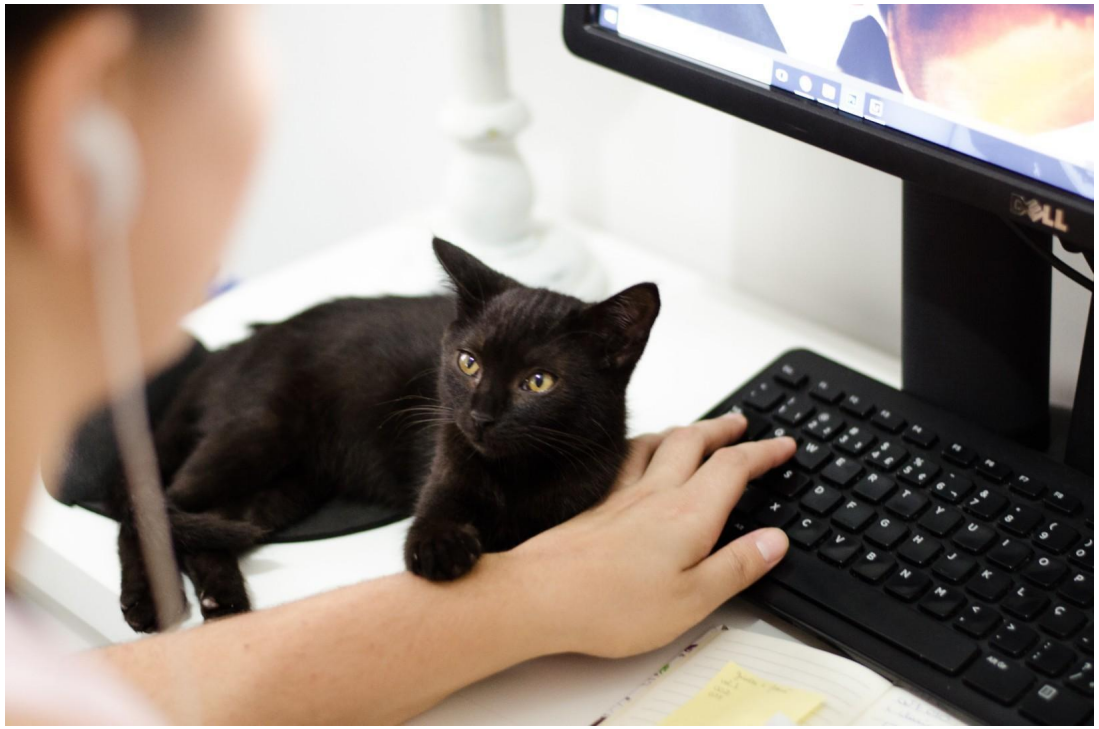
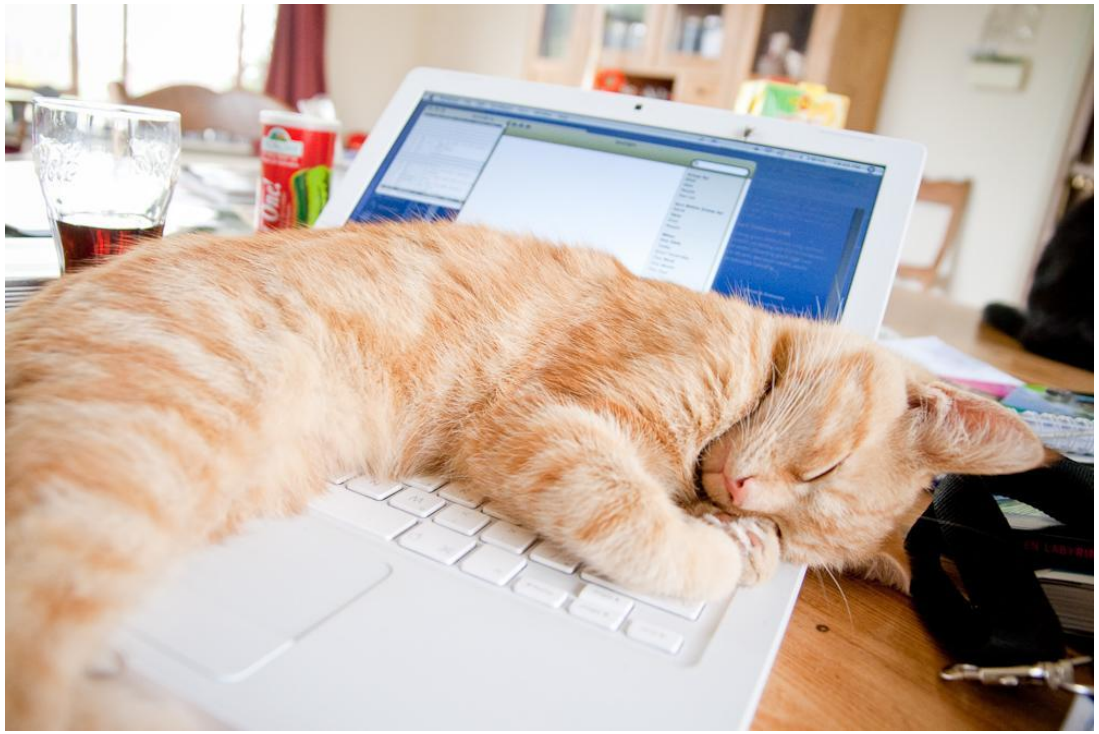
Современные информационные технологии для АВТОМАТИЗАЦИИ ПЕРЕВОДА

