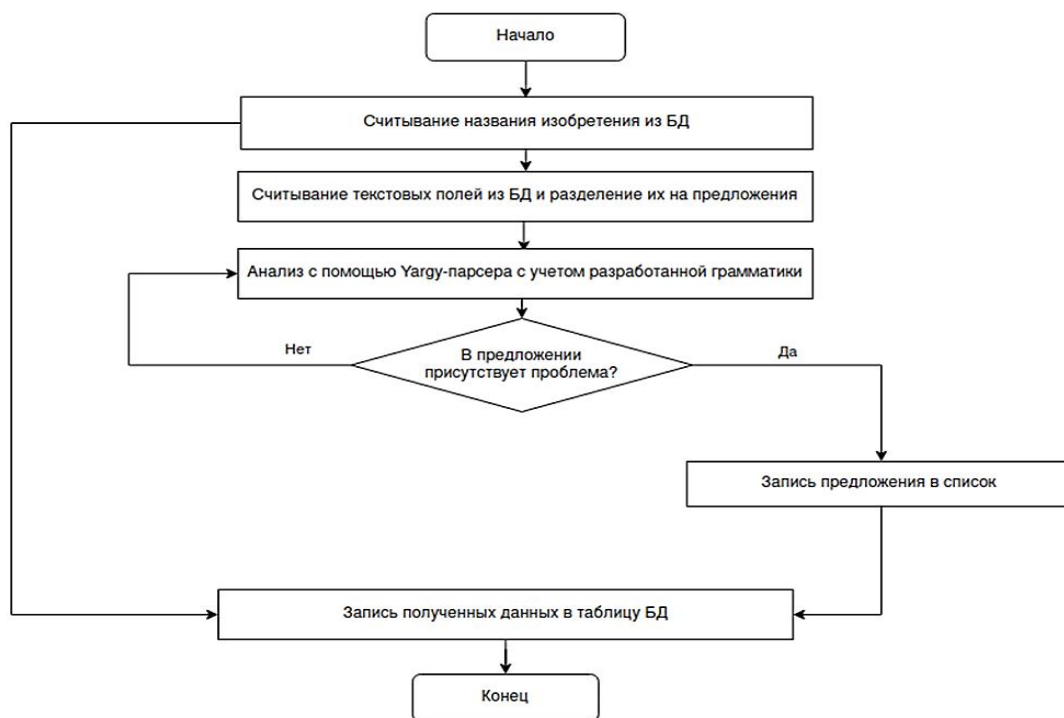


Рис. 3. Алгоритм выделения структур «Проблема-Решение»

Fig. 3. Algorithm of structures selecting “Problem-Solution”

Чтобы сравнить решаемые проблемы, описанные в патентах предприятий, для дальнейшего поиска партнеров, а также выявления «технологического вакуума», производим построение деревьев зависимостей [10] при помощи библиотеки Stanza [11]. Из этого дерева необходимо получить следующие данные:

- вершина дерева, обычно сказуемое (root);
- первый уровень – слово/слова, зависящие от сказуемого (первые дочерние элементы согласно дереву зависимостей), имеющие семантическую роль nmod;
- второй уровень – слово/слова, завершающие основную мысль предложения (дочерние элементы второго уровня дерева зависимостей), имеющие семантические роли pmod, amod, obj.

При сравнении всех слов, за исключением сказуемого, также применяется библиотека RuWord Net [12, 13] для получения гиперонимов, в связи с тем, что один и тот же термин в полученном дереве зависимости можно описать различными синонимичными словами, при этом их будет объединять некий общий термин.

Сравнение происходит в три этапа, на каждом этапе при успешном соответствии повышается коэффициент сравнения.

Первый этап – сравнение сказуемого («Действия»), при этом, если слова не совпали, можно заявить, что «Проблемы» не совпадают. В противном случае коэффициент k_1 становится равным 3.

Второй этап – сравнение первого уровня дерева зависимостей, при совпадении коэффициент второго этапа k_2 становится равным 2. Если слова не совпали даже с учетом гиперонимов, проблемы не совпадают.

Третий этап – сравнение второго уровня дерева зависимостей, при полном совпадении всех слов коэффициент третьего этапа становится равным 1. Если слова не совпали даже с учетом гиперонимов, то все равно в данном случае сравнение будет валидным.

Итоговый коэффициент рассчитывается по следующей формуле:

$$k_f = \frac{k_1 + k_2 + k_3}{6},$$

где k_1, k_2, k_3 – коэффициенты первого, второго и третьего этапа соответственно.

Алгоритм сравнения проблем представлен на рис. 4.

