

Внедрение технологий PROMT в ФИПС

Синописис

Документ описывает опыт внедрения продуктов компании PROMT для автоматизации процессов перевода в ФИПС, цели внедрения, особенности итогового решения и его достоинства.

О ФИПС

Федеральное государственное бюджетное учреждение «Федеральный институт промышленной собственности» (далее - ФИПС) Федеральной службы по интеллектуальной собственности решает следующие основные задачи:

- Прием и экспертиза заявок на объекты патентного права, в т.ч. международных заявок и проведение международного поиска по международным заявкам, рассмотрение предусмотренных законодательством Российской Федерации документов, представленных на государственную регистрацию договоров о распоряжении исключительным правом, и сделок, предусматривающих использование единой технологии за пределами Российской Федерации, выпуск официальных бюллетеней о зарегистрированных объектах патентных прав, поданных заявках и выданных по ним патентах, и других изданий, рассмотрение заявлений, ходатайств, касающихся продления срока действия исключительного права, рассмотрение возражений, касающихся решений по результатам экспертизы заявок на объекты патентного права, признания недействительным предоставления или досрочного прекращения действия правовой охраны, подготовка проектов решений.
- Прием и экспертиза заявок на государственную регистрацию средств индивидуализации, выполнение работ по международным заявкам на товарный знак, рассмотрение заявлений о признании товарного знака или обозначения общеизвестным в Российской Федерации, рассмотрение предусмотренных законодательством Российской Федерации документов, представленных на государственную регистрацию договоров о распоряжении исключительным правом, средствах индивидуализации, выпуск официальных бюллетеней о зарегистрированных средствах индивидуализации и других изданий, рассмотрение заявлений, ходатайств, касающихся продления срока действия исключительного права на средства индивидуализации, рассмотрение возражений и заявлений, касающихся решений по результатам экспертизы заявок на средства индивидуализации, признания недействительным предоставления или досрочного прекращения действия правовой охраны средств индивидуализации, подготовка проектов решений.
- Прием и проверка заявок на государственную регистрацию программы для электронных вычислительных машин, базы данных и топологии интегральных микросхем, выпуск официальных бюллетеней о зарегистрированных свидетельствах и других изданий, рассмотрение предусмотренных законодательством Российской Федерации документов, представленных на государственную регистрацию договоров о распоряжении исключительным правом, и сделок, предусматривающих использование единой технологии за пределами Российской Федерации, подготовка проектов решений.

- Библиотечное, библиографическое, справочно-информационное, научно-методическое обслуживание на базе государственного патентного фонда и единой системы автоматизированных банков данных.

О PROMT

PROMT - ведущий российский разработчик лингвистических IT-решений для корпораций и частных пользователей с 25-ти летним опытом работы на мировом рынке. Компания предлагает системы перевода документов для 20 языков (английский, русский, болгарский, испанский, итальянский, казахский, китайский в двух вариантах, латышский, немецкий, польский, португальский, турецкий, украинский, французский, японский, арабский, финский, венгерский, корейский). Клиентами компании являются более 15 000 крупных компаний в сферах IT, промышленности, финансов, торговли, науки, образования, переводческой деятельности, в государственных структурах по всему миру. Среди них - всемирно известные компании PayPal, TripAdvisor, Amadeus, NASA, «Лаборатория Касперского», «Норникель», «Лукойл», «Полиметалл Инжиниринг», и многие другие.

PROMT предоставляет весь спектр лингвистических решений: системы автоматического перевода, словари и системы анализа текстов, мобильные приложения для всех платформ. Компания занимает лидирующее положение в области разработки решений по автоматическому переводу в России и мире.

Цели проекта

Патентные эксперты ФИПС при анализе патентов осуществляют поиск информации в более чем 110 млн. патентов, которые хранятся как в собственных базах данных, так и в базах данных международных патентных ведомств или патентных ведомств других стран.

Задачей эксперта является определение патентоспособности поступившего на анализ патента. Соответственно, необходимо найти все имеющиеся в базах патенты, относящиеся к сути анализируемого патента. Так как патенты в базах данных хранятся на различных языках, то для повышения эффективности и точности работы эксперта было необходимо их обеспечить возможностью автоматического перевода патентной информации.