Can cats help in translation?





Can cats help in translation?

Can cats help in translation?

А вы знакомы с «кошками»?

Перевод - процесс и продукт

Этапы ПРОЦЕССА перевода – общее представление

1) **восприятие** текста (устного или письменного) = ПОНИМАНИЕ, т.е. декодирование языковых знаков – в смысл!

(+Промежуточный, невидимый этап: внутренняя работа мозга по переключению с одного языка на другой и обратно)

2) **Порождение** / **создание** текста с аналогичным содержанием – но на другом языке, т.е. перекодирование смысла – в абсолютно другие знаки, знаки другого, переводящего языка (ПЯ)

Группы программ для автоматизации перевода

1) Электронные словари – обеспечивают **быстрый поиск:**

- *значений лексических единиц (ЛЕ) и фразеологизмов (ФЕ) в любом языке,*
- *соответствий для ЛЕ в ИЯ (исходном языке) и ПЯ (переводящем языке / языке перевода)*

(т.е. позволяют **понять текст на ИЯ**);

2) Корпусы текстов и корпусные менеджеры – дают **примеры реального (нормативного и узуального) использования** ЛЕ в ПЯ

(т.е. позволяют **создавать текст, отвечающий нормам ПЯ**)

3) Средства распознавания текстов / речи

- **Optical Character Recognition** (**OCR**) –технология, которая позволяет преобразовывать различные типы документов (отсканированные документы, PDF-файлы или фото с цифровой камеры, а также рукописные тексты) - в редактируемые форматы с возможностью поиска.
- **Automatic Speech Recognition (ASR)** — компьютерная технология автоматического распознавания речи (чаще –для других ситуаций, но иногда и для перевода, в частности – для создания субтитров к видеозаписи и их последующего перевода).

4) Программы машинного перевода (Machine translation) –

без участия переводчика: выполняют **полный перевод** текста

(НО: требуется пост-редактирование человеком).

(PROMT, Google-Translate, Яндекс-переводчик и др.)

5) Программы автоматизированного перевода (Computer–assisted / aided translation = CAT-Tools) – при участи переводчика:

помогают человеку в переводе, выполняют за него только **некоторые рутинные операции –**

НА ОСНОВЕ ТЕХНОЛОГИИ **Translation Memory (TM)**

ТМ не выполняет перевод за переводчика, а является мощным инструментом для сокращения затрат при переводе повторяющихся фрагментов текстов.