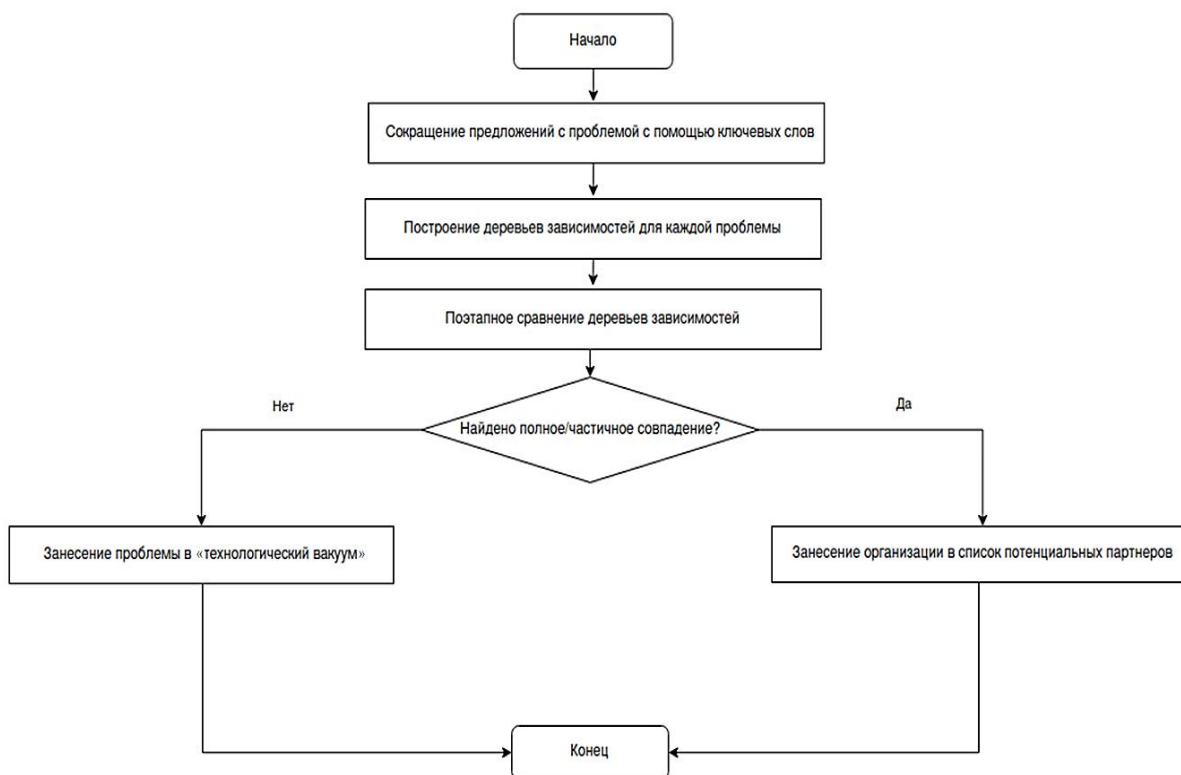


Рис. 4. Алгоритм сравнения решаемых проблем

Fig. 4. Algorithm for problems comparing

Чтобы сформировать «технологический вакуум» для предприятий РФ, следует в первую очередь получить текст иностранных патентов на русском языке при помощи разработанного переводчика. Следующим шагом является выделение структур «Проблема-Решение» из переведенных текстов, чтобы приступить к искомой задаче. Ре-

шаемая проблема иностранного предприятия будет считаться частью «технологического вакуума», если при ее сравнении с проблемами организаций РФ не будет найдено ни одного соответствия. Алгоритм определения «технологического вакуума» представлен на рис. 5.

