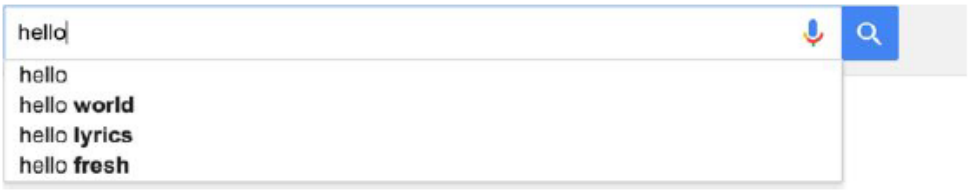
An auto complete example: Google Suggestion



Probability of a word appearing after a phrase

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

High-level design: 2 Map Reduce, one for building the N-gram, one for calculating the probabilities.

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