НЕ ПУТАТЬ!

Предпосылки разработки программ автоматизированного перевода:

1) **Бурное развитие технического прогресса**

→ увеличение **числа** технических устройств, машин и пр.

→ увеличение **объема** документации (европейский самолет Airbus, российские вертолеты - десятки тысяч страниц!)

+ большинство производителей активно осваивают региональные рынки → **локализация** продукции, в том числе перевод описания продукта на местный язык – одно из обязательных условий для выхода на новый рынок

→ = **увеличение объема документов, требующих перевода**.

2) производителям приходится **регулярно обновлять** свои продукты, т.к. новые версии продаются лучше!

НО: часто новая модель – это лишь рестайлинг (слегка измененная) предыдущая модель.

<u>В результате:</u> документация по каждому из таких продуктов часто **на 70-90% совпадает** с той, что была у предыдущей версии.

→ **высокая повторяемость текстов:**

В основе CAT-Tools – 3 основные технологии:

- **Память** (или база) **переводов** (Translation Memory, TM);
- **Управление терминологией** (Terminology Management);
- **Управление проектами** (Project Management, Translation Management System, TMS).

Технология TRANSLATION MEMORY (TM) -

ключевая для переводчика технической документации.

База Translation Memory - лингвистическая база данных, в которой переводы сохраняются в рамках определенной структуры по мере их создания пользователем-переводчиком.

Обычный состав базы TM - исходные сегменты и эквивалентные им целевые сегменты (как правило, за сегмент принимается предложение).

КАК это работает?

При обработке нового текста, поступившего на перевод, система сравнивает каждое его предложение с сохраненными в базе - и извлекает соответствующие совпадения.

В результате не требуется многократно переводить одно и то же предложение, но можно повторно использовать ранее переведенные фрагменты текста.

Управление терминологией (Terminology Management)

Точная терминология - критически важное требование к качественному переводу.

Для обеспечения точности и согласованности терминологии- разные решения:

- единые централизованные хранилища терминов = универсальные электронные словари - поиск непосредственно из основного редактора системы
- средства управления терминологией.

Управление проектами (Project Management, Translation Management System, TMS)

Современный процесс перевода - <u>большое количество задач,</u> связанных с подготовкой файлов и проектов.

Часто - выполняются вручную и занимают массу времени.

Для оптимизации данных задач - **встроенные** в современные системы САТ **модули для управления** переводами:

- автоматизация настройки проектов,
- выбор требуемых баз переводов и терминологии,
- проверка качества,
- подсчет статистики
- создание отчетов.

Этапы в процессе перевода с помощью CAT-tools

1) Создание проекта - шаги:

- Присвоение проекту названия
- Привязка к определенному клиенту (не обязательно) и тематической области (базе)
- Выбор языков оригинала и перевода
- Создание и подключение памяти переводов (одной или нескольких)
- Загрузка файла / файлов оригинала (и автоматическое преобразование их в спец. формат)
- Создание / подключение терминологических баз / лексиконов
- Подключение машинного перевода (при наличии опции)

2) автоматическое сегментирование ИТ (по умолчанию – автоматическая разбивка на предложения с ориентацией на точку / в очень длинных предложениях - на точку с запятой)

3) автоматический поиск совпадений между сегментами ИТ и сегментами, хранящимися в базе переводов.

