

Научная статья

Original article



УДК 347.771

EDN <https://elibrary.ru/pjxprz>

Патентный поиск по химическим соединениям с использованием сервиса «Синтелли» в цифровой поисковой платформе Роспатента

Анна Евгеньевна Пинигина* [✉], Алина Альбертовна Мухамеджанова*,
Наталья Геннадьевна Шимкина**, Дмитрий Сергеевич Золкин**

*ООО «Синтелли», **Федеральный институт промышленной собственности

[✉]pinigina.a.e@syntelly.com

Аннотация: статья посвящена решению актуальной проблемы, возникшей перед Федеральным институтом промышленной собственности (ФИПС), – необходимостью создания отечественного инструмента для проведения патентного поиска по химическим соединениям. Актуальность задачи обусловлена сложностью и высокой значимостью такого поиска для патентной экспертизы, а также уходом с российского рынка многих зарубежных коммерческих систем в 2022 году. Проведение качественного патентного поиска в области химии является нетривиальной задачей, поскольку одно и то же соединение может быть представлено в документах различными способами: структурной формулой, химическим названием, брутто-формулой и т. д., что требует от эксперта использования специализированных инструментов. В качестве решения данной проблемы была выбрана российская платформа искусственного интеллекта для органической химии «Синтелли». В рамках совместного проекта ФИПС и «Синтелли» при поддержке фонда «Сколково» была проведена доработка платформы и ее интеграция в цифровую поисковую платформу Роспатента. Ключевой задачей стало создание инструмента для извлечения химической информации из текстов патентов и научных статей. Для этого был разработан пайплайн на основе технологии распознавания именованных сущностей (NER), позволяющий находить русскоязычные химические названия в документах и преобразовывать их в машиночитаемые структурные формулы. Для обеспечения полноценного поиска были доработаны модули подструктурного поиска, поиска по химическому подобию и текстового поиска. Главным результатом работы стала успешная интеграция сервиса «Синтелли» в платформу Роспатента, что предоставило экспертам ведомства современный и независимый инструмент для поиска патентной и непатентной литературы по химическим структурам. Внедрение такого отечественного решения позволяет укрепить технологический суверенитет страны. В перспективе планируется дальнейшее развитие системы, включая улучшение распознавания структур, реализацию поиска по структурам Маркуша и по реакциям.

Ключевые слова: патентный поиск, химический поиск, платформа «Синтелли», искусственный интеллект, цифровая поисковая платформа Роспатента.

Для цитирования: Пинигина А. Е., Мухамеджанова А. А., Шимкина Н. Г., Золкин Д. С. Патентный поиск по химическим соединениям с использованием сервиса «Синтелли» в цифровой поисковой платформе Роспатента // Вестник ФИПС. 2025. Т. 4, № 14 (4). С. 386–395.

Благодарности: работа выполнялась в рамках гранта между ФИПС и некоммерческой организацией «Фонд развития Центра разработки и коммерциализации новых технологий» (фонд «Сколково»), источником финансового обеспечения которого является субсидия, предоставленная из федерального бюджета от 29.09.2023 г. № 30601/0006–2023. Коллектив авторов благодарит руководителя проекта со стороны ФИПС – заместителя директора А. А. Полякову и научного руководителя работ доктора хим. наук, члена-корреспондента РАН, исполняющего обязанности директора Института проблем передачи информации РАН им. А. А. Харкевича М. В. Федорова.

Patent search for chemical compounds using the "Syntelly" service in the Rospatent digital search platform

Anna E. Pinigina* [✉], Alina A. Muhamedzhanova*, Natalia G. Shimkina**, Dmitry S. Zolkin**

*Syntelly LLC, **Federal Institute of Industrial Property (FIPS)

[✉]pinigina.a.e@syntelly.com

Abstract: this article addresses a pressing issue faced by the Federal Institute of Industrial Property (FIPS) – the need to create a domestic tool for patent search on chemical compounds. The relevance of this task is determined by the complexity and high importance of such searches for patent examination, as well as that many foreign commercial systems leaved Russian market in 2022. Conducting a high-quality patent search in the field of chemistry is a non-trivial task, since the same compound can be represented in documents in various ways: by a structural formula, chemical name, molecular formula, etc., which requires the expert to use specialized tools. The Russian artificial intelligence platform for organic chemistry, "Syntelly", was chosen as a solution to this problem. As part of a joint project between FIPS and "Syntelly" with the support of the "Skolkovo" foundation, the platform was refined and integrated into the Rospatent Digital Search Platform. A key task was the creation of a tool for extracting chemical information from patent texts and scientific articles. For this purpose, a pipeline based on Named Entity Recognition (NER) technology was developed, enabling the identification of Russian-language chemical names in documents and their conversion into machine-readable structural formulas. To ensure a comprehensive search, modules for substructure search, chemical similarity search, and text search were enhanced. The main result of the work was the successful integration of the "Syntelly" service into the Rospatent Platform, providing the agency's experts with a modern and independent tool for searching patent and non-patent literature based on chemical structures. The implementation of such a domestic solution helps to strengthen the country's technological sovereignty. Future plans include further development of the system, including improved structure recognition, implementation of Markush structure search, and reaction search.