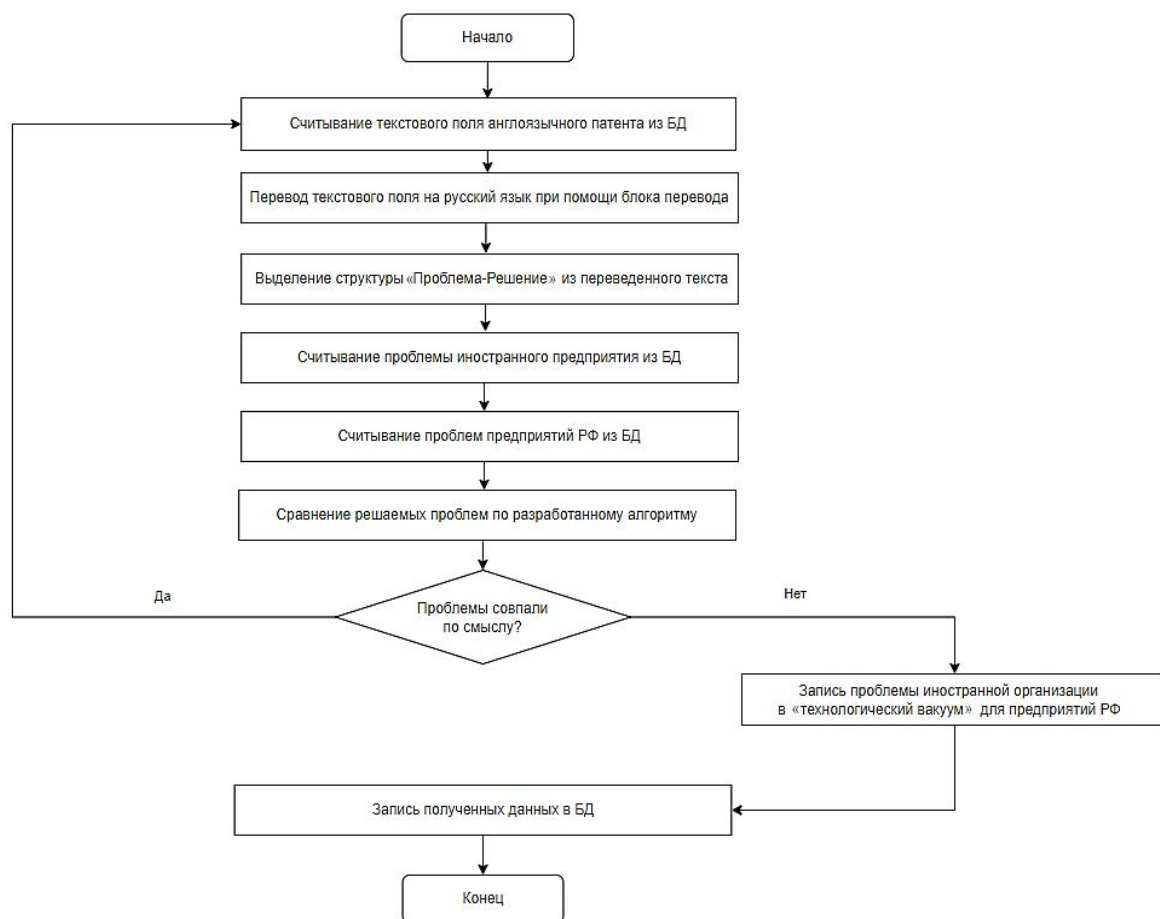


Рис. 5. Алгоритм определения «технологического вакуума»

Fig. 5. “Technological vacuum” formation algorithm

Проектирование модуля поиска технологических возможностей

В качестве среды для разработки модуля поиска технологических возможностей выбран PyCharm Community Edition [14]. Для хранения данных, а также взаимодействия с ними задействована российская СУБД Clickhouse, в качестве языка программирования для создания данного программного продукта выбран Python [15]. Для получения содержимого страницы патента в виде HTML-верстки используется библиотека *requests* [16]. Для парсинга html-документов задействована библио-

тека Beautiful Soup [17]. Если данные на сайте генерируются автоматически, необходимо воспользоваться Selenium WebDriver [18]. Для перевода англоязычных текстовых полей выбрана библиотека *deep-translator* [19].

Программный модуль условно можно разделить на три основных функциональных компонента: парсинг, выделение структур «Проблема-Решение», поиск технологических партнеров и «технологического вакуума».

Архитектура модуля приведена на рис. 6, структура базы данных – на рис. 7.

