

ИНТЕЛЛЕКТУАЛЬНЫЙ КОРПОРАТИВНЫЙ ПОИСК В КРУПНЕЙШЕЙ МЕТАЛЛУРГИЧЕСКОЙ КОМПАНИИ

В. А. Веденеев (ПАО «Северсталь»)

Описан проект по организации корпоративного поиска, основанный на технологиях оптического распознавания символов, компьютерного зрения и обработки естественного языка. Проект реализован в ПАО «Северсталь» с помощью решения ABBYY Intelligent Search. В результате внедрения удалось в 1,7 раза ускорить поиск документов и данных из разных источников предприятия.

Ключевые слова: корпоративный поиск, база знаний, хранилище файлов, оптическое распознавание символов, компьютерное зрение и обработка естественного языка.

ПАО «Северсталь» — это вертикально интегрированная горнодобывающая и металлургическая компания с самым высоким показателем рентабельности среди сталелитейных предприятий. С 2018 г. организация реализует программу цифровой трансформации. Ключевыми ее направлениями стали создание превосходного клиентского опыта, лидерство по снижению затрат в части производства основных продуктов по сравнению с конкурентами, а также получение доступа и возможность успешно интегрировать передовые технологии в новые бизнес-модели, которые появляются на рынке. В выборе ИТ-решений и проектов компания руководствуется именно этими тремя стратегическими приоритетами.

Проблема поиска информации в ПАО «Северсталь»

В 2020 г. одной из задач компании стало ускорение поиска информации о новых видах продукции и технологиях на основе данных и формализованных знаний сотрудников, сохраненных в корпоративных информационных системах. В ПАО «Северсталь» за 66 лет существования предприятия был накоплен обширный массив знаний и научно-технических разработок — примерно несколько десятков терабайт отсканированных и цифровых документов, которые хранятся в базе научно-технической информации. Это патенты, научные статьи, отчеты НИОКР и другие данные. Поиск документов, необходимых для работы, был затруднен, так как эта информация хранилась на разных ресурсах и в файловых системах, не все сотрудники помнили или знали, где находятся нужные им документы, а с уходом отдельных специалистов из компании данные могли потеряться, что влекло за собой дополнительные временные затраты и финансовые риски.

Сотрудникам Дирекции по техническому развитию и качеству, которые отвечают за разработку но-

вых сплавов и видов стали, была необходима новая единая интеллектуальная система корпоративного поиска по результатам научно-исследовательских работ и технологических проектов, которые были получены ранее, а также по нереализованным экспериментам, чтобы не повторять прошлых ошибок и не тратить средства организации на заведомо неудачные проекты. Например, если сотрудник работает над снижением коррозии металла в том или ином сплаве, ему нужно убедиться, что его задачу не пытались решить ранее в рамках других НИОКР. Зная эту информацию и изучив отчеты, специалисту не потребуется заново обращаться к институтам-партнерам, что позволит сэкономить ресурсы на реализацию проекта.

Нужно было найти решение, которое способно собрать нужные документы, обработать и структурировать данные, а затем вывести результаты для пользователей в понятном виде. Для сокращения времени на сбор и обработку данных и таким образом обеспечения более быстрого выпуска продукции на рынок ИТ-дирекция ПАО «Северсталь» приняла решение реализовать проект по организации корпоративного поиска, основанного на технологиях оптического распознавания символов (Optical Character Recognition, OCR), компьютерного зрения (Computer Vision, CV) и обработки естественного языка (Natural Language Processing, NLP). OCR был необходим для того, чтобы распознавать и извлекать текст из сканкопий бумажных документов, CV — для классификации документов по их внешнему виду, а NLP — для автоматического определения общего смысла и содержания документов.