Целью проекта было создание и развертывание встроенной в информационную среду ФИПС платформы перевода, решающей следующие бизнес-задачи:

- предоставление эксперту возможности получать мгновенный качественный перевод патентного документа или его фрагмента на русский язык;
- автоматический перевод запроса эксперта к поисковой системе PatSearch для возможности поиска релевантной информации в базах патентов на различных языках;
- возможность создания экспертами терминологических баз данных для повышения качества перевода патентной информации.

Технические требования к платформе перевода

ФИПС сформулировал следующие требования к платформе перевода:

- Платформа должна представлять собой клиент-серверное решение с легким клиентом, обеспечивающим доступ к переводу из всех наиболее распространенных браузеров для обеспечения независимости доступности функции перевода от операционной системы, версии используемого клиентского ПО;

- Платформа перевода должна быть легко масштабируемой для обеспечения быстрого перевода документов, одновременно поступающих с нескольких сотен рабочих мест экспертов;
- Платформа должна предоставлять программный доступ к функции перевода, чтобы обеспечить использование функции перевода как backend процесса в работе информационных систем, функционирующих в ФИПС, в частности в разработанной в ФИПС системе патентного поиска Patsearch;
- Платформа должна обеспечивать перевод с английского, немецкого, французского, испанского, итальянского языков на русский и с русского языка на перечисленные;
- Должна быть обеспечена возможность перевода не только текстовой информации, но и документов, хранящихся в графических форматах, что характерно для работы экспертов ФИПС, так как огромное количество информации хранятся в базах данных в виде предварительно сканированных документов в форматах PDF, JPEG;
- Персонализация и авторизация доступа к платформе перевода. С одной стороны для части пользователей должна быть обеспечена возможность персонализированного доступа к платформе перевода, при этом необходимые личные настройки должны сохраняться между сеансами работы. С другой стороны должна оставаться возможность анонимного доступа. При этом доступ должен осуществляться только из внутренней сети ФИПС, но из различных иногда не связанных доверительными отношениями доменов;
- Должна быть проведена инициальная настройка платформы перевода на особенности терминологии и структуры, характерные для патентных текстов и документов; должны быть предоставлены средства и технологии для повышения качества перевода патентной документации;
- Кроме непосредственно перевода платформа должна выполнять и роль терминологической базы данных, в том числе обеспечивать ввод новой терминологии специалистами ФИПС;
- Платформа перевода должна содержать средства администрирования для возможности эффективного управления пользователями и качеством перевода.

Продукт PROMT Translation Server как основа для платформы перевода и необходимые доработки и настройки

В качестве основы для реализации платформы перевода ФИПС был взято семейство продуктов PROMT Translation Server как отвечающее основным требованиям к платформе перевода.

С **системно-архитектурной** точки зрения продукт является клиент-серверным с возможностью масштабирования, дает возможность получения перевода как с помощью пользовательского интерфейса (в частности, через браузеры), так и через программный интерфейс, реализованный как кросс-платформенный web-сервис. Также в продукте предусмотрена аутентификация и авторизация пользователей, правда в недостаточном для платформы перевода объеме.

Для обеспечения необходимого качества перевода и настройки на предметную область в продукт было включено более 100 словарей, содержащих специализированную терминологию по различным областям знания, что дает возможность сразу же получить качество, достаточное для перевода документов из многих патентных областей.