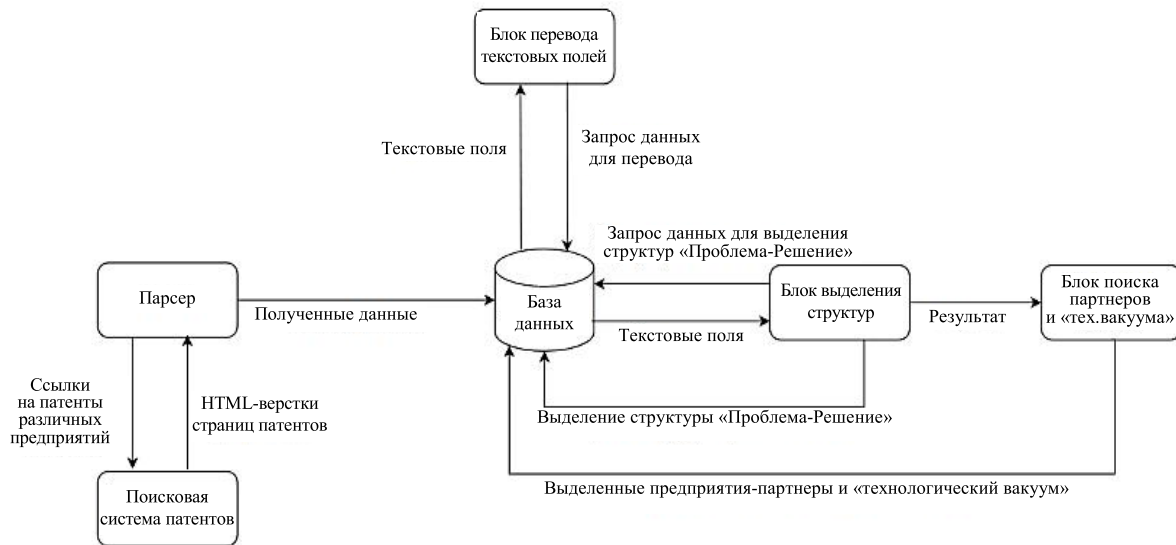


Рис. 6. Архитектура ПО

Fig. 6. Software Architecture

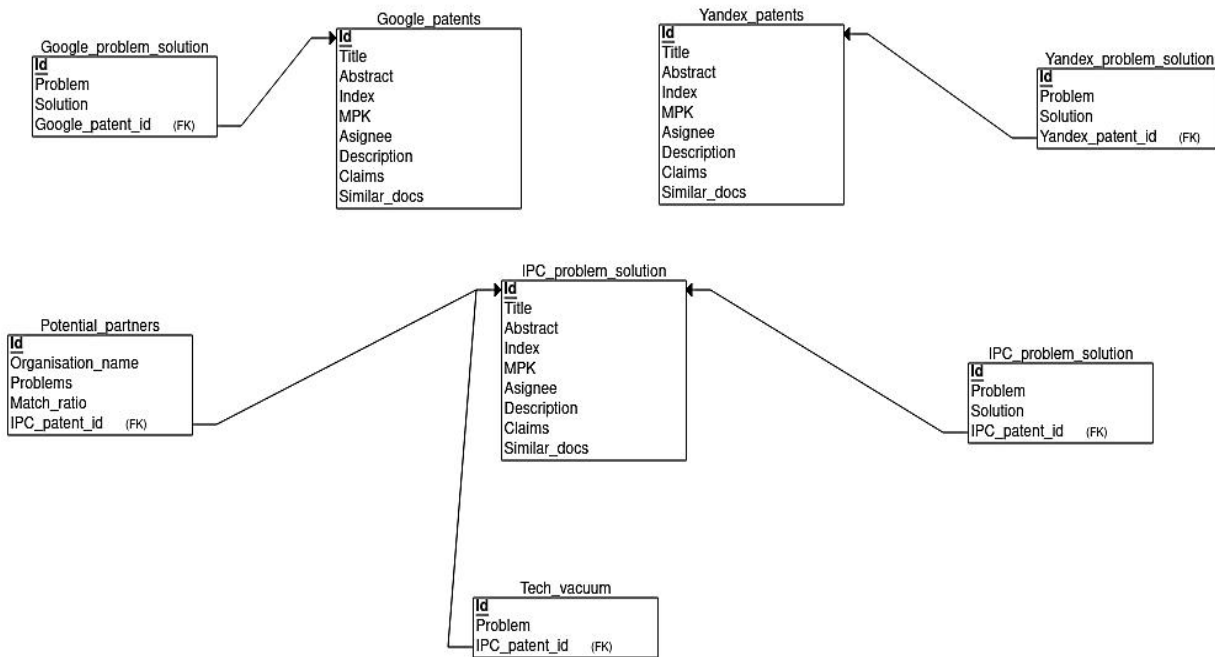


Рис. 7. Структура БД

Fig. 7. Database structure

Полученные при парсинге данные сохраняются в БД в соответствующую таблицу, столбцы которой хранят основные данные, необходимые для

дальнейшего анализа. Пример сохраненных патентов для предприятия «Волжский трубный завод» в таблице «yandex_patents» приведен на рис. 8.

	123 Id ↑	ABC Title	ABC Abstract	ABC Index	ABC IPC
1	1	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛ	Изобретение относится к трубному произво	RU 2 308 338 C1	B21C 37/12¶
2	2	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛ	Изобретение относится к трубному произво	RU 2 334 577 C2	B21C 37/12¶
3	3	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛ	Изобретение относится к трубному произво	RU 2 320 438 C2	B21C 37/12¶
4	4	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛ	Изобретение относится к трубному произво	RU 2 320 439 C2	B21C 37/12¶
5	5	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛ	Изобретение относится к трубному произво	RU 2 308 339 C1	B21C 37/12¶
6	6	СПОСОБ ЭЛЕКТРОДУГОВОЙ СВАРКИ	Изобретение может быть использовано для	RU 2 410 217 C2	B23K 9/18¶B2
7	7	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛ	Изобретение относится к трубному произво	RU 2 308 337 C1	B21C 37/12¶
8	8	УСТАНОВКА ДЛЯ ИСПЫТАНИЯ ТРУБ НА СМЯ	Изобретение относится к испытательной тех	RU 2 381 468 C1	G01M 3/08¶
9	9	УСТРОЙСТВО ДЛЯ ОПРЕДЕЛЕНИЯ ПРОЧНОС	Изобретение относится к испытательной тех	RU 2 394 224 C2	G01N 19/04¶
10	10	СПОСОБ ГОРЯЧЕЙ ПРОКАТКИ БЕСШОВНЫХ Т	Изобретение относится к обработке металл	RU 2 368 440 C1	B21B 19/02¶

Рис. 8. Пример патентных данных, полученных при парсинге

Fig. 8. Example of patent data obtained by parsing

При извлечении структуры «Проблема-Решение» сохраняются в соответствующие поля БД предложения с описанием проблемы, на решение которой направлено данное изобретение, извле-

ченные из текстовых полей патента, а также название запатентованной разработки. Пример извлеченной структуры «Проблема-Решение» приведен на рис. 9.

	123 Id ↑	Problem	ABC Solution	123 Patent_id
1	1	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛЕНИЯ СПИ	1
2	2	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛЕНИЯ СПИ	2
3	3	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛЕНИЯ СПИ	3
4	4	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛЕНИЯ СПИ	4
5	5	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛЕНИЯ СПИ	5
6	6	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	СПОСОБ ЭЛЕКТРОДУГОВОЙ СВАРКИ	6
7	7	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	ТЕХНОЛОГИЧЕСКАЯ ЛИНИЯ ДЛЯ ИЗГОТОВЛЕНИЯ СПИ	7
8	8	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	УСТАНОВКА ДЛЯ ИСПЫТАНИЯ ТРУБ НА СМЯТИЕ ВНЕШ	8
9	9	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	УСТРОЙСТВО ДЛЯ ОПРЕДЕЛЕНИЯ ПРОЧНОСТИ СЦЕПЛ	9
10	10	{Задачей данного изобретения является усовершенствование описанной линии для повышения качества поперечных швов и, следовательно, работоспособности спирально-шовных труб в целом.}	СПОСОБ ГОРЯЧЕЙ ПРОКАТКИ БЕСШОВНЫХ ТОНКОСТЕЙ	10

