Chapter No. 11 Probabilistic Information Retrieval Model

Note: We have discussed the articles 11.1, 11.2, and 11.3 from this chapter.

<Food for Thoughts>

1. How do you define Probabilistic Information Retrieval Model? How it is better than vector space model?
2. Define the Probability relevance ranking? Explain what are its assumptions?
3. Define Binary Independence Model? Explain what are its assumptions?
4. Consider a corpus of three documents, which comprises of Vocabulary = {w_1 w_2 w_3 w_4 w_5}, assume that the subscript dictate order of dimension:

   d_1= {w_1 w_2 w_2 w_1 w_5}
   d_2= {w_3 w_2}
   d_3= {w_1 w_2}
   Relevant Documents = d1, d3
   Non-Relevant Document= d2
   q = {w_1 w_2 w_3}

   Using the Probability Ranking Principle (PRP) rank these documents. Using Prior from the given information. Assume a document d_4= {w_4 w_2 w_3} check whether it is relevant or not?
5. How relevance feedback mechanism different in probabilistic information retrieval from vector space model for IR?
6. Consider a corpus of three documents, which comprises of Vocabulary = {w_1 w_2 w_3 w_4 w_5}, assume that the subscript dictate order of dimension:

   d_1= {w_1 w_2 w_2 w_1 w_5}
   d_2= {w_3 w_2}
   d_3= {w_1 w_2}
   q = {w_1 w_2 w_3}

   Using the Probability Ranking Principle (PRP) rank these documents. As no relevance information is given, assume that d2 and d3 are relevant. Assume a new document d_4= {w_4 w_2 w_3} check whether it is relevant or not?