

Системы автоматизированного перевода и машинный перевод

Каримова Румия Фаилевна

студентка факультета иностранных языков

Елабужского института К(П)ФУ,

*Научный руководитель: кандидат физико-математических наук,
доцент кафедры математики и прикладной информатики ЕИ К(П)ФУ*

Миронова Юлия Николаевна

Аннотация: *В статье раскрыты такие понятия, как автоматизированный и машинный переводы, описаны их принципы работы и приведены соответствующие примеры.*

Принцип работы современного машинного перевода

В основе современных систем машинного перевода лежит алгоритм перевода, использующий формальную грамматику языков и статистические данные. Чтобы выучить язык, система сравнивает тысячи параллельных текстов, содержащих одну и ту же информацию, но на разных языках. Для каждого изученного текста система строит список уникальных признаков. Например, редко используемые слова и специальные знаки, которые встречаются в тексте с определенной частотой.

В системах машинного перевода, как правило, три основные части: модель перевода, модель языка и декодер. Модель перевода — это таблица, в которой для всех слов и фраз на одном языке перечислены возможные переводы на другой язык с указанием вероятности этих переводов. Система сравнивает не только отдельные слова, но и словосочетания из нескольких слов, идущих подряд. Модели перевода для каждой пары языков содержат миллионы пар слов и словосочетаний. Что касается модели языка, то она создается системой на этапе изучения текстов.

Переводом занимается декодер. Он проводит морфологический и синтаксический анализ текста и для каждого предложения подбирает все варианты перевода с сортировкой по убыванию вероятности. Затем все полученные варианты декодер оценивает с помощью модели языка на частоту употребления и выбирает предложение с наилучшим сочетанием вероятности и частоты.

Использование статистических данных позволяет системам машинного перевода меняться вместе с языком. Если люди начинают писать какое-то слово по-другому, система видит это, как только к ней попадают новые тексты. Чтобы улучшать качество перевода, систему регулярно обновляют и проводят проверки. Впрочем, высококачественный машинный перевод текстов по-прежнему недостижим. Однако он значительно облегчает и ускоряет работу переводчикам.

Вместо «*машинный*» иногда употребляется слово **автоматический**, что не влияет на смысл. Однако термин **автоматизированный перевод** имеет совсем другое значение — при нём программа просто помогает человеку переводить тексты.

Автоматизированный перевод – перевод текстов с одного языка на другой человеком с использованием специализированных программ, приложений.

Машинный перевод – процесс перевода текстов с одного языка на другой машиной посредством специальной компьютерной программы.

Между ними есть три главных различия:

- Трудозатраты переводчика.
- Специализированное программное обеспечение.
- Качество.

Так, при автоматизированном переводе основную работу выполняет переводчик, а специальные программы выступают только в качестве вспомогательного инструмента. Его главной целью является сокращение

времени процесса, обеспечение единообразия терминов и общего соответствия (качества).

Машинный же перевод производит сама программа, а человек только редактирует полученный результат. И здесь уже цель – заменить труд человека, получая при этом быстрый перевод низкого качества.

Специальные программы

Системы **автоматизированного перевода** – это собирательное определение для специализированных программ и интернет-сервисов, которые используют переводчики в процессе работы. Они хорошо подходят для работы над художественными, юридическими и техническими переводами. К ним относятся:

- Отдельные либо встроенные редакторы, которые помогают автоматически проверить грамматику текста и правописание слов.
- ПО, таблицы, текстовые редакторы, обеспечивающие управление терминологией (MultiTerm, Termex и т.д.).
- ПО, с помощью которого осуществляется менеджмент переводческих проектов.
- CAT-инструменты, использующие ТМ (TranslationMemory – Память переводов), которые включают в себя образцы ранее переведённых текстов или предложений. Яркими их представителями стали программы Trados, DéjàVu, MemoQ, MemSource, Wordfast.
- Корпусы, представляющие собой большой ряд документов с использованием одного или нескольких языков. С их помощью составляется сжатое описание употребления слов и выражений в общих случаях или с учётом какой-либо определённой предметной темы.

