

Text Suggestion. Отчет

Общее

В данной работе реализован веб-интерфейс для автодополнения текста с использованием библиотеки Reflex. Пользователь вводит текст в специальное поле, после чего система в реальном времени предлагает несколько вариантов продолжения. Подсказки формируются на основе n-граммной модели.

Архитектура:

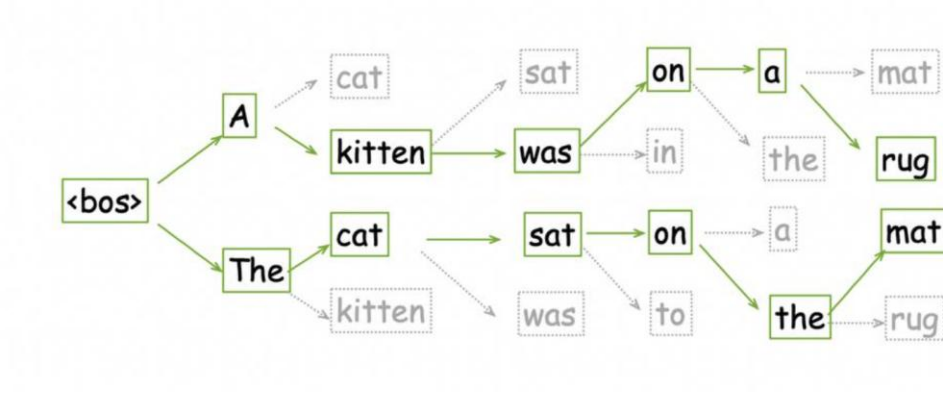
- WordCompleter — формирует словарь и префиксное дерево, а так же умеет быстро находить все возможные продолжения слова вместе с их вероятностями.
- NGramLanguageModel — простой подсчёт n-грамм (без сглаживания), возвращает распределение следующего слова по префиксу.
- TextSuggestion — комбинирует дополнение последнего слова (WordCompleter) и генерацию следующих слов n-граммой. Для генерации нескольких вариантов реализован beam search (параметр n_texts отвечает за количество вариантов на каждом шаге).

Реализация (Подробнее)

Реализация поддерживает фильтрацию дубликатов и сортировку предложений по вероятности. Для генерации нескольких возможных продолжений был использован подход Beam Search. На первом шаге генерируются слова-дополнения на основе нескольких предыдущих. Затем выбираются топ-2 (например, если n_texts=2) слова по вероятности. Далее для каждого из этих двух слов генерируются ещё по два слова-дополнения (при использовании контекста длины n). И выбираются топ-2 последовательности по вероятности. И так далее.

На каждом шаге сохраняются наиболее вероятные последовательности и их вероятности. Эти последовательности сохраняются в порядке убывания их вероятностей.

Beam Search



Интерфейс

Интерфейс включает текстовое поле и блок с подсказками. Подсказки отображаются в виде кнопок, продолжения упорядочены по вероятности. При нажатии на кнопку выбранное слово автоматически добавляется к тексту. Это позволяет пользователю быстро составлять фразы без необходимости полностью печатать каждое слово.

Text Suggestion

Введите текст, и система подскажет продолжение.

we

we celebrate

we celebrate the

we go

we celebrate the near

we go to

were

we celebrate the launching