Рис. 9. Пример предложения с извлеченной структурой «Проблема-Решение»

Fig. 9. Example of a sentence with the extracted “Problem-Solution” structure

Изначально извлеченные иностранные патенты, найденные по необходимому классу МПК, хранятся в оригинальном виде в общей таблице. Текстовые данные, полученные после перевода, сохраняются в отдельную таблицу с теми же идентификационными номерами, что и в основной. Таким образом, сохранится структурированность информации, что позволит избежать лишних запросов в БД.

изобретениях, происходит запись в БД искомых потенциальных организаций (возможных партнеров) совместно с их решаемыми проблемами, а также коэффициентом соответствия их задач с задачами предприятий Волгоградской области.

В результате сравнения решаемых предприятиями проблем, описанных в их запатентованных

К примеру, рассмотрим решаемую вышеупомянутой организацией «Волжский трубный завод» проблему (рис. 10).

9	8	{Использование предлагаемого изобретения позволяет...}	УСТАНОВКА ДЛЯ ИСПЫТАНИЯ ТРУБ НА СМЯТИЕ ВНЕШ	8
10	9	{}	УСТРОЙСТВО ДЛЯ ОПРЕДЕЛЕНИЯ ПРОЧНОСТИ СЦЕПЛ	9
11	10	{Задача настоящего изобретения состоит в создании...}	СПОСОБ ГОРЯЧЕЙ ПРОКАТКИ БЕСШОВНЫХ ТОНКОСТЕЙ	10
12	11	{Задачей настоящего изобретения является создание...}	СПОСОБ ПРОИЗВОДСТВА БЕСШОВНЫХ ТОНКОСТЕННЫХ	11
13	12	{Задачей настоящего изобретения является создание...}	СПОСОБ ИСПЫТАНИЯ ТРУБ НА СМЯТИЕ В КОНТЕЙНЕРЕ	12
14	13	{Задачей настоящего изобретения является создание...}	ДУГОВОЙ СВАРКИ КОЛЬЦЕВЫХ ШВОВ ПО	13
15	14	{Задачей настоящего изобретения является создание...}	УСТРОЙСТВО	14
16	15	{Задачей настоящего изобретения является создание...}	ПРОИЗВОДСТВА СПИРАЛЬНОШОВНЫХ ТРУБ	15
17	16	{}	СПОСОБ ПРОИЗВОДСТВА СТАЛИ	16
18	17	{Использование предлагаемого способа позволит п...}	СПОСОБ ЦЕНТРИРОВАНИЯ ПРОБКИ ВЕРХНЕЙ ПРИ КРЕ	17
19	18	{Техническим результатом использования предлага...}	СПОСОБ ЭЛЕКТРОДУГОВОЙ СВАРКИ СПИРАЛЬНО-ШОИ	18

Рис. 10. Пример проблемы, решаемой предприятием Волгоградской области

Fig. 10. An example of the problem solved by an enterprise of the Volgograd region

Согласно рисунку одним из направлений деятельности данного предприятия является создание способов производства бесшовных труб. Запустим модуль поиска партнеров и рассмотрим найденные

организации со схожим классом МПК (B21) и их решаемые проблемы. На рис. 11 показан найденный на основе сравнения решаемых проблем потенциальный партнер из Китая.

▶ 10	9	XIANYANG TIANCHENG TITANIUM INDUSTRY	{Три, краткое изложение изобретения}	482	0,916666667
▶ 11	10	Hengyang Valin Steel Tube Co Ltd	{Краткое изложение изобретения}	720	0,916666667
▶ 12	11	Wuxi Baoshun Stainless Steel Co Ltd	{Раскрытие изобретения}	744	0,916666667
▶ 13	12	Открытое акционерное общество "Синарски	{Задачей настоящего изобретения являет	848	0,916666667
▶ 14	13	Dalian University of Technology	{Раскрытие изобретения}	1 270	0,916666667
▶ 15	14	Daye Special Steel Co Ltd	{Раскрытие изобретения}	1 296	0,916666667
▶ 16	15	Inner Mongolia North Heavy Industries Group	{Раскрытие изобретения}	1 303	0,916666667
▶ 17	16	Jiangyin Shanhou Pipe Fittings Co Ltd	{Раскрытие изобретения}	1 662	0,916666667
▶ 18	17	Wuhan Bojin N		1 846	0,916666667
▶ 19	18	Handan Xinxin	{Раскрытие изобретения	2 075	0,916666667
▶ 20	19	Nanjing Iron ai	{Изобретение направлено на создание способа изготовления бесшовной переходной трубы, который может быть использован для изготовления переходной трубы большого диаметра.}	2 142	0,916666667

Рис. 11. Пример найденного технологического партнера

Fig. 11. An example of a found technology partner

При решении задачи поиска «технологического вакуума» для предприятий РФ в патентах иностранных фирм выявляются «Проблемы-Решения», которые не запатентованы российскими патентами. Например, проблема, связанная с созданием различных гибочных устройств для изготовления ав-

томобильных петель, описанная в китайском патенте CN111054869A: «Гибочное устройство для изготовления автомобильных дверных петель и способ гибки» (патентообладатель – компания Nanjing Liju Precision Forging Co., Ltd.) (рис. 12).

