Input Text Predictor

167

For this project, we will be building our own input text predictor, similar to ones you may have seen in Google Instant. We will build this input text predictor from a text corpus.

The steps involved in building this input text predictor are:

- Generate a list of **n-grams**, which is simply a list of phrases in a text corpus with their corresponding counts.
- Generate a **statistical language model** using the n-grams. The statistical language model contains the probability of a word appearing after a phrase.
- Create a **user interface** for the input text predictor, so that when a word or phrase is typed, the next word can be predicted and displayed to the user using the statistical language model.

We'll start the project by generating n-grams from a text corpus for this module. The second and third tasks above will be in later modules of this project. Click through when you are ready to begin.