Выбор поставщика решения

Специалисты «Северстали» проанализировали решения, доступные на российском рынке. Технология должна была отвечать следующим критериям:

- сквозной поиск по всем источникам данных, без необходимости переключаться между системами;
- удобный пользовательский интерфейс: поисковая строка, фильтры по дате создания документа, названию компании, суммам и т.д., поисковые подсказки на основе запросов, введенных ранее;
- возможность поиска данных по смыслу, а не только по ключевым словам, за счет применения NLP-технологий. Это помогло бы сотрудникам находить документы, содержащие синонимы и гипонимы. Пример: запросы «автосталь», «автомобильная сталь», «сталь для автомобилестроения»;
- информационная безопасность и разграничение прав доступа для разных категорий сотрудников, возможность выдавать разные результаты в поисковой выдаче в зависимости от разрешений и политик;
- простое масштабирование решения на другие бизнес-процессы и подразделения, возможность быстрой разработки коннекторов к другим информационным системам.

Данным требованиям удовлетворяла система компании АВВУ — мирового разработчика решений в области интеллектуальной обработки информации и анализа бизнес-процессов. Ранее АВВУ уже реализовала несколько подобных проектов на других промышленных предприятиях. В частности, в 2020 г. было успешно завершено крупное внедрение АВВУ Intelligent Search в НПО Энергомаш, компании, входящей в корпорацию Роскосмос, специализирующейся на разработке и производстве ракетных двигателей.

Пилот: поиск по 18 тыс. документов

Для проверки возможностей технологии в задачах предприятия дирекция по техническому развитию и качеству ПАО «Северсталь» и сотрудники компании АВВУ провели пилотный проект с подключением базы научно-технической информации. Поиск был организован по архиву технической документации, в котором содержалось около 18 тыс. документов. Для этого ИТ-дирекция «Северстали» организовала удаленный доступ специалистов АВВУ к своей инфраструктуре. Инженеры АВВУ установили АВВУ Intelligent Search на выделенный сервер. После загрузки документов в систему был построен полнотекстовый индекс архива. Из документов извлекли текст и мета-данные, после чего они стали доступными для поиска.

Одна из особенностей интеллектуального поиска заключается в том, что для системы нет разницы, работает ли она со структурированными документами, такими как счета-фактуры или товарные накладные, или с неструктурированными — патенты, чертежи, научно-исследовательские отчеты и т.д. В

неструктурированных документах система больше учитывает контекст, что отражается на результатах поисковой выдачи.

Интерфейс решения АВВУ включает поисковую строку и набор фильтров, по которым можно сужать результаты выдачи. Система интегрирована с Sharepoint, что позволяет быстро получать необходимые документы из файлового хранилища. В ходе пилота несколько сотрудников «Северстали» сравнивали время, затраченное на поиск документа стандартными инструментами Sharepoint¹ и с помощью АВВУ Intelligent Search. По их оценке, скорость выполнения этой задачи удалось увеличить в 1,7 раза. Релевантные результаты поиска оказались уже на первой странице выдачи, в то время как в Sharepoint пользователю было необходимо пролистать несколько страниц, чтобы найти нужную информацию. По итогам тестирования компания приняла решение подключить к поиску другие корпоративные источники.

Боевой проект

Специалисты «Северстали» назвали системы, по которым хотели бы организовать интеллектуальный поиск в первую очередь: файловое хранилище (сетевая папка), Sharepoint и корпоративная база знаний Confluence. На первом этапе проекта компания решила сосредоточиться именно на внутренних ресурсах.

При развертывании АВВУ Intelligent Search в операционной среде специалисты использовали кластерную конфигурацию из трех узлов. Это позволило повысить надежность и отказоустойчивость системы. Для доступа по единому адресу применили аппаратный балансировщик Citrix Netscaler. Работоспособность узла он определяет по ответу метода REST API, который не требует аутентификации.

Гибкие возможности настройки поискового решения позволили учесть различные пожелания пользователей по отображению результатов поиска. Например, в ходе настройки индексации содержимого портала Sharepoint возникла необходимость обеспечить фильтрацию найденных документов по названиям корневых каталогов, в которых они расположены и упростить отображение пути к документу, чтоб он выглядел привычным для пользователя. Это удалось реализовать с помощью встроенной в продукт поддержки регулярных выражений.

