

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЦИФРОВЫХ ТЕХНОЛОГИЯХ НА ПРИМЕРЕ ЭЛЕКТРОННЫЙ ПЕРЕВОД С УЗБЕКСКОГО НА КЫРГЫЗСКИЙ ЯЗЫК

*Назирова Элмира Шодмоновна д.т.н., профессор, декан факультета Телевизионные технологии
ТУИТ имени Мухаммада аль-Хорезмий*

*Абидова Шахноза Бахадировна доцент, PhD кафедры Мультимедийные технологии
ТУИТ имени Мухаммада аль-Хорезмий,*

*Узакова Мамура Абдурайимовна старший преподаватель PhD кафедры Искусственного
интеллекта ТУИТ имени Мухаммада аль-Хорезмий,*

Самадов Дилишод Усмон угли магистрант 1-курса ТУИТ имени Мухаммада аль-Хорезмий

Аннотация

В данной статье рассматривается роль искусственного интеллекта (ИИ) в развитии цифровых технологий с особым акцентом на системах машинного перевода между узбекским и кыргызским языками. Анализируются современные подходы к нейронному машинному переводу (НМП), архитектурные решения на основе трансформеров, а также специфические лингвистические особенности тюркских языков, определяющие направления разработки. Обосновывается актуальность создания высококачественных систем автоматического перевода для поддержки межкультурной коммуникации, деловых связей и развития информационного общества в Центральной Азии.

Ключевые слова: *искусственный интеллект, машинный перевод, нейронные сети, трансформер, узбекский язык, кыргызский язык, тюркские языки, цифровые технологии, обработка естественного языка.*

Введение

Цифровая эпоха кардинально изменила способы коммуникации между народами и государствами. Искусственный интеллект занял центральное место в этой трансформации, предоставив беспрецедентные инструменты для обработки и передачи информации через языковые барьеры. Среди наиболее востребованных приложений ИИ - системы машинного перевода, которые обеспечивают мгновенный доступ к контенту на иностранных языках, содействуют международному сотрудничеству и расширяют горизонты образования [1].

Узбекистан и Кыргызстан - два соседних центральноазиатских государства с тесными историческими, культурными и экономическими связями. Несмотря на типологическое родство узбекского и кыргызского языков, принадлежащих к тюркской языковой семье, между ними существуют значительные фонетические, морфологические и лексические различия, которые существенно затрудняют взаимопонимание. В условиях интенсификации двусторонних отношений потребность в качественном автоматическом переводе между этими языками становится всё более острой [2].

В данной статье приведены цели такие как

- которые обозначить ключевые тенденции применения ИИ в сфере машинного перевода;
- охарактеризовать лингвистические особенности узбекского и кыргызского языков, релевантные для разработки переводных систем;
- описать современные технологические решения и их потенциал применительно к данной языковой паре;

- обозначить перспективы и вызовы дальнейших исследований направлений обработка естественного языка занимает особое место, поскольку язык является главным инструментом человеческой коммуникации, хранения знаний и культурной идентичности [3].

Прорыв в этой области напрямую влияет на доступность информации и равенство возможностей в цифровом мире.

Революционным сдвигом стало появление нейронного машинного перевода (NMT) в середине 2010-х годов. Архитектура «кодировщик - декодировщик» на основе рекуррентных нейронных сетей (RNN) с механизмом внимания (attention mechanism) позволила учитывать глобальный контекст предложения и существенно улучшила качество перевода. В электронном переводе трансформеры тоже имеют значительную роль. Трансформер основан на механизме самовнимания (self-attention), позволяющем модели одновременно учитывать взаимосвязи между всеми словами входного предложения. В отличие от RNN, трансформер эффективно обрабатывает длинные зависимости и допускает параллелизацию вычислений, что сокращает время обучения. На базе трансформерной архитектуры созданы крупнейшие языковые модели мира: BERT, GPT, T5, mBART и многие другие [4].

Многоязычные модели, такие как mBART-50 и NLLB-200 компании Meta AI, обучены одновременно на сотнях языков и обеспечивают перевод между любой парой включённых языков - в том числе узбекским и киргизским. Подход переносного обучения (transfer learning) позволяет дообучить предобученную модель на небольшом корпусе целевых языков и достичь приемлемого качества даже при ограниченных ресурсах.

Оба языка принадлежат к тюркской семье, что обуславливает ряд общих черт: агглютинативная морфология (добавление суффиксов для выражения грамматических значений), синтаксическая модель «субъект - объект - глагол» (Subject Object Verb), развитая система падежей, согласование по числу и лицу. Тюркское типологическое родство создаёт благоприятную почву для применения методов переносного обучения и совместного обучения многоязычных моделей [5].

Вместе с тем между языками существуют принципиальные различия. Узбекский язык относится к карлукской подгруппе и характеризуется утратой гармонии гласных, значительным пластом персидско-арабской лексики и современной системой письма на основе латинского алфавита (с 1993 года). Киргизский язык входит в кипчакскую подгруппу, сохраняет вокальную гармонию, широко использует кириллицу и располагает более архаичным лексическим фондом. Данные различия определяют специфику задачи перевода и требуют специализированных технических решений.

