

MSDN Library - это обширный ресурс технической информации, созданный компанией Microsoft в помощь специалистам, разрабатывающим программные продукты с использованием приложений и средств разработки Microsoft. До последнего времени Microsoft обеспечивал разработчиков технической документацией в основном на английском языке, хотя количество иностранных программистов, использующих MSDN Library в работе, не уступает количеству англоязычных пользователей. Российское отделение компании решило позаботиться об интересах русскоязычных разработчиков приложений для операционных систем Microsoft и сделать информацию электронной библиотеки доступной для них на родном языке. Для решения сложной задачи были привлечены специалисты компании ПРОМТ, которая является мировым лидером в области разработки технологий машинного перевода. В результате удалось создать уникальный сервис перевода, настроенный для конкретного информационного ресурса. В ходе работы специалисты ПРОМТ нашли новое применение ранее созданным технологиям машинного перевода, а также разработали ряд уникальных решений.

Одной из причин, помогающих компании Microsoft на протяжении многих лет удерживать абсолютное лидерство на рынке компьютерной техники и программного обеспечения, является открытость информации о любых ее технологиях и решениях. Эта открытость не абсолютна, и тут в качестве противоположного примера можно привести пресловутую секретность кодов ОС Windows, но это, скорее, исключение из правил. А они таковы: компания Microsoft заинтересована в привлечении лучших сил по разработке программных продуктов с использованием технологий и приложений Microsoft.

Понимая, что при работе с программным обеспечением могут возникать вопросы, ответы на которые не найти в "Справке", специалисты Microsoft создали обширную электронную библиотеку статей и материалов, посвященных исключительно технической стороне дела - MSDN (Microsoft Developer Network) Library и выложили ее в интернете по адресу <http://www.msdn.com/library>. Разработчики быстро оценили ее полезность, и "электронное справочное бюро" стало популярным, несмотря на то, что информация была представлена только на английском языке.

Здесь надо отметить, что компания Microsoft ранее никогда не предпринимала никаких усилий по переводу на другие языки хотя бы части материалов, содержащихся в MSDN Library, - с одной стороны, не желая тратить силы на постоянное обновление огромного объема материала, с другой - полагая, что нет программиста, который не владел бы технической компьютерной терминологией, которая априори англоязычна. При этом едва ли не для большинства профессиональных разработчиков продуктов на базе приложений MS английский язык не является родным. В частности, далеко не всем русскоязычным программистам понятна сложная, часто изменяющаяся техническая информация MSDN Library.

Если сейчас в нашей стране и приходится говорить о дефиците, то это, несомненно, нехватка грамотных бумажных технических словарей. Хорошо известно, что в общих словарях компьютерные термины тоже редкие гости. В результате, российские специалисты, не владеющие английским языком в совершенстве, не имели полноценного доступа к информации, содержащейся на сайте <http://www.msdn.com/library>.

Но проблемой заинтересовались сотрудники российского представительства компании Microsoft. В связи со сложностью тематики и большим объемом данных, задача требовала уникальных решений. Только одна российская компания обладала технологией онлайн-перевода, с помощью которой можно было достичь нужных результатов. Именно поэтому российское отделение Microsoft обратилось с предложением о сотрудничестве в компанию ПРОМТ, ведущему мировому разработчику технологий автоматизированного перевода.

Было принято решение организовать сервис перевода в рамках проекта GotDotNet (<http://www.gotdotnet.ru/>) – сайта, специально созданного компанией Microsoft для российских разработчиков. На этом сайте выкладываются различные справочные и учебные материалы по продуктам Microsoft, а также работают форумы для обсуждения животрепещущих вопросов относительно работы с продуктами Microsoft.

Хотя ПРОМТ давно выпускает системы перевода и специализированные словари по компьютерной тематике, было принято решение разработать специальный сервис для англо-русского перевода документации из MSDN Library. До специалистов ПРОМТ никто таких задач не выполнял, поэтому конкретика решений вдохновлялась азартом первооткрывателей.