Keywords: patent search, chemical search, "Syntelly" platform, artificial intelligence, Rospatent Digital Search Platform.

For citation: Pinigina A. E., Muhamedzhanova A. A., Shimkina N. G., Zolkin D. S. Patent search for chemical compounds using the "Syntelly" service in the Digital Search Platform of Rospatent // Bulletin of Federal Institute of Industrial Property. 2025. Vo. 4, no. 4 (14) 386–395 (In Russ.).

Acknowledgments: this work was carried out within the framework of a grant agreement between FIPS and the non-profit organization Foundation for the Development of the Center for the Development and Commercialization of New Technologies ("Skolkovo" foundation), the financial source for which is a subsidy provided from the federal budget from 29.09.2023 № 30601/0006–2023. The authors thank the project lead from FIPS – Deputy Director A. A. Polyakova and the scientific supervisor of the work – Dr. Sci. (Chemistry), Corresponding Member of the Russian Academy of Sciences, Acting Director of the A. A. Kharkevich Institute of Information Transmission Problems of the Russian Academy of Sciences, M. V. Fedorov.

Введение

Федеральный институт промышленной собственности (ФИПС), являясь подведомственным учреждением Федеральной службы по интеллектуальной собственности (Роспатент), осуществляет экспертизу заявок на объекты патентного права, в том числе международных заявок, и проводит международный поиск по ним. Более 800 экспертов ФИПС проводят экспертизу заявок по всем основным направлениям науки и техники. Процесс рассмотрения патентной заявки включает в себя выполнение патентного поиска. Информационные поиски в мировых базах научной, патентной и общедоступной информации являются нетривиальной задачей, поскольку эксперт

патентного ведомства не имеет права пропустить, возможно, единственный документ, который может послужить причиной отказа или, наоборот, выдачи патента.

Для проведения поиска эксперты используют не только общедоступную информацию и патентные базы данных, но и специализированные базы. В области химии поиск особенно сложен. Для этого необходим специальный патентный поиск по химическим соединениям, когда с каждым документом связаны химические структуры, по которым необходимо осуществлять поиск. Такой поиск является сложной задачей, поскольку одно и то же соединение может обозначаться в уровне техники разными способами:

в виде структурной формулы, химического наименования, брутто-формулы, формулы Маркуша и т. д. [1]. Экспертам необходимы различные механизмы поиска по всем возможным идентификаторам.

В мире существуют различные решения для подобного поиска, большинство из них коммерческие, в России у ФИПС собственного инструмента поиска по химическим структурам не было. С 2022 года использование экспертами ведомства иностранных систем стало затруднительным, поскольку многие из них ушли с российского рынка. Для ФИПС остро встала проблема замены ушедших решений и создания собственного независимого инструмента, который позволил бы эффективно проводить патентный поиск по заявкам, содержащим химические соединения. Созрела необходимость в создании отечественной системы, способной проводить специализированный поиск по химическим структурам.

Эксперты ФИПС остановили свой выбор на продукте компании «Синтелли». «Синтелли» – это российская платформа искусственного интеллекта для органической химии. Для решения задачи создания инструмента проведения патентного поиска по химическим соединениям ФИПС и «Синтелли» совместно подготовили заявку в фонд «Сколково» на условиях совместного финансирования. Целью проекта была доработка продукта и интеграция сервиса «Синтелли» в цифровую платформу поиска патентной и непатентной информации ФИПС¹. Заявка прошла все необходимые экспертизы и была одобрена экспертным советом фонда «Сколково». Реализация решения осуществлялась на стороне «Синтелли», а затем была проведена интеграция в платформу Роспатента. В итоге на платформе Роспатента появился патентный поиск по химическим соединениям, а для «Синтелли» же были проведены доработки платформы, которые доступны всем ее пользователям.

ОБЗОР ЛИТЕРАТУРЫ

Системы для проведения патентного поиска по химическим соединениям

Для поиска патентной и непатентной литературы по химическим структурам существует несколько специализированных поисковых систем [2].

1. Patentscope² – химический поиск на сайте Всемирной организации интеллектуальной собственности (ВОИС). В результате поиска можно получить полные тексты патентов и заявок РСТ с графическим представлением структуры. Патенты Российской Федерации и Евразийского патентного ведомства (ЕАПВ) проиндексированы для структурного поиска. Также доступны рефераты журналов от Nature и MDPI.

2. PubChem³ – химический поиск на сайте Национального центра биотехнологической информации (NCBI). Эта база данных состоит из трех взаимосвязанных разделов: Substance, Compound и BioAssay. В PubChem содержится более 338 млн описаний химических веществ, 122 млн уникальных химических структур и 298 млн описаний биологических последовательностей, а также 44 млн статей из химических журналов.

