Тукеев У.А., Рахимова Д.Р., Жуманов Ж.М., Сундетова А.М.

МАШИННЫЙ ПЕРЕВОД КАЗАХСКОГО ЯЗЫКА НА АНГЛИЙСКИЙ И РУССКИЙ ЯЗЫКИ (И ОБРАТНО) НА БАЗЕ ПЛАТФОРМЫ APERTIUM

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Активная интеграция Казахстана в мировое сообщество и увеличивающимся объемом информационных потоков между нашей страной и ее зарубежными партнерами, реальная потребность для различных слоев населения в оперативном компьютерном переводе при работе в Интернете определяют актуальность вопросов машинного (компьютерного) перевода казахского языка на различные ведущие мировые языки, такие как английский, русский, французский, немецкий, и последнее время, китайский языки, а также и обратного машинного перевода. Первоочередными задачами для информационного взаимодействия населения Казахстана с зарубежными партнерами и внутри страны определены взаимодействия по трем языкам: казахскому, английскому и русскому. В связи с этим является весьма актуальным является высокоэффективная инструментальная поддержка машинным переводом такого трехязычного языкового взаимодействия. В связи с чем актуальным являются исследования и разработка систем машинного перевода промышленного качества с казахского языка на английский и русский языки и обратно.

Анализ состояния исследований в области машинного перевода с казахского языка на английский, русский языки и обратно показывает, что исследования в данной области практически отсутствуют, несмотря на наличие двух-трех коммерческих программных продуктов, качество перевода которых является не достаточно высоким.

Область машинного перевода в мировой науке является достаточно зрелой, разработаны многие формальные модели, алгоритмы, существуют достаточно высокого уровня системы машинного перевода. Однако, большинство результатов машинного перевода получены и применены к ведущим мировым языкам, таким как английский, французский, немецкий, русский, китайский. В приложении к казахскому языку разработанных моделей и алгоритмов, или применений уже разработанных моделей и алгоритмов весьма мало.

Анализ реализации машинного перевода естественных языков показывает всю сложность проблемы машинного перевода естественных языков, как одну из главных проблем создания систем искусственного интеллекта. Выявилось, что основной проблемой машинного перевода естественных языков является недостаточное качество перевода. На это влияют множество вопросов таких как неоднозначность слов в естественных языках, различие грамматических структурах предложений в различных языках, вопросы определения имен собственных, вопросы различного представления темпоральных выражений в различных языках.

Кроме этих языковых проблем, при реализации систем машинного перевода возникают и проблемы реализации программного обеспечения машинного перевода промышленного качества, что требует разработки или выбора соответствующей платформы, использующей уже принятые стандарты хранения и обмена данными в области машинного перевода.

Для решения данной проблемы авторы остановили свой выбор на свободной/открытой платформе машинного перевода Apertium. Выбор свободной(бесплатной) открытой платформы машинного перевода был обусловлен тем, использование коммерческих систем требуют значительных первоначальных вложений. И кроме этого, оказалось, что среди свободных открытых платформ машинного перевода

система Apertium является наиболее известной платформой, основанной на правилах, и которая активно расширяется, исследуя и внедряя статистические методы машинного перевода. Разработка собственной платформы машинного перевода является, с нашей точки зрения, нецелесообразным по соображениям существенных дополнительных значительных финансовых и временных затрат.

В данной работе авторы описывают опыт реализации системы машинного перевода казахского языка на английский и русский языки (и обратно), включая модели, алгоритмы и программное обеспечение.

Благодарности. Авторы выражают благодарность МОН РК за финансовую поддержку данного исследования (проект 0749/ГФ4), а также всех участников проекта за их вклад в разработку проекта.