Первый этап работы заключался в организации собственно службы перевода на основе .NET сервиса. Из множества разделов библиотеки для перевода была выбрана документация по продукту Microsoft .NET Framework SDK. Здесь надо отметить, что реализовать сразу сервис перевода всего контента библиотеки MSDN было невозможно в силу колоссального объема и разнородности информации, поэтому для первого этапа был выбран раздел по наиболее современной тематике – технологии Microsoft .NET.

Когда программисты ПРОМТ выстроили каркас будущего сервиса перевода, на первый план вышла задача по настройке и оптимизации сервиса. Поэтому был применен профессиональный инструмент для анализа текстов и извлечения терминологии - PROMT Terminology Manager. С помощью этого инструмента был проведен статистический анализ терминологии в документации MSDN Library, подготовлен список терминологических кандидатов, на основе которого был составлен словарь терминов. Данный терминологический словарь стал основным рабочим инструментом сервиса перевода.

Как известно, в технических статьях, как правило, одно и то же слово в заголовке, в самом тексте и в списке программных кодов может означать совершенно разные вещи. Следовательно, метод перевода зависит от структуры HTML-страницы. Техническая оптимизация заключалась в том, чтобы “научить” специальную программу различать структуру элементов страницы и сообщать о ней сервису перевода.

Во время лингвистической настройки специалисты ПРОМТ впервые использовали в масштабе целого сервиса базу данных Translation Memory (TM), отличающуюся от обычных словарей тем, что в нее заносятся не отдельные словари и их переводы, а предложения и даже целые фрагменты текста, называемые сегментами. Таким образом, в базе TM содержатся исходные сегменты текста и их перевод.

В базе TM, созданной для проекта по переводу библиотеки текстов MSDN Library содержится более 2000 сегментов - готовых предложений и фрагментов текста, характерных для технической документации. Если в требующей перевода статье содержится отрывок, на 100% совпадающий с одним из тех, что хранятся в Translation Memory, то сервис берет перевод текста из базы данных, не переводя его заново. Это обеспечивает гарантированное качество перевода стандартных фраз и выражений.

«Последняя часть настройки сервиса была самой сложной, - вспоминает Александр Серебряков», технический директор компании ПРОМТ. «Без преувеличения можно сказать, что на этом этапе трудилась команда лучших лингвистов в нашей стране. Мы действительно гордимся проделанной работой, потому что, во-первых, никто ничего подобного до нас не делал, а во-вторых, были продемонстрированы новые возможности онлайн-перевода.»

Сервис перевода документации .NET Framework SDK получил массу положительных отзывов со стороны профессиональных программистов, многие из которых были удивлены высоким качеством машинного перевода. Опираясь на успешно внедренную технологию, специалисты ПРОМТ реализовали также перевод документации по системе Microsoft SQL Server, предназначенной для создания и анализа электронных баз данных. Качественный уровень перевода, как и в случае с разделом .NET Framework SDK, вырос за счет настройки сервиса с учетом терминологии и особенностей форматирования, характерных для документов Microsoft SQL Server.

Дмитрий Старостин, представитель российского отделения компании Microsoft: «Нам приятно, что именно в российском отделении Microsoft родилась идея перевода MSDN Library с английского языка на родной язык пользователей. Безусловно, без участия ПРОМТ все осталось бы на уровне мечты. Хочу подчеркнуть, что воплотить ее в реальность могли только инновационные решения, и то, что компании ПРОМТ удалось их разработать, говорит о большом интеллектуальном потенциале и высоком уровне профессионализма российских специалистов. Интерес к результатам эксперимента со стороны других иностранных отделений Microsoft, доказывает его значимость. Если мы внимательно относимся к нуждам пользователей продуктов Microsoft в своих странах, заботимся о том, чтобы наши специалисты имели такой же свободный доступ к информации, как и их англоязычные коллеги, то в ответ повышается лояльность к имени компании и ее продуктам. Собственно говоря, наши ожидания уже оправдываются. Мы на хостинге отмечаем рост числа посетителей сайта <http://www.gotdotnet.ru> именно после внедрения сервиса перевода ПРОМТ.