Взяв в качестве стартовой точки для построения платформы перевода PROMT Translation Server, были сделаны следующие дополнительные разработки для реализации требований к платформе перевода в полном объеме:

- в конвейер обработки документов был добавлен модуль распознавания графических файлов, который конвертировал графический образ текстового документа в текст, который уже подавался на перевод;
- была модифицирована первоначальная система аутентификации, в результате чего в платформе перевода были реализованы два типа аутентификации:
 - анонимная, при которой пользователю не надо вводить никаких данных, доступ к платформе обеспечивается за счет того, что пользователь является членом локальной сети организации;
 - персонифицированная, на основе саморегистрации в базе данных пользователей платформы перевода;

при этом запоминаются такие настройки перевода как направление, в котором осуществлялся перевод, выбранные лингвистические настройки перевода (так называемый, профиль перевода);

- на сервере платформы перевода был развернут дополнительный модуль "Электронный словарь", позволяющий:
 - осуществлять сквозной поиск терминов и переводов по всем словарям платформы перевода;
 - получать грамматические признаки терминов/переводов, примеры склонений/спряжений, спеллинг терминов и другую грамматическую информацию;
 - получать список терминов или словосочетаний, содержащих искомую подстроку; т.е. пользователь может видеть все термины, куда как составная часть входит то или иное слово или словосочетание;
- конвертировать терминологические базы, созданные экспертами ФИПС, в формат, необходимый для Электронного словаря PROMT и предоставлять этот контент всем пользователям платформы перевода;
- для решения задачи инициальной настройки перевода полного корпуса патентной документации была применена уникальная гибридная технология перевода, с помощью которой был построен профиль *Патенты* для англо-русского перевода;;
- чтобы обеспечить дальнейшее повышение качества перевода, на рабочих местах экспертов, ответственных за ввод новых терминов в систему, были развернуты специальные десктопные версии, которые позволяют создавать и редактировать словари на рабочих местах с последующей публикацией их на сервере; также предусмотрены средства для периодического перестроения гибридного профиля перевода Патенты.

Гибридная технология перевода

Гибридная технология PROMT представляет собой композицию традиционного метода построения перевода на основе грамматического описания естественного языка (rule based machine translation, RBMT) и пост-редактирования полученного машинного перевода с помощью технологии статистического перевода. Такой подход позволяет, сохранив основные достоинства RBMT (смысловую точность перевода, грамматическую и терминологическую корректность), сделать автоматический перевод более "гладким" и по стилю более соответствующим переводу, сделанному человеком. Материалом для создания модуля статистического пост-редактирования была база русских патентов и их переводов на английский язык, предоставленная ФИПС, с параметрами:

Количество патентов	511 620
Объем патентов на английском языке (в	65 413 474

словах)	
Объем патентов на русском языке (в словах)	63 962 350

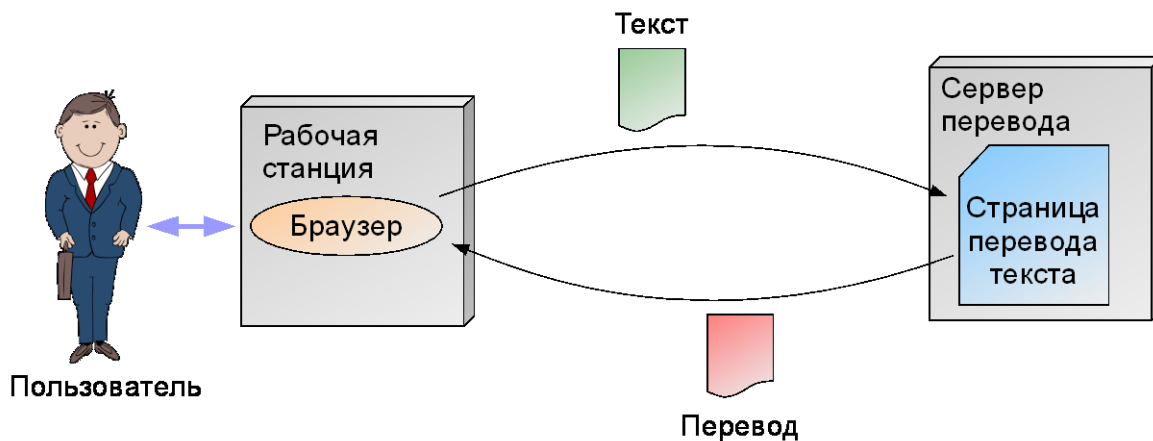
Результат перевода оценивался по методике BLEU Score и достиг значения более 26 пунктов, что для англо-русского перевода является очень хорошим результатом.

