

Разработка казахско-русского словаря поддерживающего морфологический анализ казахского теста

СЕРЖАНОВ АЗАТ БАХЫТЖАНОВИЧ

Новосибирский Государственный Университет (Новосибирск), Россия

e-mail: azatserzhan@gmail.com

Казахский язык – является одним из самых трудных тюркских языков, в котором есть девять частей речи, два существительного, четыре глагола, разбор частей речи и морфологии трудоемки процесс. Язык характеризуется большим числом словоформ для каждого слова, образованных путем добавления к его концу суффиксов и окончаний. Суффиксы относятся к семантической категории и при образовании новых слов часто изменяют части речи, к которой относится корневое слово или основа. Отсюда выходит нужда в автоматизированном определении частей речи, глагола, наречия, местоимения и. т. д. Дополнительный переводчик на русский язык позволяет увидеть точный перевод слово на русском языке.

Отсюда возникает нужда в словаре, с помощью которого можно получить не только перевод, но и морфологическую структуру слово. Разрабатывается алгоритм, который позволит автоматический определять суффиксы и их принадлежности к частям речи.

Исследуется предметная область и методы морфологического разбора и алгоритм нахождения корня слово. Существует формализация правил добавления окончаний, семантическую нейронную сеть, осуществляющую синтез словоформ казахского языка, что позволяет порождать структуру словаря начальных форм в виде синхронизированного линейного дерева.

Так же ожидается построение базы русско-казахских слов, интерфейса на языке C# и связка с базой данных.