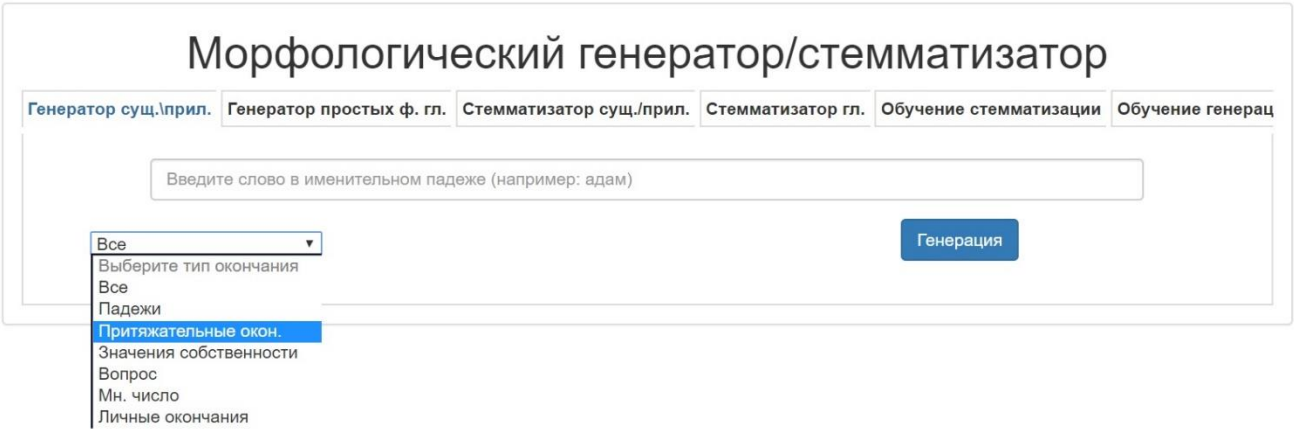
Программная система
«Стемматизация и генерация словоформ казахского языка»

АВТОРЫ: чл.-корр. РАН Федотов А.М., д.т.н. Барахнин В.Б., к.филол.н. Кожемякина О.Ю., Бакиева А.М., Лукпанова Л.Х.

Программная система решает задачу стемматизации и генерации словоформ изменяемых частей речи казахского языка: существительных, прилагательных и глаголов. Новизна программной системы заключается в том, что в её основе лежат оригинальные алгоритмы синтеза и анализа словоформ казахского языка, базирующиеся на принципах разбиения слов на флективные классы. Поскольку казахский язык является агглютинативным, подключать словарь словоформ для автоматизации морфологического анализа нецелесообразно. Значительно эффективнее пользоваться словарями аффиксов и наборами правил. С использованием предложенных алгоритмов создана база данных PostgreSQL, содержащая в себе все виды аффиксов казахского языка (в общей сложности более 5500).

Разработанные алгоритмы могут применяться на этапе морфологического анализа в поисковых системах, системах автореферирования и вопросно-ответных системах, системах автоматического анализа поэтических текстов, при построении тезаурусов и онтологий, а также для изучения морфологии казахского языка. Программная система реализована в форме веб-приложения, доступного по адресу http://db4.sbras.ru/morpher



ПУБЛИКАЦИИ:

1. Барахнин В.Б., Бакиева А.М., Бакиев М.Н., Тажибаева С.Ж., Батура Т.В., Лукпанова Л.Х. Стемматизация и генерация словоформ в казахском языке для систем автоматической обработки текстов // Вычислительные технологии. – 2017. – Т. 22. – № 4. – С. 11-21.
2. Барахнин В.Б., Федотов А.М., Бакиева А.М., Бакиев М.Н. Тажибаева С.Ж., Батура Т.В., Кожемякина О.Ю., Тусупов Д.А., Самбетбаева М.А., Лукпанова Л.Х. Алгоритмы генерации и стемматизации словоформ казахского языка // Cloud of Science. – 2017. – Т. 4. – № 3. – С. 434-449.

3. Barakhnin V.B., Fedotov A.M., Bakiyeva A.M., Bakiyev M.N., Tazhibayeva S.Zh., Batura T.V., Kozhemyakina O.Yu., Tussupov D.A., Sambetbaiyeva M.A., Lukpanova L.Kh. The software system for the study the morphology of the Kazakh language // The European Proceedings of Social & Behavioural Sciences. – 2017. – V. XXXIII. – P. 18-27.
4. Барахнин В.Б., Кожемякина О.Ю., Бакиева А.М., Содбоев М.К. Алгоритмы автоматизированной обработки поэтических текстов на казахском языке // Материалы II Международной научной конференции «Информатика и прикладная математика». Алматы, 27-30 сентября 2017 года. – Часть II. – С. 55-64.
5. Бакиева А.М. Морфологическая система «Стемматизация и генерация словоформ казахского языка» // Свидетельство о государственной регистрации программы для ЭВМ № 2018614456 от 06 апреля 2018 г.