Для обеспечения полноты поиска к АВВУ Intelligent Search также подключили словарь синонимов с терминами, характерными для металлургической отрасли. При указании в запросе одного из подобных терминов он автоматически дополняется всеми возможными синонимами. Например,

¹ Microsoft SharePoint Foundation — бесплатное приложение к Windows Server; предоставляет базовую инфраструктуру для совместной работы: редактирование, хранение документов, контроль версий и т. д. Также он включает в себя такую функциональность, как «маршруты» движения документов (платформа для документооборота), списки заданий, напоминания, онлайн-дискуссии.

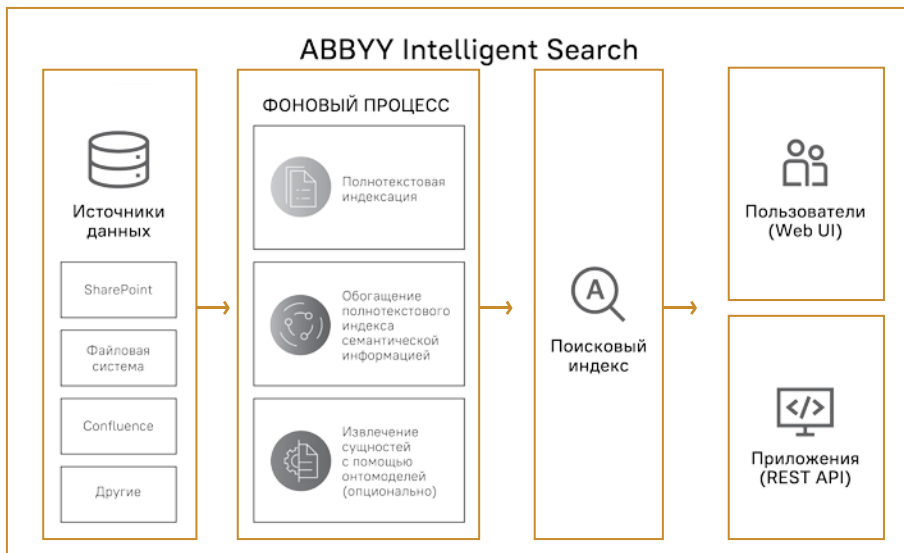


Рис. 1. Схема работы решения ABBYY в ПАО «Северсталь»

при вводе поискового запроса со словом «коррозия» также будут выдаваться результаты со словами «разрушение», «разъедание» и рядом других.

Одновременно с этим команда ABBYY начала разрабатывать интерфейс к базе знаний Confluence. Интеграция с этой системой входила в планы компании по развитию продукта, и ПАО «Северсталь» стала первым заказчиком, который смог протестировать все возможности работы ABBYY Intelligent Search в связке с порталом корпоративной документации на базе этой технологии. Первую версию интерфейса для Confluence удалось разработать за один месяц. В мае 2020 г. специалисты ABBYY продемонстрировали первый прототип решения (рис. 1).