Гибридная технология по мере увеличения базы патентов, используемой для обучения статистической модели перевода, позволяет эффективно увеличивать качество перевода без значительных затрат.

Общая схема процессов, выполняющихся с помощью платформы перевода ФИПС

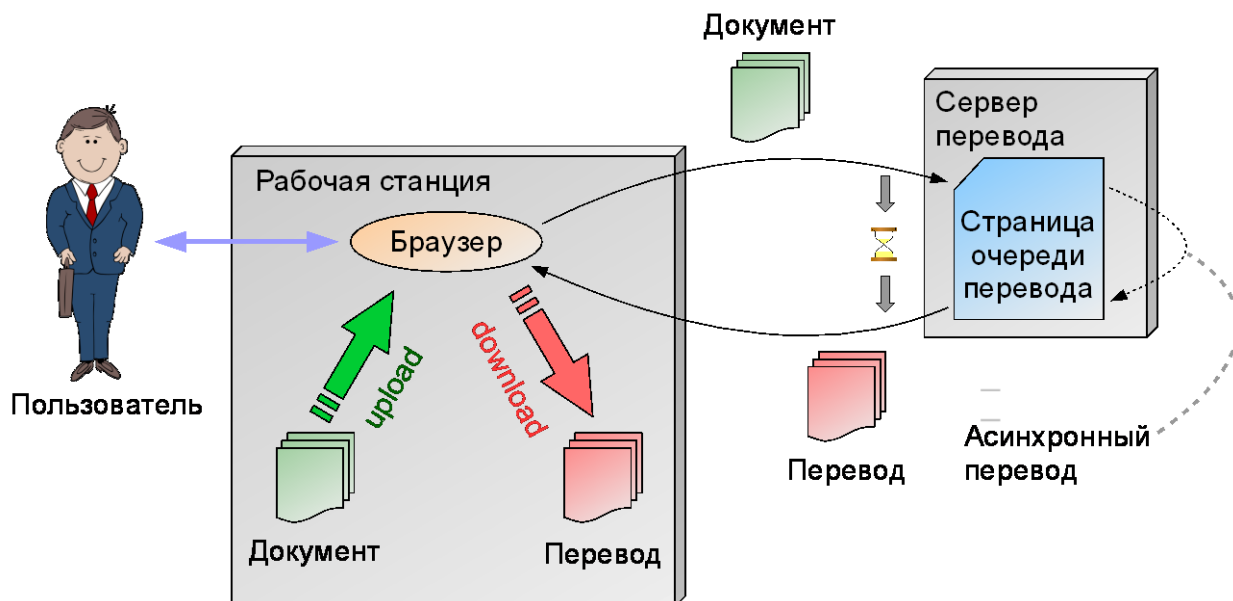
Платформа перевода, развернутая в ФИПС, дает разнообразные возможности применения автоматического перевода в работе экспертов ФИПС. Ниже описаны основные способы использования, поддерживаемые платформой перевода.

Перевод фрагмента текста



В качестве клиента платформы перевода эксперт использует любой браузер, помещая в специальное окно на странице интересующий его текст и посылая его на сервер платформы перевода, дополнительно указывая направление и профиль перевода, чтобы получить перевод максимально возможного качества. Сервер платформы перевода возвращает переведенный текст в другое окно браузера на той же странице, что и окно с исходным текстом.

Перевод документа



Эксперт может послать на перевод и документ целиком (как во всех офисных форматах, так и представляющий собой графический образ в JPEG, PDF). Так как документ может быть достаточно большим, то чтобы не заставлять эксперта ожидать окончания перевода, этот процесс организован как асинхронный процесс следующим образом:

- эксперт в специальном окне на странице ставит документ в очередь переводов, при этом указывая все необходимые условия перевода (направление перевода, профиль перевода), после чего может переключиться на другие задачи;
- далее эксперт периодически проверяет статус перевода (отображается в окне очереди переводов);
- когда перевод интересующего его документа получает статус "Перевод завершен", эксперт скачивает переведенный документ на свой компьютер и открывает его в соответствующем формате документа редакторе (следует отметить, что в случае перевода графического образа документа, перевод формируется в формате RTF, позволяющем отредактировать переведенный документ).