3. STNnext⁴ от Clarivate Analytics' Derwent World Patent Index (DWPI) – это универсальная база данных, доступная в крупнейших патентных ведомствах мира. STNnext содержит информацию по химическим структурам, неорганическим сплавам, полимерам и генетическим последовательностям. Она используется учеными, специалистами и патентными экспертами для оценки абсолютной новизны объектов интеллектуальной собственности. Основные поисковые системы в базе данных – REAXIS и Chemical Abstract. STN предоставляет доступ к мировому непатентному химическому контенту из Chemical Abstracts Service (CAS) и патентам из DWPI на интегрированной платформе. 95 % патентных заявок в мире рассматриваются патентными ведомствами, которые предоставляют свои данные в STN. Кроме того, 25 крупнейших фармацевтических компаний мира поставляют свои данные в эту базу.

4. Reaxys⁵ от Elsevier – база данных, которая позволяет находить по химической структуре статьи, опубликованные в журналах, а также патентную литературу. До недавнего времени доступ к Reaxys был возможен на рабочих местах экспертов ФИПС, однако с уходом компании с российского рынка возникла необходимость в создании отечественного инструмента для проведения химического структурного поиска на цифровой поисковой платформе Роспатента.

Методы извлечения информации о химических соединениях из документов

При создании с нуля решения, позволяющего искать патентную и непатентную литературу по химическим соединениям, самой важной частью являются данные – связь документов и химических структур в них. Часть таких данных можно найти в открытых источниках, однако для получения большей части необходим специальный инструмент извлечения химических соединений из патентов и статей.

Информация в документах находится в разных модальностях: картинки, текст, таблицы. Для получения всей полноты информации необходимо извлекать данные из всех этих модальностей. Чаще всего эти модальности взаимосвязаны, и необходимо их объединять, устанавливать связи между ними, однако это является

¹ Цифровая поисковая платформа Роспатента, доступ: <https://searchplatform.rospatent.gov.ru/> (дата обращения: 11.11.2025).

² Patentscope, доступ: <https://patentscope.wipo.int> (дата обращения: 11.11.2025).

³ PubChem, доступ: <https://pubchem.ncbi.nlm.nih.gov> (дата обращения: 11.11.2025).

⁴ STNnext, доступ: <http://www.stn-international.com> (дата обращения: 12.11.2025).

⁵ Reaxys, доступ: <https://www.reaxys.com> (дата обращения: 11.11.2025).

очень сложной задачей. Попытки мультимодального извлечения химической информации из документов предпринимались в инструменте OpenChemIE [3]. Он объединяет в себе модели текстового и оптического распознавания молекул, текстового и оптического распознавания реакций, а также модель, связывающую эти модальности. Для мультимодального распознавания могут подходить большие языковые модели, однако обработка документов с их помощью требует большого количества вычислительных мощностей и занимает много времени.

Если говорить про извлечение химических структур из отдельных модальностей, то в этой области существует много работ и инструментов с открытым исходным кодом. Для извлечения названий молекул из текста чаще всего используют технологии распознавания именованных сущностей (Named Entity Recognition – NER), например в инструментах [4, 5]. Качество извлечения названий химических веществ для таких моделей обычно достаточно хорошее. Методы компьютерного зрения на текстах неточны, и для их обучения необходимо гораздо больше данных. Для этой задачи можно использовать большие языковые модели, но это также дорого. Отдельной задачей при текстовом извлечении данных о соединениях является последующее преобразование полученных названий в химические структуры.

По оптическому распознаванию молекул с картинок документов довольно много работ [6, 7]. Этот способ извлечения использует технологии компьютерного зрения, могут использоваться сверточные нейронные сети или трансформеры для изображений. Также нужны инструменты, которые бы детектировали на странице документа сами картинки химических соединений, их меньше, но они есть с открытым исходным кодом. Однако общее качество оптического распознавания молекул очень сильно зависит от стиля отрисовки химических структур. Чаще всего такие инструменты работают хорошо на каком-то одном стиле и плохо на других. Они могут показывать хорошее качество на бенчмарках, а в реальных документах нередко можно сталкиваться с очень большим количеством неверно распознанных молекул, которые засоряют итоговые результаты. Отдельной сложной задачей является распознавание структур Маркуша (обобщенных химических структур) с картинок.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

При создании решения, которое позволяло бы делать поиск патентной и непатентной литературы по химическим соединениям, описанным в них, «Синтелли» необходимо было решить несколько задач.

