



Опыт успешного внедрения МП для перевода технической документации к продуктам Лаборатории Касперского

Лариса Поклад

Старший инженер по локализации
в Лаборатории Касперского

Юлия Епифанцева

Директор по развитию PROMT

МП при переводе документации Лаборатории Касперского



Лаборатория Касперского

- 18 лет на рынке разработки ПО
- Более 2 500 сотрудников
- Персональные и корпоративные продукты
- 35 территориальных офисов (НQ в Москве)
- Более 300 миллионов пользователей по всему миру
- Продукты локализуются на 37 языков

МП при переводе документации Лаборатории Касперского

Локализация в ЛК – команда в цифрах



- 50 штатных сотрудников
- 20 аутсорс-сотрудников
- 6 стран, 7 офисов
- 50 проектов в год
- 37 языков локализации
- 15 вендоров

МП для Лаборатории Касперского

Задачи проекта



Снижение стоимости
локализации



Увеличение скорости
локализации

МП для Лаборатории Касперского

Требования



- Достойное качество машинного перевода, пригодное для последующей редактуры
- Возможность продолжать настройку системы на стороне разработчика по результатам постредактирования
- Возможность поддерживать несколько профилей согласно проектной специфике
- Оперативная поддержка со стороны разработчика
- Интеграция со средствами автоматизированного перевода (SDL Trados Studio)
- Возможность держать сервер на стороне разработчика
- Возможность выбирать языковые пары и их количество

МП для Лаборатории Касперского

Исполнитель проекта



PROMT

- С 1991 одна из ведущих в мире лингвистических IT компаний
- 17+ языков, 64+ языковых пар
- Технологии: PROMT Rule-based MT, SMT, Hybrid MT
- 10 000+ крупных компаний-клиентов, миллионы частных пользователей и 100 000 000+ запросов на сервисе Translate.RU
- Высокие оценки экспертов на ежегодном конкурсе машинных переводчиков под эгидой ACL

МП для Лаборатории Касперского

Предложенное решение

PROMT Cloud
на платформе
PROMT Translation Server

- Клиент-серверное решение с поддержкой любых технологий перевода: Rule-based MT, SMT, Hybrid MT
- Возможность обучения системы на материалах заказчика
- Технологии: PROMT Rule-based MT, SMT, Hybrid MT
- Результаты настройки в виде профилей перевода (без ограничения их количества для одной языковой пары)
- Интеграция с SDL Trados Studio

МП для Лаборатории Касперского

Средства обучения PROMT

Специализированные словари с терминологией из определенных предметных областей

Пользовательские словари для терминологии заказчика

Традиционная технология Translation Memory (только для 100%-ных совпадений)

Правила перевода для стилистических и синтаксических предпочтений

Статистическая модель (на основе параллельных текстов от 100 000 сегментов)

МП для Лаборатории Касперского

Пилотный проект

Цели и сроки:

- Оценить качество МП после обучения на материалах заказчика и эффективность постредактирования для англо-итальянского, англо-португальского и англо-испанского
- Срок— **6-8 недель**

Входные данные:

- ТМ в формате tmx (~ 7000 сегментов для каждой языковой пары)
- Проект SDL Trados Studio в формате sdxliff (от 5 000 до 10 000 сегментов в зависимости от языковой пары)
- Экспорт из базы Passolo в формате csv (от 6 000 до 8 000 сегментов в зависимости от языковой пары)
- Небольшие по объему глоссарии для каждой языковой пары

МП для Лаборатории Касперского

Выполненные работы

- Конвертация данных из баз ТМ, проекта SDL Trados Studio и из базы Passolo в формат tmx и подключение к профилю перевода Kaspersky как баз Translation Memory
- Подключение специализированных словарей PROMT с базовой ИТ-терминологией к профилю после тестирования на материалах заказчика
- По результатам анализа текстов заказчика подбор значений правил перевода (ПП - опция для выбора предпочтительного варианта синтеза на выходной язык)
- Добавление терминов из глоссариев в пользовательские словари
- Построение частотных списков на всем множестве текстов для выделения новых терминологических кандидатов
- Для португальского языка словарные данные приведены к новой орфографии