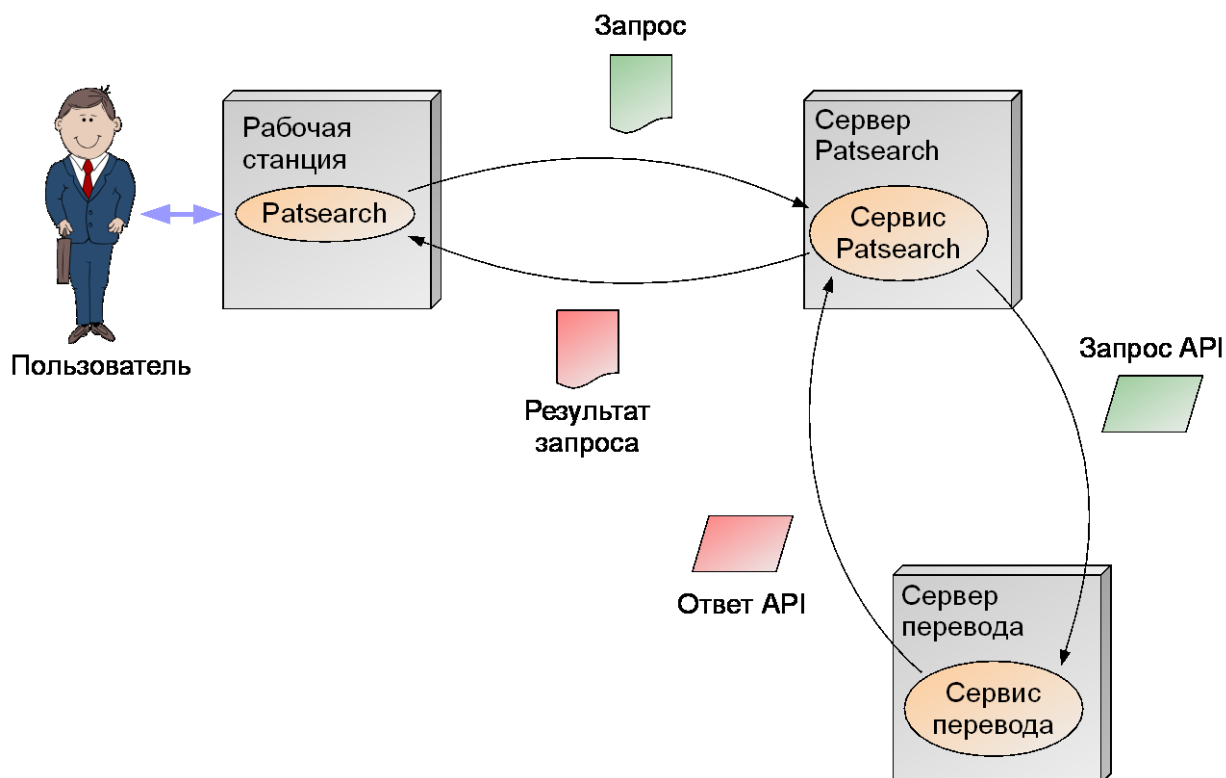
Поиск в терминологической базе данных

Эксперту предоставляется возможность по полному или частичному написанию слова или словосочетания получить исчерпывающую информацию обо всех терминах, в которые входит интересующее слово/словосочетание. При этом в строке поиска слово/словосочетание может быть введено в произвольной форме. Формируется информация по всем переводам термина, находящимся в базе, с указанием из какого словаря взят перевод термина, с полной грамматической информацией для термина и перевода.

Терминологическая база данных может расширяться за счет данных, вводимых в систему экспертами ФИПС.

Отображение информации может быть настроено таким образом, чтобы пользователь видел все переводы или только те, которые были введены в базу пользователями ФИПС.

Перевод запроса в системе Patsearch



Функция перевода используется также и при работе с другими информационными системами, используемыми в ФИПС. В частности она встроена в поисковую систему Patsearch.