Инструмент для извлечения информации о химических соединениях из документов

В первую очередь «Синтелли» было необходимо создать свой инструмент для извлечения химической информации из документов. При использовании

оптического извлечения сложно контролировать случаи неправильного распознавания, результаты могут засоряться не относящимися к документу химическими структурами. Поэтому на данном этапе работ было выбрано распознавание молекул в тексте, при этом необходима была возможность работы с русским языком. Нужна была end-to-end система, которая позволит быстро извлекать химические структуры из текста всего документа, обладать хорошим качеством и возможностью масштабирования. На основе указанных критериев была выбрана NER-технология.

В рамках NER-подхода был разработан пайплайн для обработки текстов документов, включающий несколько этапов.

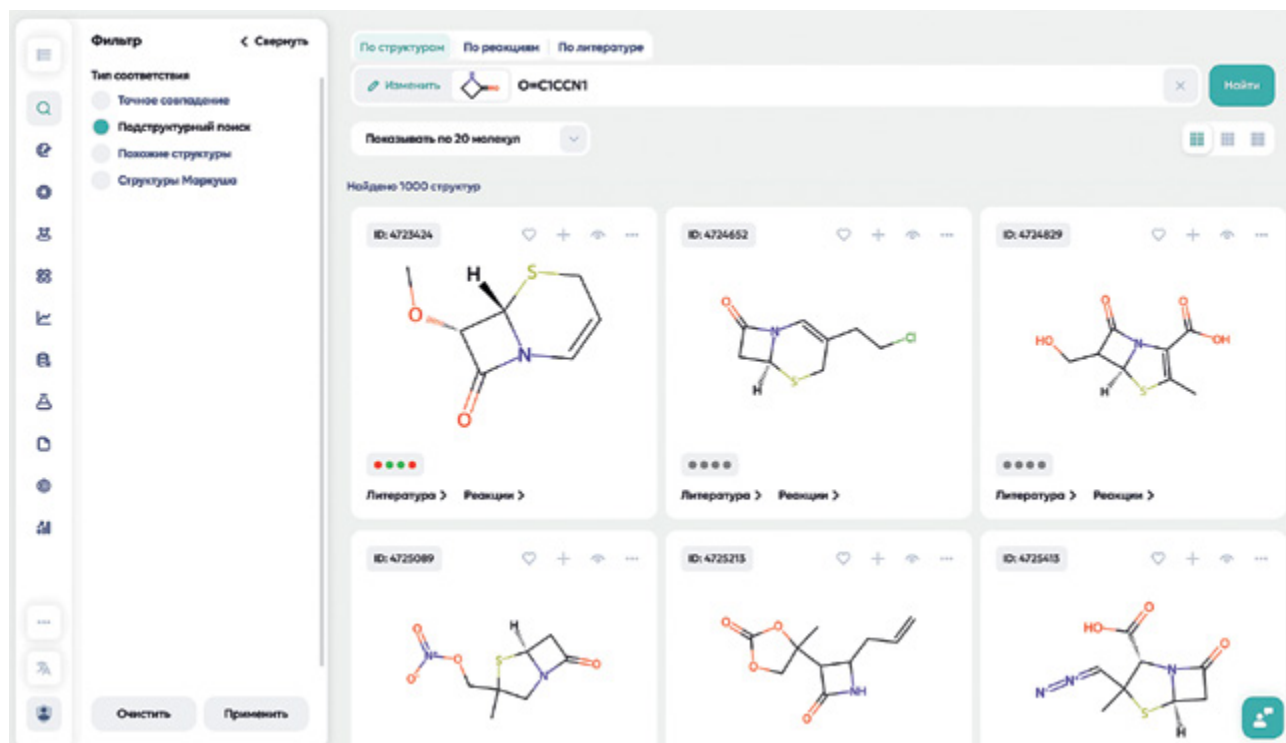
1. Извлечение и предварительная обработка текста патента или научной статьи.

2. Детекция протоколов: на этом этапе с помощью модели детектора из текста извлекаются химические протоколы. В качестве предварительно обученной модели использовалась anferico/bert-for-patents, обученная на большом корпусе патентов (более 100 млн).

3. Распознавание сущностей: из химического протокола извлекаются сущности – химические названия. В качестве предварительно обученной модели использовалась также модель anferico/bert-for-patents в рамках задачи MultiLabel Token Classification.

Данные для обучения на обоих этапах включали примерно 1000 патентов с разметкой BIO, выполненной ассессорами. Функции потерь для обучения моделей: CrossEntropyLoss для этапа детекции протоколов, MultiLabelSoftMarginLoss для этапа распознавания химических названий. Валидация моделей проводилась с помощью 3-фолдовой кросс-валидации и отложенного теста.

