Chapter No. 12 Language Model for IR

Note: We have discussed the articles 12.1, 12.2, 12.3 and 12.4.

<Food for Thoughts>

1. Define a language model for IR.
2. Consider making a language model from the following training text:

   the martian has landed on the latin pop sensation ricky martin

   a. Under a MLE-estimated unigram probability model, what are P(the) and P(martian)?

   b. Under a MLE-estimated bigram model, what are P(sensation|pop) and P(pop|the)?

3. How Ponte and Croft experiments differentiate between different models of IR?
4. What are some of the problem of language model if we estimate document likelihood for a query? Illustrate with some arguments.