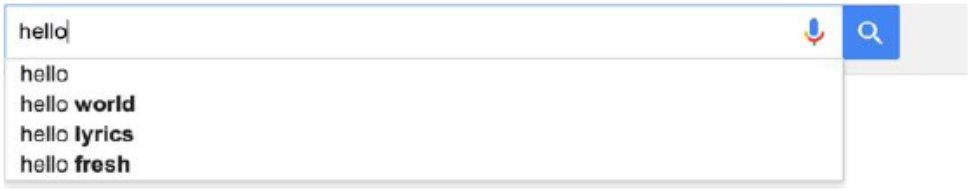
An auto complete example: Google Suggestion



Mathematical Theory: Probability of a word appearing after a phrase

$$\Pr(\text{word} \mid \text{phrase}) = \frac{\text{Count}(\text{phrase} + \text{word})}{\text{Count}(\text{phrase})}$$

Two Map Reduce:

NGramLibraryBuilder.java	To build the N-Gram based on input files.
LanguageModel.java	Build the language probabilities based on N-Gram.

Steps:

- Read a large-scale document collections
- Build n-gram library
- Calculate probability
- Run the project on Mapreduce

2-gram	
want to	200
eat apple	120
eat shit	1

want	to = 200
eat	apple = 120 shit = 1