4. Преобразование химических названий в соединения: перевод русскоязычных названий веществ в структуры (SMILES строки). Использовался также инструмент, разработанный в «Синтелли». В его основе лежит алгоритмический инструмент с открытой лицензией OPSIN [8], который предназначен для перевода англоязычных ИЮПАК названий в SMILES. Он пытается поделить слово на морфемы, несущие определенный химический смысл с точки зрения структуры вещества, определить их структурное значение, а затем по определенным правилам собрать из них молекулу. Задача состояла в переводе большого количества морфем с английского языка на русский. Дополнительно необходимо было обработать все исключения, связанные с разницей английского и русского языков. Например, под один структурный фрагмент на русском языке может быть больше вариаций написания, или, наоборот, разные фрагменты на английском языке имеют одинаковый перевод на русский. Дополнительной сложностью был учет окончаний в русском языке – под все падежи и числа добавлялись альтернативные морфемы – концы слов. Полученные SMILES затем стандартизировались и канонизировались.



Источник: платформа «Синтелли» Source: Syntelly platform

Рисунок 1.

Пример выполнения подструктурного поиска на платформе «Синтелли». В качестве запроса использовался бета-лактамыный фрагмент – обязательная часть структуры бета-лактамыных антибиотиков

Figure 1.

An example of performing a substructural search on the Syntelly platform – a beta-lactam fragment was used as a query – an obligatory part of the structure of beta-lactam antibiotics

Процесс обработки документов

Для выполнения работ «Синтелли» был необходим сам массив документов, из которого должны были извлекаться химические соединения. В зависимости от способа извлечения информации нужны полные тексты документов или PDF файлы (или набор картинок). Данные по патентам для извлечения химических структур «Синтелли» были переданы непосредственно ФИПС в формате ВОИС XML ST.96 [9]. Были отобраны российские патенты с химическими МПК: C07*, A61K*, C02F*, C05*, C08*, C12*, C13*, A01N*, A23*, G01N* – около 400 тыс. документов. Для обработки необходимы определенные вычислительные мощности, в том числе доступ к графическим процессорам.

Хранение полученных данных и поиск

Далее «Синтелли» была необходима система хранения всех данных по соединениям и документам, для этого использовался PostgreSQL. Необходимо было также доработать модуль поиска по литературе для запросов по этим данным – искать по структурам связанные документы, сохранять возможность поиска по текстовым запросам. Кроме того, для ФИПС требовалось сделать

подструктурный поиск и поиск по похожим структурам, позволяющий найти близкие по химической структуре соединения, для оценки патентоспособности соединений по критерию «изобретательский уровень». Поисковый движок должен выдавать по запросу информацию точно, быстро и релевантно. Для эффективного поиска также требуется наличие возможности поиска по дополнительным условиям и ограничивающие фильтры. Разные виды поиска по структурам и поиск по тексту отличаются по механизму, поэтому для этих целей нужны разные поисковые движки. Разработка с нуля таких решений потребовала бы большого количества ресурсов, однако существуют инструменты с открытыми лицензиями, которые можно использовать под свои нужды. Для подструктурного поиска и поиска по подобию структур можно использовать специальные картриджи для баз данных, один из самых известных открытых инструментов – Bingo⁶ от EPAM. Для текстового поиска известное и проверенное открытое решение – это Elasticsearch⁷.

⁶ Bingo // Life sciences open source: сайт. URL: <https://lifescience.opensource.epam.com/bingo/> (дата обращения: 16.11.2025)

⁷ Elastic search: сайт. URL: <https://www.elastic.co> (дата обращения: 16.11.2025).

Результаты исследования

В работе по гранту ФИПС с фондом «Сколково» по апробации технологий искусственного интеллекта в приоритетных отраслях в рамках реализации федерального проекта «Искусственный интеллект» национальной программы «Цифровая экономика Российской Федерации» были выполнены следующие основные работы.

