

Разработка модуля предиктивного ввода текста для виртуальной клавиатуры операционной системы Sailfish

Е. А. Филиппова
efilippo@cs.petrso.ru

Аннотация

В работе представлен обзор существующих методов предиктивного ввода текста, преимущества и недостатки методов и их применимость к русскому и татарскому языкам. Целью работы является создание прототипа системы предиктивного ввода текста на основе N-грамм на русском и татарском языках для мобильной ОС Sailfish. Для достижения поставленной цели были составлены корпуса русского и татарского языков на основе текстов книг, статей, новостей. Разработаны алгоритмы для автодополнения вводимого пользователем слова и составления корпуса языка.

Введение

В связи с появлением мобильных устройств с сенсорным экраном появилась необходимость в создании систем предиктивного ввода. В большинстве случаев данные устройства обладают небольшим экраном, из-за чего ввод текста может быть затруднительным.

Прогнозирование следующего слова относится к задачам обработки естественного языка, где основное внимание уделяется пониманию и оптимизации взаимодействия человека и компьютера. Основная задача систем предиктивного ввода – сократить количество вводимых символов с клавиатуры: ускорив ввод текста с помощью подсказки готового варианта окончания слов или фраз. Подсказки новых слов, которые могут быть использованы, генерируются на основе имеющихся в словаре.

Существует несколько методов предиктивного ввода текста. В основном эти методы основаны на использовании словарей, содержащих слова и вероятности их встречи в тексте. В работе будет использоваться метод N-грамм, так он является простым в реализации и может быть применен к любому языку. Недостаток метода – большой объем памяти данных словарей, в связи с этим словари будут ограничены до триграмм.

Ввиду отсутствия клавиатур для Sailfish ОС с возможностью предиктивного ввода текста и поддержкой русского и татарского языков, тема данной работы является актуальной.

В статье представлен процесс разработки программного модуля для виртуальной клавиатуры Sailfish ОС. Для разработки модуля были поставлены следующие задачи:

- Анализ готовых решений с похожим функционалом;
- Поиск литературы для составления корпусов русского и татарского языков;
- Разработка программы для составления корпусов;
- Разработка программного модуля для виртуальной клавиатуры;
- Тестирование модуля.

В первом разделе содержится описание существующих методов предиктивного ввода текста и описывается процесс разработки модуля. Во втором разделе содержится описание и результаты тестирования созданного модуля.

1 Разработка модуля предиктивного ввода

1.1 Обзор методов предиктивного ввода текста

В подразделе содержатся методы предиктивного ввода, с описанием преимуществ и недостатков.

1.2 Описание алгоритмов

В подразделе описываются алгоритмы, используемые для предсказания окончания слова и следующего слова.

1.3 Хранение данных

В подразделе описывается структура данных для хранения словаря.

2 Тестирование

В данном разделе описаны данные (словари), которые использовались при тестировании, и полученные результаты. Сделаны выводы о работе алгоритма.

Заключение

В результате работы был создан модуль предиктивного ввода текста для виртуальной клавиатуры для ОС Sailfish. Недостатком разработанного алгоритма предиктивного ввода является наличие зависимости качества его работы от объема корпуса языка.

В дальнейшем для улучшения прогнозирования слова возможен переход к модели с использованием морфологических параметров.

Список литературы

- [1] Филатов С.Ю. Обзор методов предиктивного ввода. // Новые информационные технологии в автоматизированных системах. — 2017 — С. 55-61.
- [2] Martin J. H., Jurafsky D. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. — Prentice Hall, 2000.