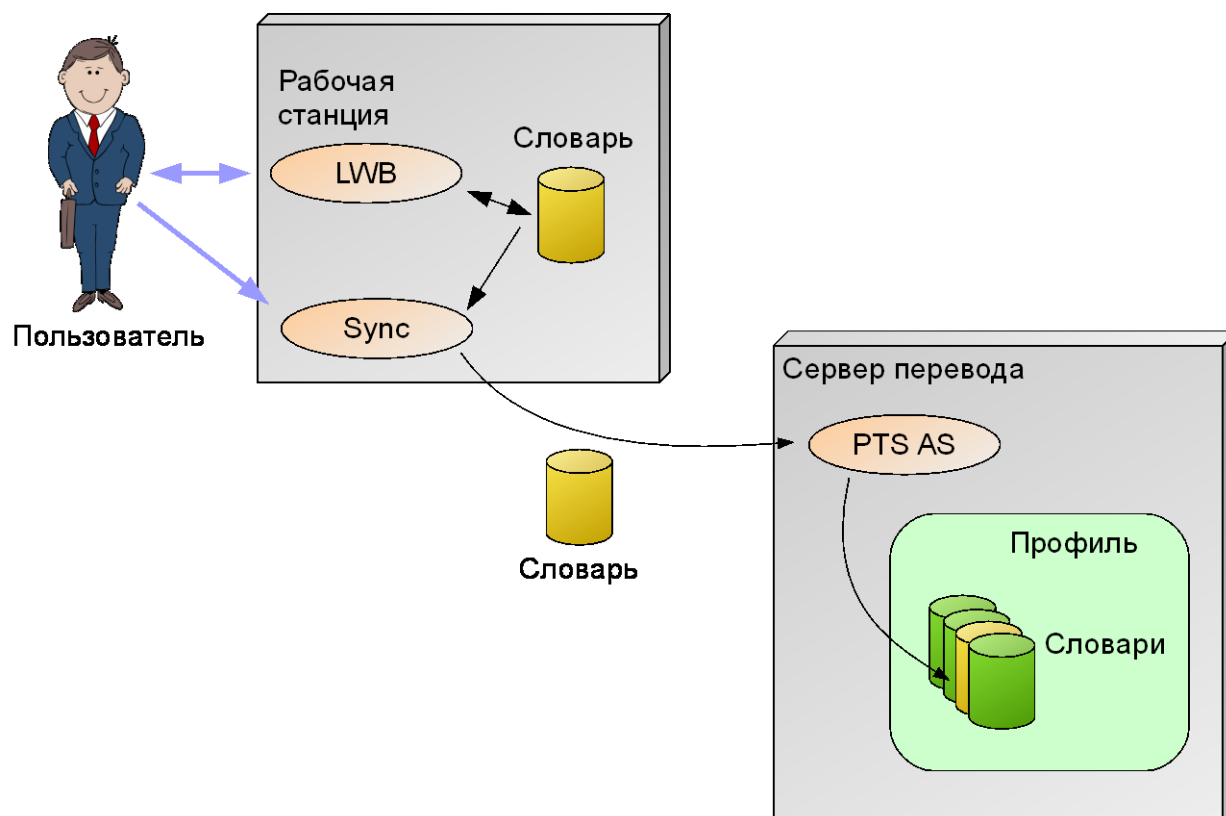
Способ использования в этом процессе - следующий:

- эксперт средствами, предоставляемыми Patsearch, формирует запрос на русском языке и указывает дополнительно, что поиск должен быть осуществлен и в базе патентов, содержащих документы на другом языке;
- В системе Patsearch происходит обращение к платформе переводов, в результате чего формируется запрос на целевом языке;
- Далее поиск производится с помощью переведенного запроса в базе документов на целевом языке.

Особенностями такого способа использования является то, что:

- взаимодействие Patsearch и PROMT Translation Server реализовано на API, предоставляемого сервером перевода. API в свою очередь реализовано как платформо-независимый web-сервис;
- PROMT разработал специальный модуль перевода, учитывающий специфику формирования запроса в системе Patsearch, а именно то, что большинство запросов формируются как логическая функция над патентными терминами.