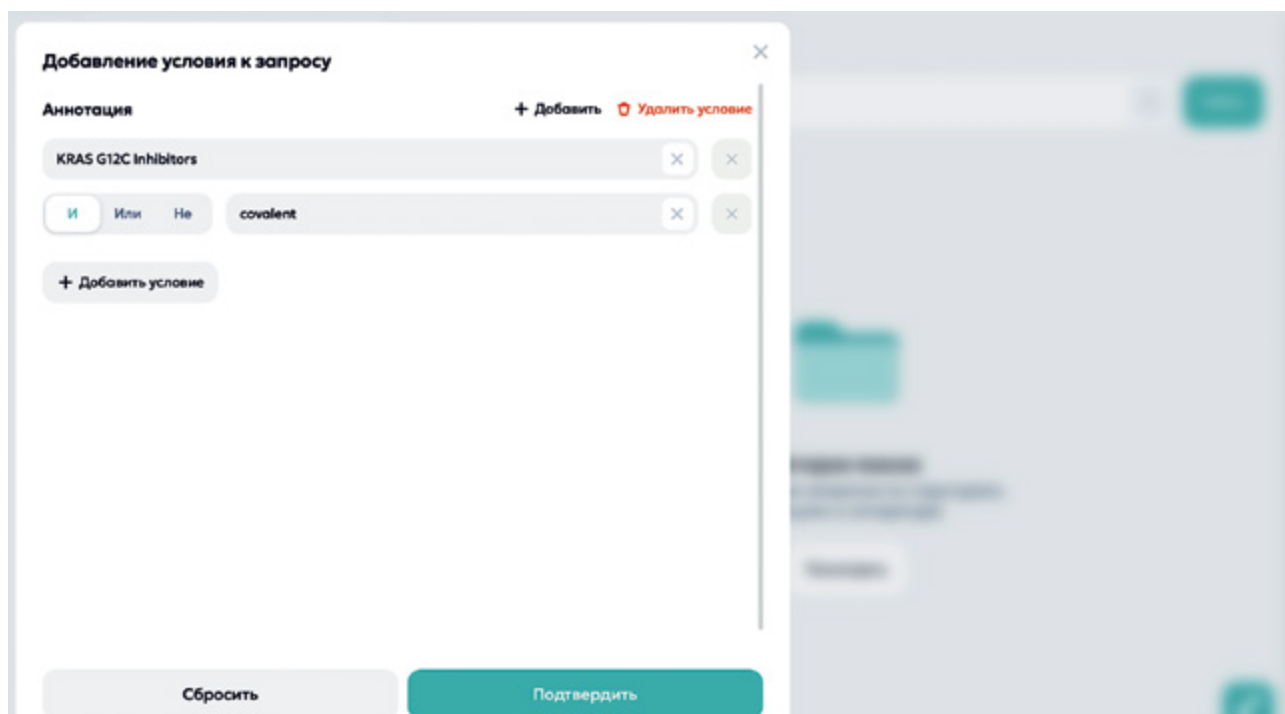
Была проведена доработка платформы «Синтелли» в части функционала загрузки и автоматического распознавания химической информации из патентных документов России. Был реализован инструмент для распознавания русскоязычных названий химических соединений из текстов документов на основе NER-технологии, обработан массив патентных документов от ФИПС – извлечены химические структуры из документов и загружены в базу данных «Синтелли».

Проведена доработка функционала поиска платформы «Синтелли». Для подструктурного поиска и поиска по подобию был использован `bingo` в качестве картриджа для PostgreSQL. На рисунке 1 показан пример выполнения подструктурного поиска. В качестве запроса использовался бета-лактамный фрагмент – обязательная часть структуры бета-лактамных антибиотиков, которые можно найти среди поисковой выдачи. Для

поиска по подобию на платформе «Синтелли» можно задать диапазон интересующего подобия в процентах.

Для текстового поиска в «Синтелли» используется Elasticsearch. На платформе предусмотрены дополнительные условия поиска, такие как заголовок, аннотация, автор, DOI, номер патента, журнал, издатель, патентообладатель, заявитель, полный текст и формула изобретения. Есть возможность составлять сложные комбинированные запросы из условий с использованием логических операторов между ними (пример на рисунке 2). Также доступны фильтры по типу документа (статья или патент) и языку (русский или английский). Для удобства поиска можно задать диапазон годов публикации.

Главная часть поиска – поиск по документам напрямую по химическим структурам – реализована следующим образом. На платформе вводится структура химического соединения с помощью молекулярного редактора или идентификатор молекулы: «Синтелли» ID, SMILES, тривиальное название, название ИЮПАК, CAS, InChI, InChIKey, идентификаторы других баз данных. С помощью PostgreSQL или Elasticsearch структура находится в базе данных молекул «Синтелли», а затем осуществляется запрос в связанные таблицы с патентами и научными статьями. Таким образом, в поисковой



Источник: платформа «Синтелли» Source: Syntelly platform

Рисунок 2.

Пример составления комбинированного текстового поискового запроса на платформе «Синтелли»

Figure 2.

An example of creating a combined text search query on the Syntelly platform

выдаче показываються документы, для которых есть связи с молекулой-запросом.

Также была проведена доработка платформы «Синтелли» для просмотра информации об источнике – в интерфейсе отображена вся доступная информация, а также показаны структуры, связанные с документом. Если был сделан поисковый запрос, то искомая молекула подсвечивается первой. Также было реализовано сохранение истории поиска пользователя (поисковые запросы и результаты).

Главный результат выполнения работ – была проведена интеграция сервиса химического поиска «Синтелли» в цифровую платформу Роспатента. «Синтелли» предоставляет Роспатенту доступ к своему API. Через интерфейс платформы Роспатента пользователь (эксперт) делает поисковый запрос, он передается в API «Синтелли», на стороне «Синтелли» запрос обрабатывается, и данные передаются обратно на платформу Роспатента. Для доступа к «Синтелли» для ФИПС был зарегистрирован пользователь, доступ осуществляется по подписке, как и для остальных пользователей платформы. Кроме того, ФИПС ежемесячно передает «Синтелли» актуальный массив российских химических патентов через FTP-сервер, таким образом, интеграция является частично двусторонней.