МП для Лаборатории Касперского

Результаты пилотного проекта

Результаты анализа контента и обучения системы

- Профили перевода для англо-итальянского, англо-португальского и англо-испанского со всеми словарями, базами ТМ и другими настройками
- Профили подключены к PROMT Translation Server 10 на площадке PROMT
- Количественные характеристики настройки

	EP	EI	ES
Словарь Kaspersky	146 статей	298 статей	438 статьи
Переработанный словарь по IT	4065 статей	6623 статьи	8931 статьи
BLEU Score (до настройки)	16,7	10,7	25,3
Editing Distance (до настройки)	66	69,2	52,9
BLEU Score (после настройки)	26,4	22,5	41,8
Editing Distance(после настройки)	54,3	56,2	39,6

- Возможность дальнейшей работы по улучшению качества перевода с гибридной настройкой при наличии нужного объема данных

МП для Лаборатории Касперского

Внедрение технологии МТ

Использование в локализации пользовательской документации
для флагманского продукта KAV/KIS/KTS/KSOS



Общий объем - 120 тыс. слов



Из них 5-10% на перевод при каждой итерации



МТ используется для профессиональных целей, то есть
не «как есть», а с последующей редактурой

МП для Лаборатории Касперского

Этапы внедрения

1. Внедрение web-приложения на проекте KAV/KIS/KTS/KSOS 2015:

- Необходимость дополнительных действий при подготовке пакета на перевод
- Параллельная доработка профиля перевода по результатам постредактирования
- **Экономия до 15%**

2. Внедрение плагина для SDL Trados Studio на проекте KAV/KIS/KTS 2016:

- Удобство и простота использования в связи с интеграцией с SDL Trados Studio 2014
- Доработка профиля перевода по результатам постредактирования
- **Экономия до 15%**

МП для Лаборатории Касперского

Языковые пары

- English (US) – Italian (Italy)
- English (US) – Portuguese (Portugal)
- English (US) – Spanish (Spain) – в процессе обучения

МП для Лаборатории Касперского

Процесс локализации пользовательской документации

Создание проекта в SDL Trados Studio с применением проектных ТМ и плагина ProMT



Подготовка пакета на перевод и постредактирование



Локализация на стороне LSP: обработка fuzzy-сегментов и постредактура АТ-сегментов, редактура, QA checks и подготовка отчета о текущей итерации постредактуры



QA checks на стороне KL, пополнение ТМ и отправка отчета о текущей итерации постредактуры

МП для Лаборатории Касперского

Приоритетная поддержка

Техническая поддержка

- Консультации специалистов PROMT по техническим вопросам в режиме 24/7

Лингвистическая поддержка

- Анализ поступающего фидбэка от постредакторов: добавление в словарь новых терминов, улучшение перевода конкретных конструкций
- Консультации по возможному улучшению качества перевода в рамках работающей версии
- Предложение по дальнейшему развитию и настройке системы на материалах заказчика

МП для Лаборатории Касперского

ИТОГИ И ВЫВОДЫ



- Успешная интеграция с SDL Trados Studio и простота в использовании
- Использование и редактирование специализированных словарей и профилей перевода



Повышение
производительности
переводчиков



Снижение стоимости
локализации



Качественное
выполнение больших
проектов в сжатые сроки

МП для Лаборатории Касперского

Задачи и перспективы

- Выработать систему оценки качества внутри ЛК
- Выработать систему оценки стоимости постредактуры в зависимости от качества
- Внедрить новые языковые пары:
 - English (US) – Spanish (Spain)
 - English (US) – Spanish LATAM
 - English (US) – Portuguese Brazilian
 - English (US) – Turkish
 - English (US) – Simplified Chinese
- Внедрить использование МТ на других проектах
- Продолжить настройку системы перевода: создание и пополнение словаря UI-элементов, создание разных лингвистических профилей для разных продуктов



Спасибо за внимание!

Юлия Епифанцева

Директор по развитию PROMT

Julia.Epiphantseva@promt.ru

Лариса Поклад

Старший инженер по локализации
в Лаборатории Касперского

Larisa.Poklad@kaspersky.com