Создание собственных словарей и их публикация

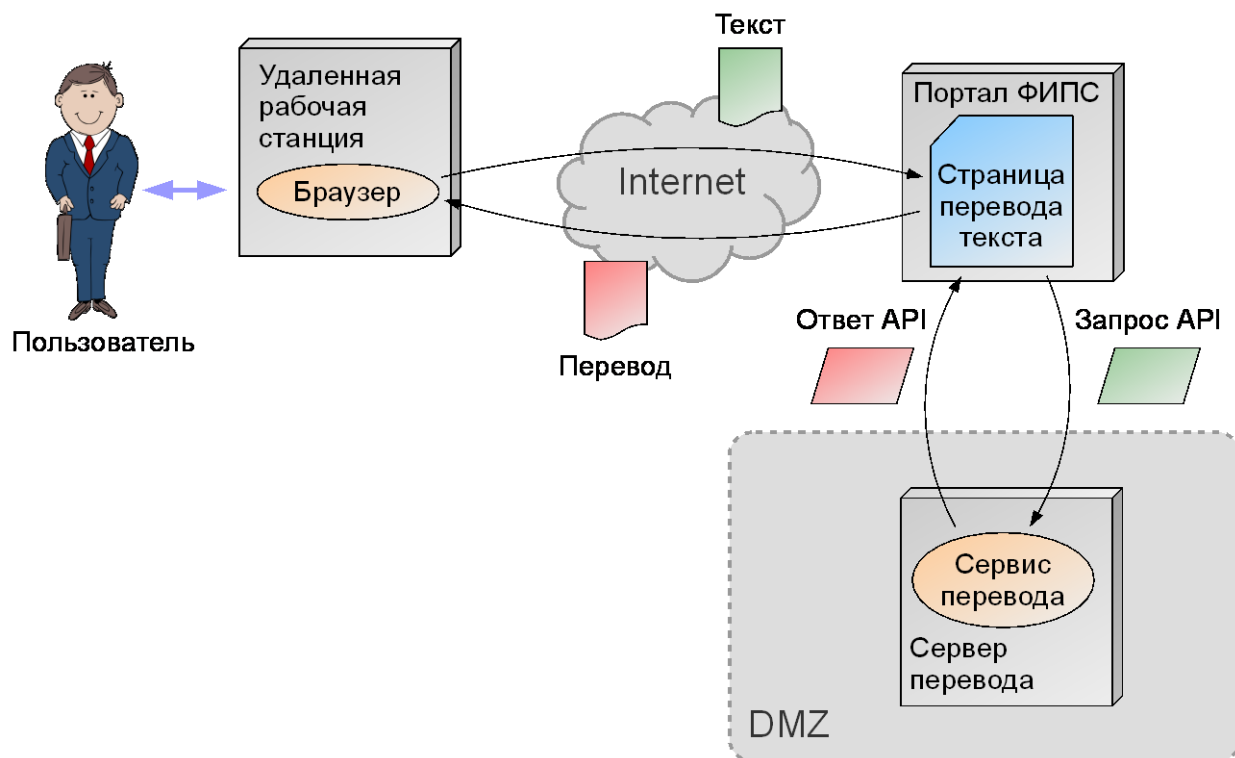


Важной особенностью развернутой в ФИПС платформы перевода, является ее "обучаемость". В частности это происходит за счет того, что эксперты имеют возможность дополнять имеющиеся в платформе перевода словари новыми терминами и их переводами. В результате растет качество перевода, генерируемое платформой перевода.

Процесс ввода новой терминологии организован следующим образом:

- у эксперта на рабочей станции устанавливается специальная программа PROMT Linguistic Workbench (LWB);
- с помощью нее эксперт создает новый терминологический словарь или дополняет терминами и их переводами уже существующий словарь; этот процесс происходит локально на рабочей станции эксперта;
- для того, чтобы сделать введенную информацию доступной всем при переводе, необходимо выполнить публикацию словаря на сервере платформы перевода;
- после принятия решения о том, в каких профилях перевода будет использоваться словарь, эксперт с помощью программы Sync передает словарь на сервер;
- после публикации этот словарь начинает использоваться в процессе перевода с теми профилями, в которые он включен.