189	188	Положительные эффекты настоящего изобретения по сравнению с предшествующим уровнем техники.	{Цель и
190	189	Раскрытие изобретения	{Изобретение направлено на создание гибочного устройства для изготовления автомо-
191	190	Раскрытие изобретения	{Изобретение направлено на создание гибочного устройства для изготовления
192	191	Изобретение направлено на создание гибочного устройства для изготовления	{автомобильной дверной петли и способа ее гибочной обработки, которые можно
193	192	использовать для надежного и точного гибания и формирования предварительно	{формированной дверной петли, которая имеет неправильную форму
194	193	использовать для надежного и точного гибания и формирования предварительно	{формированной дверной петли, которая имеет неправильную форму
195	194	кованой заготовки дверной петли.	{и другое разгрузочное устройс
196	195	В частности, в настоящем изобретении устройство 2 прокатки фартового листа дополнительно содержит двиг	{атель, решаемой проблемы, заключающейся
197	196	Однако, когда усилие прокатки широкого и толстого листа прогнозируется с использованием модели одномерн	{
198	197	Раскрытие изобретения	{Изобретение направлено на создание способа изготовления кольцевого выпуклого м
199	198	РЕЗЮМЕ ПОЛЕЗНОЙ МОДЕЛИ	{Целью настоящего изобретения является предоставление предмета для решени

Рис. 12. Пример проблемы в «технологическом вакууме»

Fig. 12. Example of the problem in “technological vacuum”

Результаты

Рассчитаем точность и полноту автоматизированного извлечения семантических структур «Проблема-Решение» из текстовых полей патентных документов [20]. Данная оценка является ключевой для определения эффективности поиска партнеров и формирования «технологического вакуума», ведь именно извлеченные структуры «Проблема-Решение» являются входными данны-

ми к вышеописанным процедурам.

Для расчета была произведена выборка из 1 000 предложений. Предложение будет считаться релевантным, если в нем содержится проблема, решаемая данной запатентованной разработкой, нерелевантным – если в предложении нет семантической структуры «Проблема-Решение». Результаты приведены в таблице.

Результаты проверки эффективности разработанного программного обеспечения

Results of the software’s effectiveness check

Предложение	Релевантное	Нерелевантное	Всего
Найденное в патентном тексте	653	80	733
Не найденное в патентном тексте	267	–	267
<i>Всего</i>	920	80	1 000

Коэффициент полноты рассчитывается по следующей формуле:

$$R = \frac{a}{a+c},$$

где a – количество релевантных предложений, найденных в тексте; c – количество не найденных релевантных предложений.

Коэффициент точности рассчитывается по следующей формуле:

$$R = \frac{a}{a+b},$$

где b – количество найденных нерелевантных предложений.

Коэффициент полноты: $R = \frac{653}{920} = 0,7$.

Коэффициент точности: $R = \frac{653}{733} = 0,89$.

Заключение

В данной работе рассматривается автоматизация поиска патентообладателей, которые могут быть партнерами предприятий Волгоградской области (решают схожие технологические проблемы). В качестве партнеров рассматриваются компании не только из РФ, но и из Китая, Индии, других дружественных стран. Процесс выявления партнеров строится на основе сходства решаемых технологических проблем, извлеченных из патентных документов, патентообладателями которых являются данные предприятия. Посредством анализа деревьев зависи-

мых, извлеченных из полей патента «Область применения изобретения» и «Название изобретения», формируются структуры «Проблема-Решение». На основе анализа патентных документов потенциальных партнеров выявляются «Проблемы-Решения», не запатентованные в России и, соответственно, определяющие «технологический вакуум» для предприятий РФ.

Теоретическая ценность работы заключается в разработанном методе выявления технологических возможностей на основе анализа мирового патентного массива. Новизна метода, обеспечивающего поиск потенциальных вакантных технологических областей для ключевых предприятий Волгоградской области, заключается в использовании технологий глубокого обучения применительно к анализу естественно-языковых данных мирового патентного массива.

Практическая значимость работы заключается в разработанном программном модуле поиска технологических возможностей для предприятий Волгоградской области на основе схожести решаемых технологических проблем. Для выявления технологических возможностей для предприятий Волгоградской области было проанализировано 6 785 патентов.

Основным направлением совершенствования и дальнейшего развития разработанной системы является повышение точности выделения семантических структур «Проблема-Решение» из текстовых полей патентных документов путем применения нейросетевых технологий на обучающей выборке, составленной путем ручного анализа патентного массива.