С точки зрения машинного перевода, ключевые трудности связаны со следующими факторами:

- различие алфавитов и орфографических систем: латиница в узбекском и кириллица в киргизском требуют предварительной транслитерации или работы систем на уровне символов;
- богатство морфологии: один корень может порождать десятки словоформ, что требует от модели глубокого понимания морфологических правил;
- дефицит параллельных корпусов: в отличие от европейских языковых пар, объём параллельных текстов узбекского и киргизского весьма ограничен, что затрудняет обучение систем NMT;

- диалектное разнообразие: оба языка имеют региональные варианты с существенными фонетическими и лексическими расхождениями;
- неологизмы и заимствования: стремительное обновление терминологии в сфере технологий, экономики и права опережает пополнение учебных корпусов.

На сегодняшний день наиболее доступными системами перевода в рассматриваемой языковой паре являются Google Translate и Yandex Translate. Обе системы основаны на нейронном машинном переводе и поддерживают как узбекский, так и кыргызский языки. Тем не менее, по оценкам лингвистов и пользователей, качество перевода между этими языками остаётся значительно ниже, чем для пар с развитыми корпусами, таких как английский–испанский или русский–немецкий.

Методология разработки специализированной системы перевода

Разработка высококачественной системы электронного перевода с узбекского на кыргызский язык предполагает поэтапный подход:

На первом этапе необходимо сформировать репрезентативный параллельный корпус, включающий тексты различных жанров и тематических областей: официальные документы, художественная литература, публицистика, технические тексты, разговорная речь.

На втором этапе производится дообучение предобученной многоязычной модели (например, mBART-50 или NLLB-200) на сформированном корпусе. Техника дообучения (fine-tuning) позволяет адаптировать уже освоенные моделью общелингвистические закономерности к специфике конкретной языковой пары, значительно сокращая требования к объёму данных по сравнению с обучением с нуля.

Третий этап - комплексная оценка качества перевода с применением автоматических метрик (MRS_{hb} - meteor rough suffix *hybrid*) и экспертной лингвистической экспертизы. Автоматические метрики обеспечивают воспроизводимость и масштабируемость оценки, тогда как экспертная оценка позволяет зафиксировать семантические и прагматические аспекты, недоступные количественным показателям [4].

Проблема малоресурсных языков - одна из центральных в современной NLP-исследовательской повестке. Для её преодоления применяются техники аугментации данных: обратный перевод (back-translation), переформулировка предложений (paraphrasing), синтез обучающих примеров с помощью LLM. Обратный перевод предполагает автоматический перевод целевых монологических текстов на исходный язык и использование полученных пар для обучения - это позволяет многократно расширить обучающую выборку без привлечения профессиональных переводчиков.

Развитие качественных систем электронного перевода между узбекским и кыргызским языками имеет широкое практическое значение для различных сфер общественной жизни.

Заключение

Искусственный интеллект кардинально трансформирует сферу цифровых коммуникаций, выводя машинный перевод на качественно новый уровень. Применительно к языковой паре «узбекский - кыргызский» современные технологии нейронного перевода и большие языковые модели создают реальные предпосылки для преодоления языкового барьера между двумя соседними народами.

Вместе с тем реализация этого потенциала требует целенаправленных усилий: формирования полноценных параллельных корпусов, разработки специализированных

моделей с учётом лингвистической специфики тюркских языков, государственной поддержки и международного научного сотрудничества. Инвестиции в эту область окупятся многократно - в виде укрепления двусторонних экономических связей, расширения доступа к знаниям и сохранения языкового и культурного разнообразия региона. Электронный перевод с узбекского на кыргызский язык - не просто технологическая задача, но и стратегический приоритет цифровой интеграции стран. Его решение будет способствовать сближению народов, развитию региона и построению инклюзивного информационного общества.

Список литературы

1. Zaynidinov X.N., Maxkamov B.Sh., Nurmurodov J.N 1 “Sun’iy intellekt asoslari” /. Toshkent: “Aloqachi”, 2024.-210 b
2. E. Nazirova, M. Uzakova. F.Boymurodov, Y.Xolmo’minov Turkiy tillar oilasiga mansub o‘zbek va qirg‘iz tillarida elektron tarjimaning matematik modelini ishlab chiqish Amaliy matematikaning zamonaviy muammolar va istiqbollari Respublika ilmiy - amaliy konferensiya, 24-25 may 2024 yil, Qarshi, -B.326-328
3. Uzakova M.A. Turkiy tillar oilasiga mansub qirg‘iz tilida so‘z birikmalarini hosil qilishning lingvo-matematik modeli va algoritmi “TATU xabarlarlari” ilmiy-texnika va axborot-tahliliy jurnali. №2 (70), (05.00.00;№31).2024. – B 75-87. <http://jurnal.tuit.uz>
4. Elmira Nazirova, Shakhnoza Abidova, Mamura Uzakova, Khurshida Karabayeva Mathematical Software for Creating a Trilingual Electronic Translator Between the Languages of the Turkish-Language Family Journal of “Computational Analysis and Applications”. ISSN 15211398, 15729206, SJR 0.19. VOL. 33, 2, 2024. -P.220-223.
5. E. Nazirova, Sh.Abidova, M. Uzakova “Mathematical Models of Complex Word Formation in Uzbek, Karakalpak and Kyrgyz Languages belonging to the Turkic Language Family” International conference “Modern Problems of Applied Science and Engineering” – 2024, 2-3 may