Перевод на портале ФИПС в рамках программы Patent Prosecution Highway



Платформа перевода используется не только экспертами ФИПС, но и пользователями международного пилотного проекта по программе Patent Prosecution Highway (см. <http://www.rupto.ru/activities/inter/bicoop/pph>). Эта программа предназначена для того, чтобы зарубежные ведомства в рамках проекта могли работать с российскими полностью англоязычными документами.

Взаимодействие между порталом и сервером платформы перевода осуществляется с помощью кода встроенной в страницы портала кнопки, при нажатии на которую содержимое страницы передается на сервер платформы перевода, переводится и перевод страницы отображается пользователю.

Выводы

Разработанная платформа перевода успешно используется в ФИПС в течение 15 лет, пройдя путь от разрозненной закупки десктопных продуктов до клиент-серверной платформы, встроенной в информационную среду организации, постоянно совершенствуясь и давая новые и удобные инструменты для работы с патентной информацией на основных европейских языках. В планах расширения функциональности - подключение китайского и японского языков к платформе перевода.

Использование автоматизированного перевода значительно повысило эффективность работы экспертов, сделало возможным оперативное предоставление патентной информации для пользователей портала ФИПС на родном языке.

Отличительной особенностью платформы перевода, развернутой в ФИПС, является реализация всего спектра возможностей, заложенных в продуктах серии PROMT Translation Server:

- перевод текста;
- перевод документов с сохранением форматирования;

- перевод документов, сохраненных в виде графического образа;
- терминологическая база и поиск по ней;
- настройка качества перевода;
- специальный перевод запросов, учитывающий структуру и правила построения запроса в поисковой системе;
- перевод с использованием пользовательского интерфейса и через API;
- гибкая авторизация (анонимная и персонифицированная).

Также необходимо отметить активное участие в развитии качества перевода экспертов ФИПС, которые создали несколько десятков словарей по конкретным областям знаний.

ПРОМТ, А. Серебряков, Руководитель отдела корпоративных решений

«Для нас проект с ФИПС необыкновенно важен и интересен по нескольким причинам. Во-первых, ФИПС был одной из организаций, которая, будучи заинтересована в большом количестве переводов с целью ускорения процессов обработки и поиска информации, осознала эффективность развертывания централизованного решения и для нужд конечного пользователя и для встраивания в другие информационные системы организации, во-вторых, для решения задач, поставленных перед нами как поставщиком решения, пришлось использовать весь потенциал разработок компании, пересмотреть некоторые подходы в архитектуре системы, и в-третьих, благодаря тому, что ФИПС предоставил большие массивы параллельных данных (оригинальных патентов и их переводов), удалось применить технологию гибридного перевода, доказав ее эффективность для быстрой настройки "под пользователя" и получения качественного перевода в тех случаях, когда у организации накопилось много документов и их переводов. Такой комплексный подход позволяет, по нашему мнению, извлечь максимальную пользу от применения технологий автоматического перевод»

Денис Владимирович Виноградов, директор ФИПС:

«Практика нашего многолетнего сотрудничества с компанией ПРОМТ подтвердила правильность нашего первоначального решения о выборе партнера для решения задач машинного перевода текстов патентных документов. Благодаря высокому качеству перевода ПРОМТ мы даже сейчас, когда широко доступны бесплатные средства машинного перевода, не ставим вопрос об отказе от услуг ПРОМТ. Компания ПРОМТ всегда оперативно откликалась на наши пожелания, иногда специфические для технологии работы в ФИПС. Это касалось и создания гибридной системы перевода патентных документов, и специальной функциональности по переводу запросов на поиск в англоязычных базах данных, и автоматического подключения словарей с учетом тематики документов. Более 600 экспертов ФИПС активно используют машинный перевод от компании ПРОМТ в своей ежедневной работе».