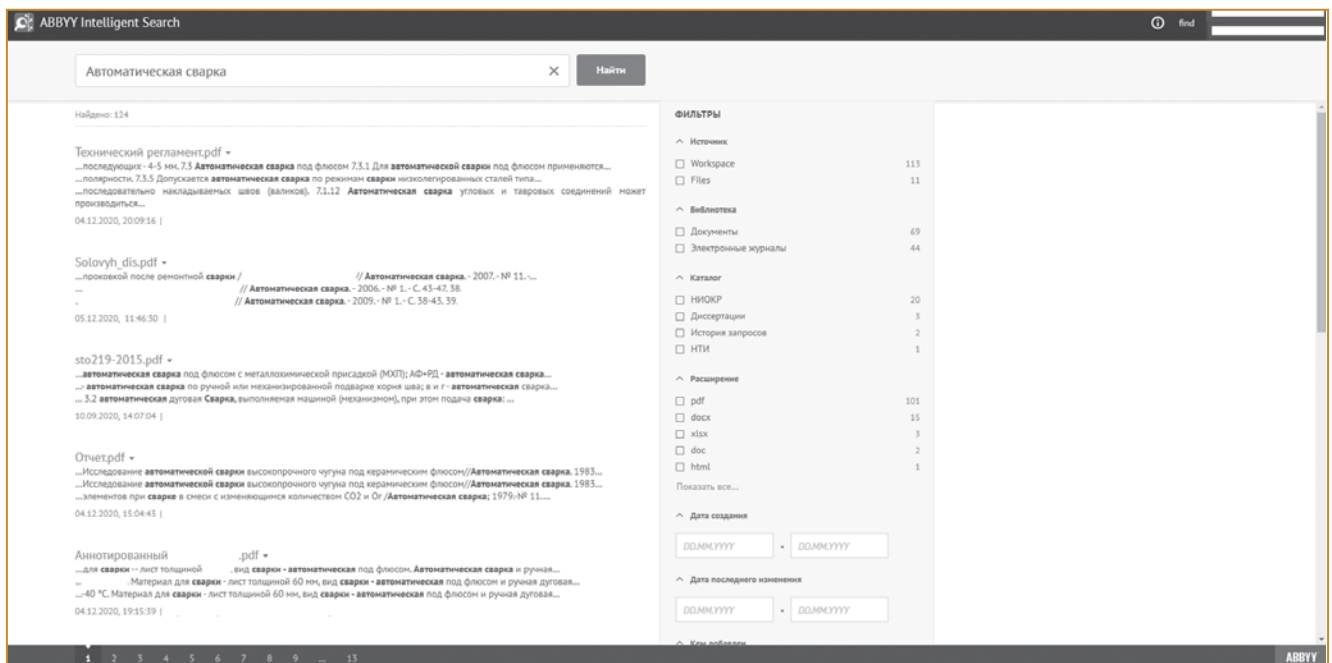
Результаты

На данный момент к системе ABBYY Intelligent Search подключены 300 тыс. файлов из ресурсов «Северстали». Системой пользуются технологи и проектные команды. Корпоративным поиском активно пользуются 150 специалистов компании из разных подразделений. Индекс активно наполняется.

Успешность проекта оценивали по следующим критериям: первый — это ускорение поиска и сбора информации. За счет того, что теперь на предприятии есть единая поисковая система, ответственный сотрудник может ввести запрос и увидеть весь список докумен-

тов, в которых содержатся интересующие его данные. Фрагменты, которые содержат слова запроса или похожие на него сущности, дополнительно выделяются жирным шрифтом. Для взаимодействия с системой не требуется специального обучения, так как по своему интерфейсу и принципу работы она похожа на популярные Internet-поисковики, которыми большинство людей пользуется каждый день. Сортировать информацию можно с помощью фильтров по источнику данных, формату файлов, дате, автору документа и другим параметрам.

Второй критерий — это снижение зависимости предприятия от человеческого фактора. За счет единой точки доступа информация не теряется, ее легко



Пример работы поиска в ПАО «Северсталь».

обнаружить, даже если автор документа находится в отпуске или покинул компанию. Кроме того, поиск реализован в соответствии с политиками информационной безопасности, что почти полностью исключает утечку данных.

Документы и данные в системе постоянно обновляются, что позволяет компании всегда работать с актуальной информацией. Система настроена следующим образом: есть сервер, на котором располагается программное решение АБВУУ. С помощью интерфейсов система подсоединяется к различным источникам. В качестве источников может использоваться Sharepoint, Confluence, файловый ресурс или любая другая система хранения данных, портал или база знаний. Одно из основных требований для источников — это наличие API (Application Programming Interface, интерфейс программирования приложений), с помощью которого посредством интерфейсов можно «подтягивать» и передавать информацию в систему поиска для последующей индексации. Для сотрудников ПАО «Северсталь» это работает так: человек создает или загружает документ на корпоративном устройстве, и через определенный промежуток времени система добавляет новый файл в поисковый индекс. Если в файл вносятся изменения, система автоматически заменяет его на обновленную версию документа.