Взросла активность пользователей из профессиональной среды. Им это интересно, они обмениваются мнениями о новой услуге на форуме, причем, большинство отзывов позитивны, что является редкостью для экспериментального начинания”.

Напоследок приведем цитату из форума : “Конечно, на выходе машины пока еще идет довольно сырой текст, но - сайт и сервис отличный. Видно, что разработчики технологии машинного перевода вкладывают много сил и старания в доработку технологии! То есть, этот сайт демонстрирует преимущества а) технологий вообще, б) технологий Microsoft. Даже самые закоренелые скептики, походив по сайту, придут к выводу, что машинный перевод заслуживает внимания.”

Продвижение новых технологий и решений - это общая черта лидеров в своих отраслях, какими являются компании Microsoft и ПРОМТ. Проект перевода MSDN Library демонстрирует, как поиск решения конкретных задач стимулирует появление уникальных методов. Их удачное применение, помимо достижения конечной цели, всегда приносит дополнительные результаты. Так, работая над частной проблемой перевода одного информационного ресурса, компания ПРОМТ подняла планку качества машинного перевода в целом.

Вставки:

О заказчике:

Компания Microsoft основана в 1975 году студентами Гарвардского университета Биллом Гейтсом и Полом Алленом. Разработанные компанией операционные системы MS-DOS и Windows, пакет офисных приложений Microsoft Office и сотни других программных продуктов стали основой рабочего стола большинства компьютеров мира и принесли ее создателям миллиардные доходы. Сегодня Microsoft - одна из самых крупных транснациональных корпораций, ежегодно повышающая объемы своих продаж на десятки процентов (так, объем продаж Microsoft в 2003 финансовом году в мире вырос на 13%, а в странах СНГ - на 55%). Более подробная информация на сайте компании по адресу www.microsoft.com/rus.

О партнере:

Компания ПРОМТ (www.promt.ru) — мировой лидер в области прикладной лингвистики и разработки технологий автоматизированного перевода для европейских языков. Компания сочетает научную деятельность с производственной. Уникальные технологии, которым нет аналогов в России, создаются на основе последних достижений лингвистической науки и современных программных и технических средств.

ПРОМТ ставит своей целью создание программных средств, позволяющих свободно общаться на разных языках. Наш лозунг: «Мы делаем мир понятным!».

Более подробная информация о компании ПРОМТ на сайте: www.promt.ru. Качество перевода можно оценить на сайте бесплатного онлайн-переводчика компании ПРОМТ: www.translate.ru.

Ситуация:

Российское отделение компании Microsoft решило перевести информацию, содержащуюся в MSDN Library - главном ресурсе технической информации о программном обеспечении Microsoft, - на русский язык для отечественных специалистов, разрабатывающих программные продукты на основе приложений и технологий Microsoft. В связи со сложностью тематики и большим объемом данных, задача требовала уникальных решений. Для их поиска и воплощения к сотрудничеству была приглашена компания ПРОМТ, лидер в области машинного перевода.

Решение:

Используя собственную, не имеющую аналогов, технологию онлайн-перевода, специалистам ПРОМТ удалось создать первый сервис перевода, настроенный для конкретного информационного ресурса. Был проведен анализ терминологии одного из разделов MSDN Library, учтены особенности структуры технического текста и впервые применена в таком масштабе технология ассоциированной памяти, благодаря которой отпала необходимость каждый раз заново переводить распространенные в технической литературе обороты и фразы. Инновационность метода перевода в сочетании с уникальными решениями сделала этот совместный проект успешным и подняла планку качества машинного перевода для всех участников рынка.