Пример патентного поиска по химическим соединениям в цифровой поисковой платформе Роспатента и на платформе «Синтелли»

На рисунках 3 и 4 показан интерфейс химического поиска на цифровой поисковой платформе Роспатента и пример результатов поиска патентной и непатентной литературы по химическому соединению. В качестве примера был взят оклацитиниб (или апоквел) – средство для лечения дерматита и зуда у животных. С помощью графического редактора вводится структура соединения, она преобразуется в SMILES-строку, после чего делается поисковый запрос на платформе. В поисковой выдаче находятся соответствующие патенты и статьи, среди документов можно сделать выборку по типу документа (патент или статья), а также по языку (английский или русский). На рисунке 4 показано, как отображается подробная информация об источнике вместе со связанными с документом химическими структурами.

На рисунке 5 показаны результаты выполнения того же поискового запроса на платформе «Синтелли» – пользователь получает те же результаты, что и на платформе Роспатента.

The screenshot displays the Syntelly search interface. On the left is a filter sidebar with options for document type (Article, Patent), language (Russian, English), and publication date. The main search area shows the input SMILES string: CNS(=O)(=O)CC1CCC(N(C1c2ncnc3[nH]ccc23)CC1. Below the search bar, there are tabs for 'По структурам', 'По реакциям', and 'По литературе'. The search results are displayed in a list format, showing patent titles, numbers, and abstracts. The first result is 'Palatable formulations' with patent number US-2022062287-A1, and the second is 'Dosing regimen for Janus Kinase (JAK) inhibitors' with patent number US-9522151-82. The interface also includes a search bar with a 'Найти' button and a 'Показывать по 20 источникам' dropdown.

Источник: платформа «Синтелли» Source: Syntelly platform

Рисунок 5.

Пример патентного поиска для структуры оклацитиниба на платформе «Синтелли»

Figure 5.

An example of a patent search for the oclacitinib structure on the Syntelly platform

Заключение

Внедрение сервиса «Синтелли» в цифровую поисковую платформу Роспатента стало важным шагом в развитии отечественных технологий патентного поиска. Совместная работа «Синтелли» и ФИПС позволила создать решение для поиска патентной информации по химическим соединениям, полностью базирующееся на российских разработках и серверах. Благодаря химическому поиску, комбинированным условиям поиска и использованию различных фильтров эксперты Роспатента получили инструмент для быстрого и точного нахождения актуальных документов по химическим соединениям. Это значительно повышает эффективность патентного поиска и ускоряет процесс обработки патентных заявок.

Дальнейшие разработки «Синтелли» будут направлены в сторону улучшения распознавания химических структур из документов, покрытия соединениями большего объема данных, а также доработки других механизмов поиска по структурам и документам. Очень важное значение имеет возможность искать молекулы и патенты по структурам Маркуша (обобщенным структурам), так как чаще всего подаются заявки на патент групп соединений, задаваемых общей формулой. Кроме того, полезной доработкой станет реализация напрямую поиска документов с помощью подструктурного поиска и поиска по подобию. Еще одним интересным направлением является поиск патентов не по соединениям, а по реакциям, так как в химических патентах патентуют также и способы синтеза соединений. Эти улучшения сделают химический поиск по патентам еще более удобным, точным и быстрым, что, в свою очередь, упростит работу экспертов и повысит качество патентной экспертизы.

Внедрение отечественного решения на базе платформы «Синтелли» не только укрепляет технологический суверенитет России, но и способствует развитию науки и промышленности за счет более оперативного доступа к патентной информации.

Список литературы / References

1. Lupu, M., Mayer, K., Kando, N. and Trippe, A. J. (2017), *Current challenges in patent information retrieval*, Springer, Heidelberg, Germany.
2. Ohms, J. (2021), "Current methodologies for chemical compound searching in patents: A case study", *World Patent Information*, vol. 66, no. 102055.
3. Fan, V., Qian, Y., Wang, A., Wang, A., Coley, C.W. and Barzilay, R. (2024), "OpenChemIE: An Information Extraction Toolkit for Chemistry Literature", *Journal of Chemical Information and Modeling*, vol. 64, no. 14, pp. 5521–5534.
4. Isazawa, T. and Cole, J.M. (2022), "Single Model for Organic and Inorganic Chemical Named Entity Recognition in ChemDataExtractor", *Journal of Chemical Information and Modeling*, vol. 62, no. 5, pp. 1207–1213.
5. Wang, X. et al. (2021), "ChemNER: Fine-grained chemistry named entity recognition with ontology-guided distant supervision", *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, November 7–11, 2021, Punta Cana, Dominican Republic.
6. Rajan, K., Brinkhaus, H.O., Agea, M.I. et al. (2023), "DECIMER.ai: an open platform for automated optical chemical structure identification, segmentation and recognition in scientific publications", *Nature Communications*, vol. 14, no. 5045.
7. Morin, L. et al. (2023), "MolGrapher: Graph-based Visual Recognition of Chemical Structures", *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 1–6 October 2023, pp. 19495–19504.
8. Lowe, D.M., Corbett, P.T., Murray-Rust, P. and Glen, R.C. (2011), "Chemical Name to Structure: OPSIN, an Open Source Solution", *Journal of Chemical Information and Modeling*, vol. 51, no. 3, pp. 739–753.
9. World Intellectual Property Organization (2022), WIPO Standard ST.96, Version 6.0: *Recommendation For The Processing Of Intellectual Property Information Using Xml*, доступ: <https://www.wipo.int/export/sites/www/standards/en/pdf/03-96-01.pdf> (Accessed 17 Nov. 2025).