Список источников

1. Григорян М. Р. Патентный анализ: стратегическое обоснование, применение, преимущества и ограничения // Науч.-метод. электрон. журн. «Концепт». 2015. Т. 30. С. 341–345.
2. Кашеярова Н. А., Андреева А. А., Пономарева Е. И. Цифровые инструменты патентных исследований // Вопр. инновацион. экономики. 2020. Т. 10, № 2. С. 1059–1074.
3. Коробкин Д. М., Фоменков С. А., Бородин Н. Ю., Верещак Г. А. Автоматизация поиска технологических партнеров для проведения НИОКР // Прикаспийский журнал: управление и высокие технологии. 2022. № 4 (60). С. 59–67.
4. Большаков А. П., Абдулгасис У. А. Патентные исследования в конкурентном соперничестве // Уч. зап. Крым. инженер.-педагог. ун-та. 2010. № 24. С. 79–83.
5. Николаев А. С. Патентная аналитика: учеб.-метод. пособие. СПб.: Изд-во Ун-та ИТМО, 2022. 98 с.
6. Что такое ClickHouse? URL: <https://clickhouse.com/docs/ru> (дата обращения: 01.04.2023).
7. Справочник. Токенизатор // Jupyter Nbviewer: official site. 2023. URL: <https://nbviewer.org/github/natasha/yargy/blob/master/docs/ref.ipynb> (дата обращения: 23.04.2023).
8. Наташа – библиотека для извлечения структурированной информации из текстов на русском языке // PVSM: official site. 2018. URL: <https://www.pvsm.ru/python/275248> (дата обращения: 19.04.2023).
9. Kochura O. A., Korobkin D. M., Fomenkov S. A., Kolesnikov S. G. Development of the patent array analysis module based on the “problem-solution” model // Journal of Physics: Conference Series. Series: International Scientific Conference Artificial Intelligence and Digital Technologies in Technical Systems 2020, AIDTTS 2020. 2021. P. 12–14.
10. Васильев С. С., Коробкин Д. М., Фоменков С. А. Извлечение морфологических признаков технических систем из русскоязычных патентов по деревьям зависимостей // Моделирование, оптимизация и информационные технологии. 2022. Т. 10, № 4 (39). С. 16–17.
11. Dependency Parsing // Stanford NLP: 2023. URL: <https://stanfordnlp.github.io/stanza/depparse.html#accessing-syntactic-dependency-information> (дата обращения: 01.05.2023).
12. Github: 2022. URL: <https://github.com/avidale/python-ruwordnet> (дата обращения: 01.05.2023).
13. Python – интерфейс WordNet // Coderlessons: 2019. URL: <https://coderlessons.com/tutorials/python-technologies/izu>

chite-obrabotku-teksta-na-python/python-interfeis-wordnet (дата обращения: 01.05.2023).

14. PyCharm for Productive Python Development (Guide) // Real Python. 2020. URL: <https://realpython.com/pycharm-guide/> (дата обращения: 01.04.2023).

15. Python Web Scraping Tutorial – How to Scrape Data From Any Website with Python // freeCodeCamp: 2021. URL: <https://www.freecodecamp.org/news/how-to-scrape-websites-with-python-2/> (дата обращения: 03.04.2023).

16. Python Requests Module // w3schools: 2019. URL: https://www.w3schools.com/python/module_requests.asp (дата обращения: 04.04.2023).

17. Guide to Parsing HTML with BeautifulSoup in Python // StackAbuse: 2021. URL: <https://stackabuse.com/guide-to-parsing-html-with-beautifulsoup-in-python/> (дата обращения: 03.04.2023).

18. Selenium WebDriver // Digital Ocean: 2022. URL: <https://www.digitalocean.com/community/tutorials/selenium-webdriver> (дата обращения: 06.04.2023).

19. Deep-translator // pypi: 2023. URL: <https://pypi.org/project/deep-translator/> (дата обращения: 27.04.2023).

20. Мамедалиева Ш. Э. Оценка процесса поиска в информационно-поисковых системах // Вестн. науки и творчества. 2023. № 2 (84). С. 20–17.

References

1. Grigorian M. R. Patentnyi analiz: strategicheskoe obosnovanie, primeneniye, preimushchestva i ogranicheniia [Patent analysis: strategic rationale, application, advantages and limitations]. *Nauchno-metodicheskii elektronnyi zhurnal «Kontsept»*, 2015, vol. 30, pp. 341-345.

2. Kashevarova N. A., Andreeva A. A., Ponomareva E. I. Tsfirovye instrumenty patentnykh issledovaniy [Patent analysis: strategic rationale, application, advantages and limitations]. *Voprosy innovatsionnoi ekonomiki*, 2020, vol. 10, no. 2, pp. 1059-1074.

3. Korobkin D. M., Fomenkov S. A., Borodin N. Iu., Vereshchak G. A. Avtomatizatsiia poiska tekhnologicheskikh partnerov dlia provedeniia NIOKR [Technological partners automation search for R&D]. *Prikaspiiskii zhurnal: upravlenie i vysokie tekhnologii*, 2022, no. 4 (60), pp. 59-67.

4. Bol'shakov A. P., Abdulgazis U. A. Patentnye issledovaniia v konkurentnom sopernichestve [Patent research in competitive Rivalry]. *Uchenye zapiski Krymskogo inzhenerno-pedagogicheskogo universiteta*, 2010, no. 24, pp. 79-83.

