

ЛИНГВОКУЛЬТУРОЛОГИЧЕСКИЕ АСПЕКТЫ МАШИННОГО ПЕРЕВОДА

Бул макалада машиналык котормону ар түрдүү заманбап программаларда иштеп чыгуу каралган.

В статье рассмотрены различные виды современных разработок программ машинного перевода.

The article devoted to various kinds of modern programming machine translation.

С возникновением письменности люди получили мощный инструмент для сохранения знаний и для коммуникации. Первые письмена, дошедшие до нас на стенах храмов и гробниц, повествуют о деяниях царей и полководцев, происшедших много веков назад. Кроме того, люди записывали результаты хозяйственной деятельности, для того чтобы успешно торговать, собирать налоги. Чтобы облегчить письменное общение между народами, были созданы первые словари. Один из таких словарей был написан шумерскими жрецами на глиняных табличках. Каждая табличка была поделена на две равные части. С одной стороны записывалось шумерское слово, а с другой – аналогичное по значению слово на другом языке, иногда с кратким пояснением. С тех времен до наших дней структура словарей практически не изменилась. С появлением персонального компьютера стали создаваться электронные словари, облегчившие поиск нужного слова и предложившие множество новых полезных функций (озвучивание слова, поиск синонимов и т.д.). Технология машинного перевода постепенно совершенствовалась. И если качество и скорость перевода первых систем оставляли желать лучшего, то теперь компьютер действительно может связно переводить текст с одного языка на другой. А современные системы с приемлемым качеством переводят 1 страницу текста за 1 секунду.

Что такое машинный перевод? В нашем понимании это автоматический перевод текстов, удобный для применения и использования. Однако при использовании таких переводов мы часто не задумываемся о качестве данных переводов. Итак, для начала выясним, что представляет собой машинный перевод.

Машинный перевод – процесс перевода текстов (письменных, а в идеале и устных) с одного естественного языка на другой с помощью специальной компьютерной программы. Так же называется направление научных исследований, связанных с построением подобных систем. Мысль использовать ЭВМ для перевода была высказана в 1947 году в США, сразу после появления первых ЭВМ. Первая публичная демонстрация машинного перевода состоялась в 1954 году. Несмотря на примитивность той системы (словарь в 250 слов, грамматика из 6 правил,

перевод нескольких простых фраз), этот эксперимент получил широкий резонанс: начались исследования в Англии, Болгарии, ГДР, Италии, Китае, Франции, ФРГ, Японии и других странах; в том же 1954 году и в СССР.

Существует такая разновидность машинного перевода, как статистический машинный перевод. Статистический машинный перевод – это разновидность машинного перевода текста, основанная на сравнении больших объемов языковых пар. Языковые пары – тексты, содержащие предложения на одном языке и соответствующие им предложения на втором, могут быть как вариантами написания двух предложений человеком – носителем двух языков, так и набором предложений и их переводов, выполненных человеком. Таким образом, статистический машинный перевод обладает свойством «самообучения». Чем больше в распоряжении имеется языковых пар и чем точнее они соответствуют друг другу, тем лучше результат статистического машинного перевода. Под понятием "статистического машинного перевода" подразумевается общий подход к решению проблемы перевода, который основан на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов. В качестве примера двуязычной совокупности текстов можно назвать парламентские отчеты, которые представляют собой протоколы дебатов в парламенте. Двуязычные парламентские отчеты издаются в Канаде, Гонконге и других странах; официальные документы Европейского экономического сообщества издаются на 11 языках; а Организация объединенных наций публикует документы на нескольких языках. Как оказалось, эти материалы представляют собой бесценные ресурсы для статистического машинного перевода.

Несмотря на то, что в современном мире имеется достаточное количество разработок для машинного перевода, тем не менее, зачастую качество перевода оставляет желать лучшего. Качество перевода зависит от тематики и стиля исходного текста, а также грамматической, синтаксической и лексической родственности языков, между которыми производится перевод. Машинный перевод художественных текстов практически всегда оказывается неудовлетворительного качества. Тем не менее, для технических документов при наличии специализированных машинных словарей и некоторой настройке системы на особенности того или иного типа текстов возможно получение перевода приемлемого качества, который нуждается лишь в небольшой редакторской корректировке. Чем более формализован стиль исходного документа, тем большего качества перевода можно ожидать. Самых лучших результатов при использовании машинного перевода можно достичь для текстов, написанных в техническом (различные описания и руководства) и официально-деловом стиле.

