

Business value

Создание переводчика с мансийского на русский язык — важная задача для сохранения и развития мансийского языка, который является неотъемлемой частью культурного наследия коренных народов России. Манси — один из малочисленных народов Севера, и их язык находится под угрозой исчезновения из-за ограниченного числа носителей и отсутствия широкого использования в повседневной жизни.

Создание переводчика поможет не только сохранить язык для будущих поколений, но и сделать его доступным для широкой аудитории. Современные технологии могут сыграть ключевую роль в распространении и популяризации мансийского языка, способствуя его изучению и использованию. Кроме того, переводчик откроет доступ к мансийскому культурному наследию, литературе и устным традициям, что укрепит национальное самосознание и культурную идентичность манси.

Наша цель — сохранить мансийский язык, обеспечив его актуальность и востребованность в современных условиях. Это важный шаг для того, чтобы традиции и культура манси продолжали жить и передаваться будущим поколениям.

Целевая аудитория: люди любого возраста, которые хотят получить перевод манси↔рус; научные сотрудники, которые хотят воспользоваться базами данных с целью исследования или улучшения сервиса; носители, желающие помочь в развитии сервиса.

Боль потенциальных клиентов:

1. Сложно найти информацию о мансийском языке, так как много старых учебников, которые используют различные диалекты, хотя на текущий момент мансийский имеет только 1 диалект. Мы решаем эту проблему тем, что собрали учебные и полезные материалы в одну вкладку "[Полезные материалы](#)".
2. Отсутствие веб-интерфейса с хорошим переводчиком (говорим об open source только). Исследования показали, что текущие существующие open source модели имеют небольшой скор по метрике BLEU, мы же смогли достигнуть оценки в 22.8 баллов и реализовали веб-интерфейс с понятным user-friendly дизайном.
3. Отсутствие сбора данных для дальнейших исследований. Мы предоставляем тематический корпус, словарь и базу данных с примерами предложений.

Competitors

Поскольку создание переводчика для редкого языка требует значительных усилий в области сбора данных и обучения моделей, серьёзная конкуренция здесь пока ограничена. Основное внимание уделяется научным и культурным учреждениям, а также потенциально большим компаниям, которые могли бы использовать свои технологии для поддержки таких проектов.

- TartuNLP: Исследовательская группа по обработке естественного языка в Институте компьютерных наук Тартуского университета. В сферу их деятельности входят машинный перевод, синтез речи, НЛП для эстонского языка и другие.
В том числе авторы разработали модель для перевода с мансийского языка, однако по результатам исследования оказалось, что их подход не был удачным из-за источника датасета. Подход на наших тестовых данных показал следующие метрики: BLEU: 0.0039, ChrF: 23.5380.
- Компании с доступом к бОльшим ресурсам: Яндекс, Гугл, которые могут начать исследования с целью программы сохранения языков или с целью развития сервиса.
- Научные и образовательные учреждения.
- Региональные правительства и культурные организации.