

Технологии компании ПРОМТ

Технологии и программные продукты компании ПРОМТ

На сегодняшний день компания ПРОМТ обладает колоссальным технологическим багажом, что позволяет разрабатывать системы перевода с различными функциональными возможностями.

Высококвалифицированная команда специалистов и накопленный опыт позволяют компании в кратчайшие сроки разрабатывать модули машинного перевода (МП), обеспечивающие высокое качество перевода практически для любых языковых пар. ПРОМТ предлагает пользователям, как готовые программные решения для бизнеса или использования в домашних условиях, так и решения для интеграции в корпоративные системы.

Более 13 лет все исследования и проекты компании ПРОМТ фокусировались на решении глобальной задачи по созданию средств, обеспечивающих перевод с одного естественного языка на другой. Уникальные технологии построения систем перевода и оригинальные алгоритмы работы с текстами на естественных языках стали тем базисом, на котором были созданы все программные продукты компании и который обеспечил возможность разработки широкого спектра решений для автоматизированного перевода с одного языка на другой. Программные продукты компании ПРОМТ одинаково полезны как для решения бизнес-задач, так и для домашнего использования. В последнее время особое внимание компания ПРОМТ уделяет созданию специальных средств и технологий для профессиональных переводчиков.

Передовые разработки компании ПРОМТ

В основу фундамента передовой технологии перевода PROMT были заложены следующие инновационные разработки:

- 1. Алгоритмическая основа «translation engines»**, обеспечивающая высокое качество перевода и построенная на формализме расширенной сети переходов или ATN сетей (Augmented Transition Network).
- 2. Принципиально новые и обеспечивающие легкую адаптацию системы перевода.** Благодаря документозависимым настройкам системы перевода предоставляют пользователю широкий, интуитивно понятный набор инструментов для настройки системы и обеспечивают использование корректной терминологии при переводе, настройку синтаксических и семантических правил, учитывающих структуру и формат текста в документ.
- 3. Интеграция двух «переводческих» технологий в одном продукте –** технологии машинного перевода МП и технологии Translation Memory. Синергетический эффект от взаимного применения двух технологий позволяет обеспечить практически 100%-ное качество перевода при работе с повторяющимися текстами.

4. Модульность структуры систем перевода позволяет быстро разрабатывать любые виды программ машинного перевода, как «легких» приложений, предназначенных для быстрого перевода текстов, так и высокопрофессиональных, клиент-серверных и интранет решений, предназначенных для корпоративного использования и применения при «промышленном» переводе.

5. Поддержка технологии единого кода (single code) для критически важных лингвистических модулей системы. Это позволило реализовать решения для ряда программных платформ:

- Операционные системы семейства Microsoft Windows плюс полная интеграция с приложениями Microsoft Office 2000/XP/2003,
- Macintosh,
- Linux,
- WindowsCE/Windows Mobile.

Лингвистические технологии

Компания ПРОМТ практикует научный подход к каждой своей разработке и в полной мере использует накопленный опыт при создании новых систем. Создание новых алгоритмов перевода и новой методологии в математической лингвистике стало постоянным процессом, что говорит об инновационности разработок компании ПРОМТ.

Все модули перевода разработаны на основе технологии единого исходного кода, что обеспечивает независимость от операционной системы. Модуль перевода представляет собой систему типа «transfer», основанную на описании грамматики в виде расширенных сетей переходов - ATN (Augmented Transition Network Grammar). Система работает с использованием информации, заданной в словарных статьях слов и выражений в словарях системы - части лингвистической базы данных. Основные компоненты модуля перевода:

- Препроцессор (языково-независимая часть),
- Нормализация текста (языково-независимая часть),
- Лексический анализ (зависит от входного языка),
- Семантико-синтаксический разбор (зависит от входного языка),
- Transfer (зависит от входного и выходного языка),
- Синтез (зависит от выходного языка).

Компоненты лингвистической базы данных используются на всех стадиях анализа и синтеза текста:

- На стадии предварительной обработки - ассоциированная память (если задана в системе);
- На стадии лексического анализа - древесные структуры, описывающие входную морфологию, и словари;
- На стадии семантико-синтаксического анализа - расширенные сети переходов;
- На стадии синтеза - выходные морфологические таблицы.

Система словарей, используемых при переводе, имеет трехуровневую иерархию - генеральный словарь, специализированные словари, а также словари, разработанные пользователем самостоятельно. Такая трехуровневая структура позволяет легко настраивать систему на работу в различных предметных областях.

Генеральный словарь для одной языковой пары содержит от 40 до 200 тысяч словарных статей, которые в свою очередь содержат структурированное описание различной лингвистической информации, необходимой системе для работы сложнейших алгоритмов анализа и синтеза текста. Словари по тематикам содержат специфические слова и выражения, характерные для предметной области, их объем может меняться от 5 до 50 тысяч словарных статей. Например, для англо-русской и русско-английской систем разработаны специализированные словари, охватывающие более 50 различных тематик.

