

Казахский язык могут добавить в Google.Translate уже в 2013 году

[05.02.2013] Категории: [Технологии](#) | Теги: [google](#), [Казахстан](#), [машинный перевод](#), [промт](#)

Комментарии: -

Автор: [Konstantin Dranch](#)



Об этом с декабря 2012 года сообщают казахские СМИ, ссылаясь на Рауана Кенжеханулы, главу алматинского фонда Wikibilim.kz. Фонд занимается развитием свободно распространяемого контента на казахском языке: казахской Википедии, созданием открытой библиотеки и видеопортала. Его курирует руководитель администрации президента Казахстана Карим Массимов.

По словам г-на Кежеханулы (например, на [tengrinews.kz](#)), Google Translate и фонд ведут совместный проект по включению казахского языка в машинный переводчик, и «делается все возможное, чтобы сервис заработал до конца 2013 года». В настоящее время в систему загружают корпус двуязычных текстов, и любые заинтересованные стороны могут предоставить свои материалы, чтобы помочь этому процессу. Чем больше объем текста, и чем полнее соответствие перевода и оригинала, тем более высоким будет качество машинного перевода впоследствии. Глава Wikibilim также добавил, что Google не добавляет языки самостоятельно, для этого должна поступить инициатива от заинтересованной группы носителей.

В Казахстане уже есть ряд систем машинного перевода, например, Soylem ([audaru.kz](#)), работающий в паре с русским языком. Этому проекту уже около 6 лет, и как сообщили в компании Limeon, владельце сервиса, сейчас готовится к запуску версия в паре с английским. Soylem распространяется на коммерческой основе в форме розничных и корпоративных лицензий, а крупнейшим проектом по внедрению системы был Нацбанк. Кроме Soylem на рынке присутствуют местные разработки izet и summersoft, а также российский PROMT, но он доступен только для корпоративных заказчиков.

Включение казахского в Google Translate, если случится уже в этом году, серьезно повлияет на рынок машинного перевода. Появится возможность переводить на множество новых языковых пар, кроме того, Google известен большинству рядовых пользователей интернета. Если проект в Казахстане пройдет гладко, и качество машинного перевода окажется достаточным для понимания большинства текстов, это, вероятно, станет сигналом для запуска аналогичных инициатив в других странах СНГ. На данный момент в Google Translate из языков СНГ включены грузинский, украинский, белорусский и армянский.

В Soylem сообщили, что за разработкой систем машинного перевода уже обращались заинтересованные лица из Узбекистана, Таджикистана и Китая. Но разработка решений упирается в финансовый вопрос. Google Translate, по информации Mozgorilla, делает решение бесплатным, но требует от местных партнеров корпус текстов в 2 млн статей, собрать который без финансирования проблематично.

PS: Из-за динамичного развития СП в Казахстане, спрос на казахский язык в последние годы очень высок. Российские бюро Неотек, Янус, Логрус, Палекс и ABBYY LS открыли в Алма-Ате свои подразделения. На казахский стремятся переводить и зарубежные MLV, такие как Moravia, Lionbridge, но они пока что вынуждены действовать через российских провайдеров, так как не смогли найти локальных поставщиков с достаточно высоким уровнем качества.

Комментарий Юлии Епифанцевой, директора по развитию PROMT:

Этот проект направлен на то, чтобы с помощью краудсорсинга собрать корпус параллельных текстов. У нас есть большой опыт использования этой методики на онлайн-сервисе [Translate.Ru](https://translate.ru). Уже полтора года там действует функция «Предложи свой перевод», с помощью которой пользователь может добавить свой перевод для слова или целого текста. К сожалению, анализ ее использования нашими специалистами показывает, что краудсорсинговый материал малопригоден для создания переводчика или хотя бы его существенной настройки. Пользователи пишут с ошибками и опечатками, или делают перевод не полностью, или просто дурачатся, специально путая слова и значения. Однако мы продолжаем эксперимент, надеясь, что когда-нибудь количество перейдет в качество, и из этого материала можно будет выделить полезную информацию.

Переводчик с и на казахский язык сейчас очень востребован. В прошлом году мы выпустили продукт с казахским для корпоративного рынка, а сейчас продолжаем работы по его статистической надстройке, однако параллельные корпуса собираем профессиональными методами. Так что мы надеемся, что в скором времени сможем представить этот язык не только в коммерческих продуктах, но и на сервисе Translate.Ru.