Информация об авторах:

Анна Евгеньевна Пинигина, специалист-хемоинформатик, ООО «Синтелли» (Москва, ул. Вере́йская, 29, стр. 134); ORCID: <https://orcid.org/0009-0007-0704-5956>; pinigina.a.e@syntelly.com

Алина Альбертовна Мухамеджанова, генеральный директор ООО «Синтелли» (Москва, ул. Вере́йская, 29, стр. 134); ORCID: <https://orcid.org/0009-0009-1983-6039>; muhamedzhanova@syntelly.com

Наталья Геннадьевна Шимкина, кандидат химических наук, ведущий государственный эксперт по интеллектуальной собственности отдела химии и фармацевтики, ФГБУ «Федеральный институт промышленной собственности» (Москва, Бережковская наб., д. 30, корп. 1); ORCID: <https://orcid.org/0009-0007-7260-5243>; SPIN: 8841-8908; natalia.shimkina@rupto.ru

Дмитрий Сергеевич Золкин, начальник отдела проектирования информационно-поисковых систем, ФГБУ «Федеральный институт промышленной собственности» (Москва, Бережковская наб., д. 30, корп. 1); ORCID: <https://orcid.org/0009-0002-1641-4518>; SPIN: 3304-2957; db_dept@rupto.ru

Вклад соавторов

А. Е. Пинигина – автор принимала участие в разработке инструмента «Синтелли» для извлечения химических структур из документов и отвечала за написание текста рукописи.

А. А. Мухамеджанова – автор обеспечила получение финансовой поддержки и ресурсов, необходимых для реализации проекта, а также осуществляла администрирование проекта.

Н. Г. Шимкина – автор сопроводила статью информацией о возможности использования сервиса «Синтелли» для экспертизы заявок на изобретения, связанные с химическими соединениями, и интеграции сервиса «Синтелли» в цифровую поисковую платформу Роспатента.

Д. С. Золкин – автор дополнил статью информацией о работах по доработке сервиса «Синтелли» для задач экспертизы заявок на изобретения и интеграции сервиса «Синтелли» в цифровую поисковую платформу Роспатента.

Information about the authors

Anna E. Pinigina, Chemoinformatics Specialist, Syntelly LLC (Moscow, Vereiskaya str., 29, bld. 134); ORCID: <https://orcid.org/0009-0007-0704-5956>; pinigina.a.e@syntelly.com
Alina A. Muhamedzhanova, CEO, Syntelly LLC (Moscow, Vereiskaya str., 29, bld. 134); ORCID: <https://orcid.org/0009-0009-1983-6039>; admin@syntelly.com

Natalia G. Shimkina, PhD in Chemistry, Leading State Expert on Intellectual Property of the Chemistry and Pharmaceuticals Department, Federal Institute of Industrial Property (Moscow, Berezhkovskaya emb., 30, bld. 1); ORCID: <https://orcid.org/0009-0007-7260-5243>; SPIN: 8841-8908; natalia.shimkina@rupto.ru

Dmitry S. Zolkin, Head of IT division for Design of Information Retrieval System, Federal Institute of Industrial Property

(Moscow, Berezhkovskaya emb., 30, bld. 1); ORCID: <https://orcid.org/0009-0002-1641-4518>; SPIN: 3304-2957; db_dept@rupto.ru

Contributions of the authors

A. E. Pinigina – The author participated in the development of the "Syntelly" tool for extracting chemical structures from documents and was responsible for writing the manuscript text.

A. A. Muhamedzhanova – The author secured the financial support and resources necessary for the project implementation and was responsible for project administration.

N. G. Shimkina – The author contributed information regarding the potential use of the "Syntelly" service for the examination of patent applications related to chemical compounds and its integration into Rospatent's Digital Search Platform.

D. S. Zolkin – The author supplemented the article with information on the work to adapt the "Syntelly" service for the tasks of examining patent applications and its integration into Rospatent's Digital Search Platform.

Авторы заявляют об отсутствии конфликта интересов.
The authors declare no conflict of interests.

Поступила в редакцию (Received): 18.11.2025

Доработана после рецензирования (Revised): 24.11.2025

Принята к публикации (Accepted): 25.11.2025