TranslationMemory – «Память перевода»

Сегодня письменный перевод нужно делать не только качественно, но и быстро. На помощь специалисту приходят программные разработки, такие

как МемоQ. Их общее распространённое название «Память перевода», принцип – «не переводи одно и то же дважды». Память перевода не следует путать с программами машинного перевода. Фактически TranslationMemory – это наработанная база переведённых слов, выражений, синтаксических целых, которая создаётся самим переводчиком. В каждом новом тексте программа определяет уже «знакомые» ей понятия и предлагает специалисту перевод из базы. Таким образом, достигается единообразие в письменном переводе документов и, разумеется, значительная экономия времени.

Машинный перевод также представлен набором программ и интернет-сервисов. Наиболее популярные из них: PROMT, Google translate, Яндекс.Переводчик.

Google translate

Переводчик Google Translate использует SMT (статистический уникальный машинный перевод текста - Statistical Machine Translation) Особенностью Google Translate является метод перевода: он не основан на анализе правил грамматики, а основан на поиске соответствий языка между переводимым текстом и гигантским массивом сервиса, который состоит из слов, вносимых пользователем ранее во время их перевода. Данные блоки текста составляются из всех возможных достоверных источников. В качестве примера можно привести документы государственных организаций. Они являются одними из наиболее проверенных источников для базы данных. Преимущество заключается в том, что такие документы содержат информацию, доступную на многих языках.

Также для развития Google Translate важны книги. В настоящий момент известно, что «корпус» Google Translate обладает более чем триллионом слов. Так сервис активно использует пользовательский уникальный перевод для повышения качественных характеристик машинного текстового перевода путем занесения пользовательских вариаций фраз в базу данных. Таким образом, такое сочетание способствует увеличению уровня качества генерации уникального текста.

Статистическому современному переводу необходим анализ параллельных языковых пар для осуществления генерации текста. Он самостоятельно приспосабливается к новой лексике. Система самостоятельно заносит в свою совершенную базу данных новые вариации перевода слов или словосочетаний, если они отличаются. Она хранит в памяти новые варианты возможного перевода и может использовать эти версии в будущем. Одним из главных достоинств SMT является то, что он обновляется и развивается параллельно с языком. То есть, если появляется новое слово или меняется его лексическое значение, то система распознает это и обновляет свою базу данных. И, как следствие, ускоряется «обучение» системы и совершенствуется качество генерации текста.

Но есть и значительный недостаток SMT. Этой системе необходимо высокопроизводительное аппаратное обеспечение. Для совершенствования программы требуется огромное число вычислений. Также для перевода SMT характерно то, что качество генерации текста полностью зависит от количества данных в корпусе сервиса. Чаще генерируемые сервисом вариации перевода текста оправдывает возложенные ожидания современного человека. Хотя предсказать заранее текстовое содержание, которое сгенерирует Google Translate, сложно.

Вывод

Описание различий автоматизированного и машинного переводов подтверждает тот факт, что инструменты профессионального переводчика – это системы автоматизированного перевода. Именно с их помощью можно сэкономить время, обеспечить единообразие текста и его качество.

В тоже время, машинный перевод может помочь, например, в экстренной ситуации, когда человек не знает языка, но срочно нужен перевод. Пусть даже он плохого качества.

Список использованной литературы

1. Разница машинного и автоматизированного перевода [Электронный ресурс] URL: <http://www.toptr.ru/library/translation-truth/avtomatizirovannyij-i-mashinnyij-perevodyi-v-chyom-raznicza.html>
2. Принцип работы онлайн-переводчика Google translate [Электронный ресурс] URL: <http://www.inteeu.com/2016/11/21/printsip-raboty-onlajn-perevodchika-google-translate/>
3. Принцип работы машинного перевода [Электронный ресурс] URL: <https://hi-news.ru/eto-interesno/kak-eto-rabotaet-mashinnyj-perevod.html>
4. Гусякова А.В. Информационные технологии и лингвистика XXI века, 2016