5. Nikolaev A. S. *Patentnaia analitika: uchebno-metodicheskoe posobie* [Patent analytics: educational and methodical manual]. Saint-Petersburg, Izd-vo Un-ta ITMO, 2022. 98 p.

6. *Chto takoe ClickHouse?* [What is ClickHouse?]. Available at: <https://clickhouse.com/docs/ru> (accessed: 01.04.2023).

7. Spravochnik. Tokenizator. *Jupyter Nbviewer: official site*. 2023. Available at: <https://nbviewer.org/github/natasha/yargy/blob/master/docs/ref.ipynb> (accessed: 23.04.2023).

8. Natasha – biblioteka dlia izvlecheniia strukturirovannoi informatsii iz tekstov na russkom iazyke. *PVSM: official site*. 2018 [Natasha – library for extracting structured information from texts in Russian. PVSM: official site. 2018]. Available at: <https://www.pvsm.ru/python/275248> (accessed: 19.04.2023).

9. Kochura O. A., Korobkin D. M., Fomenkov S. A., Kolesnikov S. G. Development of the patent array analysis module based on the “problem-solution” model. *Journal of Physics: Conference Series. Series: International Scientific Conference Artificial Intelligence and Digital Technologies in Technical Systems 2020, AIDTTS 2020*. 2021. Pp. 12-14.

10. Vasil'ev S. S., Korobkin D. M., Fomenkov S. A. Izvlechenie morfologicheskikh priznakov tekhnicheskikh sistem iz russkoiazychnykh patentov po derev'iam zavisimostei [Extraction of morphological features of technical systems from Russian-language patents by dependency trees]. *Modelirovanie, optimizatsiia i informatsionnye tekhnologii*, 2022, vol. 10, no. 4 (39), pp. 16-17.

11. *Dependency Parsing. Stanford NLP*: 2023. Available at: <https://stanfordnlp.github.io/stanza/depparse.html#accessing-syntactic-dependency-information> (accessed: 01.05.2023).

12. *GitHub*: 2022. Available at: <https://github.com/avidale/python-ruwordnet> (accessed: 01.05.2023).

13. Python – interfeis WordNet [Python – WordNet interface]. *Coderlessons*: 2019. Available at: <https://coderlessons.com/tutorials/python-technologies/izvliche-obrabotku-teksta-na-python/python-interfeis-wordnet> (accessed: 01.05.2023).

14. *PyCharm for Productive Python Development (Guide)*. Real Python. 2020. Available at: <https://realpython.com/pycharm-guide/> (accessed: 01.04.2023).

15. Python Web Scraping Tutorial – How to Scrape Data From Any Website with Python. *freeCodeCamp*: 2021. Available at: <https://www.freecodecamp.org/news/how-to-scrape-websites-with-python-2/> (accessed: 03.04.2023).

16. *Python Requests Module*. w3schools: 2019. Available at: https://www.w3schools.com/python/module_requests.asp (accessed: 04.04.2023).

17. Guide to Parsing HTML with BeautifulSoup in Python. *StackAbuse*: 2021. Available at: <https://stackabuse.com/guide-to-parsing-html-with-beautifulsoup-in-python/> (accessed: 03.04.2023).

18. Selenium WebDriver. *Digital Ocean*: 2022. Available at: <https://www.digitalocean.com/community/tutorials/selenium-webdriver> (accessed: 06.04.2023).

19. *Deep-translator*. pypi: 2023. Available at: <https://pypi.org/project/deep-translator/> (accessed: 27.04.2023).

20. Mamedaliev Sh. E. Otsenka protsessa poiska v informatsionno-poiskovykh sistemakh [Evaluation of the search process in information search engines]. *Vestnik nauki i tvorchestva*, 2023, no. 2 (84), pp. 20-17.

Статья поступила в редакцию 28.07.2023; одобрена после рецензирования 19.09.2023; принята к публикации 17.10.2023
The article was submitted 28.07.2023; approved after reviewing 19.09.2023; accepted for publication 17.10.2023

Информация об авторах / Information about the authors

Дмитрий Михайлович Коробкин – кандидат технических наук, доцент; доцент кафедры систем автоматизированного проектирования и поискового конструирования; Волгоградский государственный технический университет; dkorobkin80@mail.ru

Всеволод Сергеевич Горкин – студент по направлению «Информатика и вычислительная техника»; Волгоградский государственный технический университет; mad_gor1975@mail.ru

Сергей Алексеевич Фоменков – доктор технических наук, профессор; профессор кафедры систем автоматизированного проектирования и поискового конструирования; Волгоградский государственный технический университет; saf@vstu.ru

Dmitry M. Korobkin – Candidate of Technical Sciences, Assistant Professor; Assistant Professor of the Department of Computer-aided Design and Search Engineering; Volgograd State Technical University; dkorobkin80@mail.ru

Vsevolod S. Gorkin – Student of the direction of “Computer Science and computer Engineering”; Volgograd State Technical University; mad_gor1975@mail.ru

Sergey A. Fomenkov – Doctor of Technical Sciences, Professor; Professor of the Department of Computer-aided Design and Search Engineering; Volgograd State Technical University; saf@vstu.ru