Применение машинного перевода без настройки на тематику (или с намеренно неверной настройкой) служит предметом многочисленных бродящих по Интернету шуток. Из пространных примеров наиболее известен текст «Гуртовщики Мыши» (перевод компьютерной документации программой Poliglossum на основе медицинского, коммерческого и юридического словарей); из кратких – фраза «My cat has given birth to four kittens, two yellow, one white and one black», которую переводчик компании ПРОМТ превращает в «Моя кошка родила четырех котят, два желтых цвета,

одного белого и одного афроамериканца». Главной причиной того, почему программа перевела именно так, было то, что после слова black нужно было добавить kitten, тогда программа переведет правильно: «Моя кошка родила четырех котят: двух желтых, одного белого и одного черного котенка».

Чаще всего подобные шутки связаны с тем, что программа не распознает контекст фразы и переводит термины дословно, к тому же не отличая собственных имен от обычных слов. Тот же переводчик ПРОМТ превращает «Lie algebra» – в «алгебру Лжи», «essentricity vector» – в «вектор оригинальности», «Shawnee Smith» – в «индеец племени шони Смит» и т.п.

В настоящее время существует множество коммерческих проектов машинного перевода. Одним из пионеров в области машинного перевода была компания Systran. В России большой вклад в развитие машинного перевода внесла группа под руководством проф. Р. Г. Пиотровского (Российский государственный педагогический университет имени А.И.Герцена, Санкт-Петербург).

К середине 1960-х в США для практического использования были предоставлены две системы русско-английского перевода:

- MARK (в Департаменте иностранной техники ВВС США);
- GAT (разработка Джорджтаунского университета, использовалась в Национальной лаборатории атомной энергии в Окридже и в центре Евратома в г. Испра, Италия).

Однако созданная для оценки подобных систем комиссия ALPAC пришла к выводу, что в силу низкого качества машинно-переведенных текстов эта деятельность в условиях США нерентабельна. Хотя комиссия рекомендовала продолжать и углублять теоретические разработки, в целом ее выводы привели к росту пессимизма, снижению финансирования, часто к полному прекращению работ по этой тематике.

Тем не менее, в ряде стран исследования продолжались, чему способствовал постоянный прогресс вычислительной техники. Особенно существенным фактором стало появление мини- и персональных компьютеров, а с ними все более сложных словарных, поисковых и т.п. систем, ориентированных на работу с естественно-языковыми данными. Росла и необходимость в переводе как таковом ввиду роста международных связей. Все это привело к новому подъему этой области, наступившему примерно с середины 1970-х. В 1980-е наступило время широкого практического использования переводческих систем, сложился рынок коммерческих разработок по этой теме.

В настоящее время Cognitive Technologies разработала систему машинного перевода текстов нового поколения Cognitive Translator, в основу которой легло более глубокое представление знания о языке. Процесс разработки системы занял около 60 человеко/лет и длился более 8 лет.

Научный коллектив Cognitive Technologies (в то время ВНИИСИ АН СССР) начал заниматься задачами машинного перевода как направлением искусственного интеллекта более 20 лет назад, за 5 лет до образования компании, в рамках работ по набору первых словарей перевода.

Проект Cognitive Translator рассматривается компанией как научный, хотя некоторые из его результатов сегодня используются при построении комплексных информационных систем и при разработке тиражируемых программных продуктов.

Проект по созданию данной версии Cognitive Translator начался в 2000 году. Его участниками, помимо специалистов Cognitive Technologies, стала группа профессиональных лингвистов из МГУ, Института русского языка РАН и других организаций. Группой были разработаны уникальные алгоритмы синтаксического разбора и правила машинного перевода. На их основе была построена новая более совершенная – структурная модель представления знания о языке, которая включает, помимо описания морфологии и синтаксиса, еще и элементы анализа структурных связей. За счет этого система добивается глубокого "понимания" текста, и даже претендует на разрешение смысловых неоднозначностей ("ключ" дверной и "ключ" – источник воды).

Принципы построения модели языка, заложенные в Cognitive Translator, основаны на технологии анализа – понимания текста, что позволило в рамках исследовательского проекта при участии относительно небольшой группы разработчиков создать систему, дающую качество перевода, сравнимое с современными промышленными системами. Стоит отметить, что существующие сегодня на рынке системы машинного перевода используют подход, разработанный еще в 70-х годах прошлого века, подразумевающий непрерывное ручное наполнения базы словоформ. Для достижения высокого качества перевода и добавления новых языков этот подход требует подключения огромных ресурсов (сотен специалистов), что, естественно, ограничивает число коллективов, занятых в этой области. Предложенная компанией Cognitive Technologies модель описания "языка" практически инвариантна (по отношению к языкам) и позволяет на порядок снизить трудозатраты при описании морфологии и синтаксиса различных языков мира. Построенные на ее базе системы перевода нового поколения позволят разрешать смысловые неоднозначности не только для двух языков, но и для всего межъязыкового пространства. В настоящее время ядро системы – анализатор текста, предназначенный для описания новых языков, находится в открытом доступе.

