1. (1)0.22

   N=5, "Shenzhen" appears in doc 3, 4 and 5. So the DF is 3.

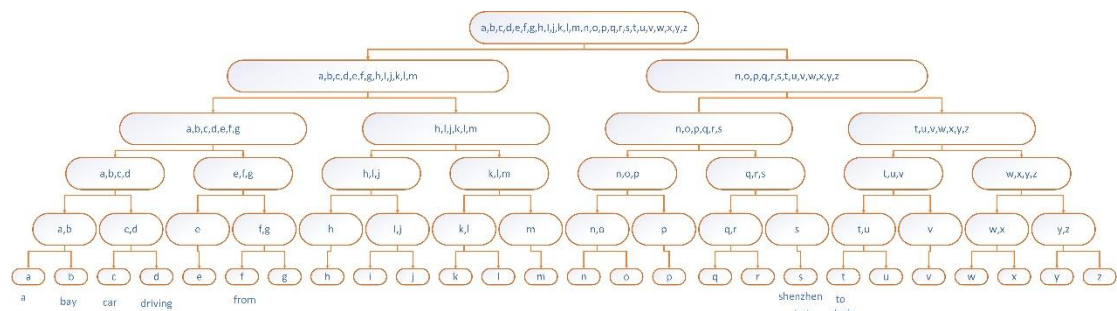   So IDF of "Shenzhen"=log(N/DF)=log(5/3) ≈0.22

   (2)0

   We can see the TF of "Shenzhen" in doc 1 is 0. So, TF-IDF of "Shenzhen" for doc 1=TF of "Shenzhen" ∗ IDF of "Shenzhen" =0∗0.22=0
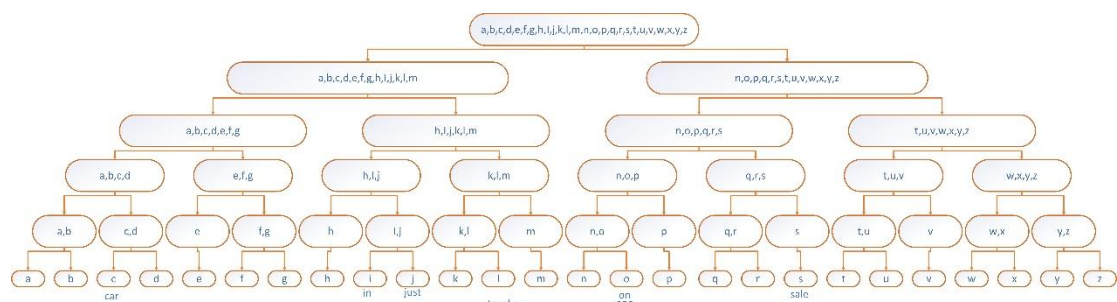
   (3)0.44

   We can see the TF of "Shenzhen" in doc 4 is 2. So, TF-IDF of "Shenzhen" for doc 4=TF of "Shenzhen" ∗ IDF of "Shenzhen" =2∗0.22=0.44

2. Doc1

   Doc4

3. the champion list of the word "Shenzhen"

   doc 3 doc 4

   the champion list of the word "car"

   doc 1 doc 2

4. (1)60%

   The number of relevant documents about solar system =500-200=300;

   And 500 documents are showed.

   Precision = The number of relevant documents / the total number =300/500=60%

   (2)30%

   The number of relevant documents about solar system =500-200=300;

   The total number of documents about solar system in the database=1000

   Recall =300/1000=30%

5.

1	a	doc3
2	on	doc1
	in	doc1 doc2 doc3
	of	doc3
3	car	doc2 doc3
4	sale	doc1 doc3
6	London	doc1 doc2
7	factory	doc2
	flowers	doc1
9	Guangzhou	doc2
12	construction	doc3