Дальнейшие планы по развитию корпоративного поиска

В течение 2021 г. ПАО «Северсталь» планирует продолжить развитие системы корпоративного поиска и постепенно подключать другие внутренние ресурсы. В первую очередь к АБВУУ Intelligent Search будет полностью подключена база знаний Confluence. Сейчас специалисты протестировали работу решения лишь на небольшой части документов из этого источника, но в планах сделать поиск полноценным инструментом для работы с этой системой. Во-вторых, у компании есть планы улучшить интерфейс решения: добавить больше поисковых подсказок и возможностей для интерпретации данных.

Существует задача масштабировать корпоративный поиск на остальную часть предприятия, инновационные центры и экспертные сообщества. В перспективе будут проиндексированы около 3 млн. документов: патенты, отчеты, НИОКР и др. Будет расширено и число пользователей: если сейчас активно применяют в своей работе систему примерно 150 сотрудников, то в течение нескольких лет ее охват может составить до 5 тыс. специалистов. Это позволит компании исключить дублирование проектов, сократить время на экспертизу идей, ускорить запуск новых продуктов и внедрение инноваций.

Веденев Виктор Алексеевич — ведущий эксперт по цифровым технологиям

Дирекции по техническому развитию и качеству ПАО «Северсталь».

Контактный телефон +7 (8202) 53-09-00.

[Http://www.severstal.com](http://www.severstal.com)

М.Видео-Эльдорадо приняла на работу нейросеть

ПАО «М.Видео» внедрила корпоративную нейросеть для автоматизации коммуникации с 28 тыс. сотрудников розницы. Искусственный интеллект через чат-ботов в Telegram и Viber помогает продавцам магазинов в решении большинства повседневных задач — от оформления отпуска до клиентской поддержки, сокращая затраты ресурсов и времени на обработку обращений пользователей в четыре раза. Сейчас нейросеть М.Видео-Эльдорадо обрабатывает около одного миллиона обращений к базам знаний в месяц.

Нейросеть М.Видео-Эльдорадо уже обрабатывает порядка 50% запросов автоматически, 30% — закрываются через правильный ответ оператором из вариантов, предложенных нейросетью. И только порядка 20% требуют погружения оператора, изучения проблематики, поиска документации и подготовки ответа. Таким образом, компания тратит примерно в четыре раза меньше ресурсов и времени на подготовку ответов, увеличивая скорость реакции на основную часть поступающих вопросов.

Самообучающаяся нейросеть на базе решения AutoFAQ интегрирована в корпоративные чат-боты, которыми сотрудники розничных магазинов пользуются для выдачи on-line заказов. В результате, компания создала

инструмент двусторонней коммуникации и оперативного ответа на вопросы о взаимодействии с ИТ-системами, кадровых политиках и процедурах обслуживания клиентов от более чем 28 тыс. человек, не увеличивая штат клиентской поддержки.

Чат-бот в Telegram и Viber с подключённой нейросетью получает запрос от пользователя и моментально запускает диалоговую сессию, анализирует вопрос и ищет ответ в загруженной базе знаний. Если вероятность верного ответа выше 80% (параметр настраиваемый), система отвечает, если ниже — переводит запрос на операторов поддержки. Вопросы, на которые нейросеть пока не в состоянии ответить самостоятельно, попадают в интерфейс оператора, при этом система предлагает ему на выбор минимум три наиболее подходящих варианта ответа. Сотрудник поддержки выбирает подходящий либо отвечает самостоятельно по-другому, за счёт чего происходит дальнейшее обучение нейросети. Нейросеть поддержки сотрудников М.Видео-Эльдорадо уже оперирует примерно 5000 статей, время дообучения новой статьёй всей нейросети составляет 10...20 с. Это позволяет максимально оперативно вносить изменения и развивать систему.

[Http://mvideo.ru](http://mvideo.ru)