Мечты, с которыми человечество полвека назад взялось за задачу машинного перевода, в значительной мере остаются мечтами: высококачественный перевод текстов широкой тематики по-прежнему недостижим. Однако прогресс не стоит на месте. Несомненно, что по истечении времени человечество найдет способ делать машинные переводы с учетом всех лингвокультурологических аспектов без участия человека-переводчика. Мировые державы прилагают массу усилий, вкладывают большое количество средств в развитие этой отрасли, что в итоге приведет к усовершенствованию техники машинного перевода.

СПИСОК ЛИТЕРАТУРЫ

1. Bar-Hillel Y. The present status of automatic translation of languages. In Alt F. L. (Ed.), *Advances in Computers*, Vol. 1, p. 91-163. New York: Academic Press, 1960.
2. Current Issues in Computational Linguistics. Proceedings. International Conference, University Sains Malaysia, Penang, Malaysia, 1991, 471 pp.
3. Hutchins W. J. and Somers H. *An Introduction to Machine Translation*. New York : Academic Press, 1992.
4. Hutchins W. J. *Machine Translation: Past, Present. Future*. Chichester, Ellis Horwood, N. Y. Etc. 1986, pp. 382.
5. Knight, K. Automating knowledge acquisition for machine translation. *AI Magazine* 18, pp.81-96, 1997.
6. Kozerenko, E.B. Cognitive Approach to Language Structure Segmentation for Machine Translation Algorithms // *Proceedings of the International Conference on Machine Learning, Models, Technologies and Applications*, June, 23-26, 2003, Las Vegas, USA.// CSREA Press, pp. 49-55, 2003.
7. Lindsay R. K. Inferential memory as the basis of machines which understand natural language. In Feigenbaum E. A. and Feldman J. (Eds.), *Computers and Thought*, p. 217—236. McGraw-Hill, New York, 1963
8. Locke W. N. and Booth A. D. *Machine Translation of Languages: Fourteen Essays*. MIT Press, Cambridge, Massachusetts, 1955.
9. MacDonald N. Language translation by machine - a report of the first successful trial. *Computers and automation*. 1954, v. 3, № 2, p. 6-10.
10. Багриновская Г. П., Кулагина О. С., Ляпунов А. А. О некоторых методологических вопросах, относящихся к машинному переводу// О некоторых вопросах теоретической кибернетики и алгоритмах программирования. – Новосибирск: Изд-во СО АН СССР, 1971.
11. Бакулов А.Д., Леонтьева Н.Н.. Теоретические аспекты машинного перевода // *Искусственный интеллект. Кн.1. Системы общения и экспертные системы*. – М., 1990. – С. 201 - 216.
12. Все о машинном переводе// *ComputerBild*. – 2007. – № 22.
13. Евдокимов А. От омонимии к полисемии. Искусство машинного перевода - 2// *Hard'n'Soft*. – 2008. – № 2.
14. Едемский М. Программы автоматического перевода // *Мир образования*. – 1996. – № 11-12. – С.54-55.
15. Кноп К. Переведутся ли переводчики?// *Мир Internet*. – 2002. – № 8.
16. Козеренко Е.Б. Логико-статистические методы представления языковых структур в машинном переводе// *Труды Международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии»*. – М.: Наука, 2005.
17. Козеренко Е.Б. Моделирование переноса функциональных значений для англо-русского машинного перевода // *Труды Международной конференции Диалог'2004 «Компьютерная лингвистика и интеллектуальные технологии»*. – М.: Наука, 2004.

18. Кулагина О. С. Исследования по машинному переводу. – М.: Наука 1979.
19. Кулагина О. С. Машинный перевод: современное состояние// Семиотика и информатика. Вып. 29. – М.: ВИНТИ, 1989.
20. Кулагина О.С. О современном состоянии машинного перевода // Математические вопросы кибернетики. Вып. 3. – М.: Наука, 1991.
21. Марчук Ю. Н. Модель «текст-текст» и переводные соответствия в теории машинного перевода. Проблемы компьютерной лингвистики. – Минск: МГЛУ, 1997. – С. 21-29.
22. Марчук Ю. Н. Проблемы машинного перевода. – М.: Наука, 1983.
23. Марчук Ю. Н. Теория и практика машинного перевода. Русский филологический вестник. – М.: Московский Лицей, том 81, 1996.
24. Моисеев А. Вавилонские рыбки: Обзор электронных словарей и переводчиков// Игромания. – 2006. – № 108.
25. Нелюбин Л. Л. Компьютерная лингвистика и машинный перевод. – М.: ВЦП, 1991.
26. Слокум Дж. Обзор разработок по машинному переводу// Новое в зарубежной лингвистике. – М., 1989.
27. Шаляпина З.М. Автоматический перевод: эволюция и современные тенденции// Вопросы языкознания. – 1996. – № 2. – С.105-117