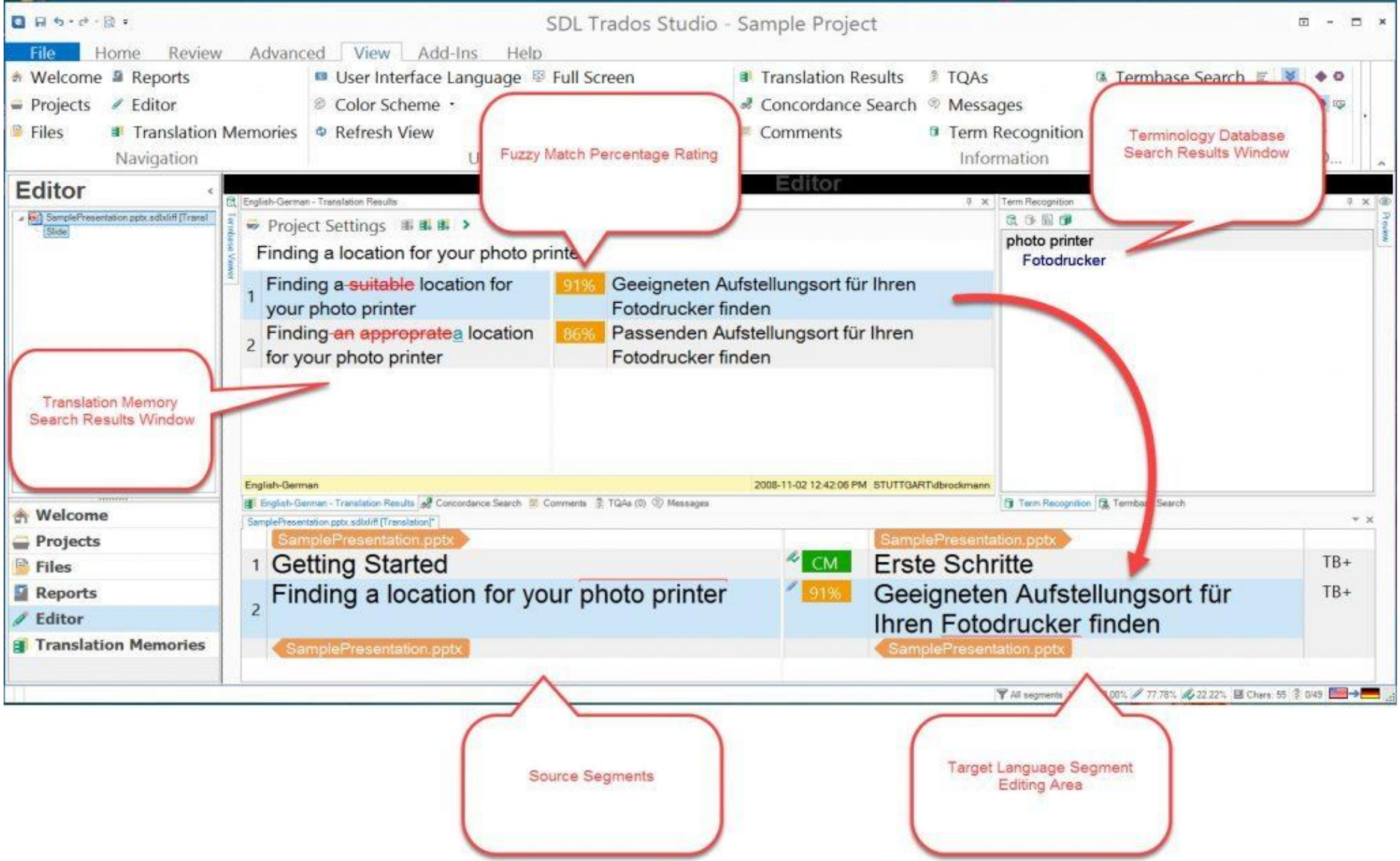
Настройка минимального процента совпадений - **75%,** если меньше – много доп. работы по редактированию!

4) Перевод не найденных сегментов и редактирование частично совпадающих сегментов (***Fuzzy matches***) – выполняется вручную!

5) Сохранение выполненных переводов (каждого сегмента!) в базе ТМ для последующего использования.