Архитектура системы МП

Система машинного перевода PROMT имеет трехслойную архитектуру:

- **Первый, "базовый" слой** – это translation engines, модули обеспечивающие перевод текстовой строки с одного языка на другой, и модули, обеспечивающие редактирование словарной статьи.
- **Translation kernel** – средний уровень в архитектуре нашего ПО. Он представляет собой набор модулей, реализующих управление всеми объектами системы, синхронизацию данных, обработку форматов документов и представление исходных документов во внутреннем формате системы. Этот слой имеет мощное API, позволяющее легко оперировать со сложными агрегированными объектами системы и унифицирует доступ к этим объектам для приложений, в которые встраиваются функции перевода документов. Разработанная нами архитектура позволяет создавать как новые программы с точки зрения функциональности и пользовательского интерфейса, так и встраивать в уже существующую архитектуру новые направления перевода.
- **Третий слой** – это широкий набор приложений для различных моделей использования, предоставляющий пользователям системы разнообразные функции, которые могут быть полезными в процессе обработки текстов на иностранных языках.

Это позволяет нам оперативно реагировать на запросы рынка, быстро выходить на новые рынки, занимать новые продуктовые ниши, разрабатывать продукты для новых операционных систем или выпускать новые направления перевода. Например, по заказу крупнейшего мирового поставщика ERP/CRM решений была разработана специализированная система МП для внутренних нужд этой компании. Несмотря на отсутствие у компании ПРОМТ каких-либо разработок для заказанных направлений перевода, эта задача была выполнена всего за несколько месяцев.

Четко выстроенный технологический процесс разработки новых направлений перевода позволяет компании ПРОМТ регулярно выпускать новинки. Так, в течение 2004 года были выпущены пять направлений перевода - французско-испанско-французская, французско-английско-французская и англо-португальско-английская, включая бразильский диалект.

Интеграция

Технологические преимущества позволяют компании полностью соответствовать как требованиям рынка, так и специфическим требованиям клиентов по совместимости, возможностям настройки, интеграции и масштабируемости, которые являются для нас основными задачами. Например, решение для перевода корпоративных Интернет-сайтов PROMT Internet Translation Server (PITS) включает SDK, которые позволяют пользователям интегрировать PITS с внешними приложениями.

Абсолютно инновационный подход использовался при создании продукта PROMT Front-End Translation Server – в этом серверном решении впервые в мире были интегрированы серверное решение TRADOS Server и система PROMT Translation Server.

Краткий обзор функциональных возможностей систем перевода

В настоящее время системы ПРОМТ выполняют перевод для следующих языков:

- английский
- испанский
- немецкий
- французский
- итальянский
- португальский
- русский.

Список поддерживаемых языков постоянно расширяется.

Архитектура систем ПРОМТ использует преимущества технологий COM и .NET и предоставляет широкие возможности использования одних и тех же модулей перевода для решения совершенно разных задач в разных программных приложениях. Одни и те же модули перевода используются и на Интернет порталах через PROMT Internet Translation Server и на домашних компьютерах. Еще одной очень важной особенностью программного обеспечения, предоставляемого компанией ПРОМТ, является то, что оно создает интегрированную среду для переводчика, соединяя процесс автоматического перевода со справочными словарными системами, с системами класса Translation Memory, со средствами автоматизации терминологической работы и т.д.

В программном обеспечении PROMT также имеются собственные дополнительные сервисные средства для анализа текстов, средства для создания и поддержки словарей и постредактирования результатов перевода. Перевод текста может осуществляться во всех офисных приложениях, а именно Microsoft Word, Microsoft Excel, Microsoft Outlook, Microsoft PowerPoint, Microsoft FrontPage. Основные компоненты системы:

- PROMT Translation Engine (PTE),
- Среда перевода (лингвистический редактор PROMT),
- Средства для редактирования словарей и баз АП,
- Средства для профессиональной работы со словарями,
- Средства для встраивания функций перевода в офисные приложения (Word, Excel, PowerPoint, FrontPage, Outlook, Internet Explorer, Acrobat),
- Средства терминологической работы,
- Средства для интеграции с системами Translation Memory (например, TRADOS),
- Средства для пакетного перевода, перевода содержимого буфера обмена и быстрого перевода,
- Сервисные средства и инструменты (Резервное копирование, Инсталляция и деинсталляция специализированных словарей).

Дополнительные компоненты системы для корпоративных решений:

- Средства администрирования для сетевого сервера переводов,
- Средства администрирования и веб-интерфейс для серверов перевода Internet/Intranet,
- Пользовательские плагины для PROMT Internet/Intranet Translation Server.

Все приложения и инструменты PROMT поддерживают работу с ядром перевода ("Translation Kernel") через API. Это позволяет всем приложениям PROMT быть языково-независимыми и значительно ускоряет разработку новых приложений. PROMT Translation Engine (PTE) включает:

- Ядро перевода с API,
- Модули перевода для всех направлений перевода,
- Лингвистические базы данных для всех направлений перевода.

Для оценки качества перевода PROMT предлагает воспользоваться услугами бесплатного онлайн-перевода на своих переводческих порталах: www.translate.ru и www.online-translator.com, посещаемость которых превышает 50.000 человек в сутки.