6) Экспорт выполненного перевода (и обратное преобразование файла в исходный формат)

7) Проверка и постредактура (форматирование, орфография)


SDL Trados Studio - Sample Project
File
Home
Review
Advanced
View
Add-Ins
Help
Welcome
Reports
Projects
Editor
Files
Translation Memories
Navigation
User Interface Language
Full Screen
Color Scheme
Refresh View
Translation Results
TQAs
Concordance Search
Messages
Comments
Term Recognition
Information
Termbase Search
Fuzzy Match Percentage Rating
Terminology Database Search Results Window
Editor
Project Settings
Finding a location for your photo printe
1 Finding a suitable location for your photo printer 91% Geeigneten Aufstellungsort für Ihren Fotodrucker finden
2 Finding an appropriate a location for your photo printer 86% Passenden Aufstellungsort für Ihren Fotodrucker finden
photo printer
Fotodrucker
Translation Memory Search Results Window
English-German
2008-11-02 12:42:06 PM STUTTGART\dbrockmann
SamplePresentation.pptx
1 Getting Started CM Erste Schritte TB+
2 Finding a location for your photo printer 91% Geeigneten Aufstellungsort für Ihren Fotodrucker finden TB+
Source Segments
Target Language Segment Editing Area

Преимущества ТМ для переводчика:

- повышение производительности труда. Подстановка даже на 80% совпадающих сегментов из базы переводов может сократить время работы над переводом на 50-60%. Как показывает практика, гораздо эффективнее править уже готовый перевод, чем переводить заново — «с нуля»;
- единство терминологии и стиля при наличии базы переводов по тематике переводимого документа. Это особенно важно при переводе узкоспециальной документации;
- организация работы коллектива переводчиков с гарантированным качеством перевода благодаря доступу к общим базам Translation Memory и Terminology Base.

2 условных этапа работы с системами автоматизации перевода

1) пользователь работает на систему, а именно:

осваивает базовую функциональность,

накапливает первоначальную базу переводов (преобразовывает ранее переведенные материалы в соответствующий формат),

создает глоссарий.

1) система начинает работать на пользователя. С увеличением базы переводов возрастает количество совпадений, следовательно, сокращается время, затрачиваемое на перевод, и увеличивается производительность.

CAT tools - неотъемлемая часть современного процесса профессионального перевода

В западных странах технология Translation Memory давно уже стала де-факто обязательным инструментом переводчика.

- средства, потраченные на создание базы переводов, рассматриваются не как затраты, а, скорее, как инвестиции в стабильную и качественную работу,
- что увеличивает не только прибыль, но и стоимость самой компании.

ТИПЫ СИСТЕМ автоматизированного перевода

По режиму доступа: десктопные (устанавливаются на компьютер)

облачные / браузерные / он-лайн системы

<u>Особенности он-лайн систем:</u>

«+» - мультиплатформенность

(совместимость с MS Windows, Mac OS, Linux)

«+» - Не требуют от устройств большого объема памяти и вычислительной мощности →

Подключение с любых мобильных устройств (Android, iOS).

«-» - зависимость от наличия доступа в Интернет, скорости соединения и пр.

«-» - частое обновление разработчиками, в т.ч. Интерфейса и алгоритмов → доп. трудозатраты переводчика на самостоятельное изучение / обращение в Службу поддержки

Наиболее популярные системы

- Trados (Традос), Deja Vu компании Artil
- Star Transit, Metatexis (Метатексис)
- Memsource OmegaT
- Smartcat Wordfast

SDL Trados - наиболее распространенная в РОССИИ и в мире - промышленный стандарт письменного перевода. По независимым оценкам, 80% переводчиков по всему миру применяют данные решения.

KILGRAY
Translation Technologies
memoQ

SDL*

ABBYY
SmartCAT

Т-Сервис
ТЕХНОЛОГИИ SDL TRADOS
SDL* Approved Reseller

STAR
Transit NXT

ATRIL
LANGUAGE ENGINEERING

déjàvu

MEMSOURCE
CLOUD

Вы знакомы с Наташиными кошками?

Что это значит?

(словарик переводчика)

Заайлайнили -

Традос -

Мемок -

Мемс -

АССОЦИАЦИЯ ПРЕПОДАВАТЕЛЕЙ ПЕРЕВОДА

Наташины «кошки»

Успехов в освоении профессии переводчика!