1.1 考虑如下几篇文档

文档1：search for new information

文档2：how to make Google your default search engine

文档3：new method for information retrieval

文档4：Google patents advanced search

a 画出该文档集的倒排索引

advanced	→	4		
default	→	2		
engine	→	2		
for	→	1	3	
Google	→	2	4	
how	→	2		
information	→	1	3	
make	→	2		
method	→	3		
new	→	1	3	
patents	→	4		
retrieval	→	3		
search	→	1	2	4
to	→	2		
your	→	2		

b 对于以下查询，给出返回结果

1). for AND (NOT method OR Google)

文档1

2). (search OR retrieval) AND information

文档1和文档3

1.2 为1.1中的文档构建双词索引（即二二元词索引）和位置信息索引。

双词索引：

advanced search	→	4
default search	→	2
for information	→	3
for new	→	1
Google patents	→	4
Google your	→	2
how to	→	2
information retrieval	→	3
make Google	→	2
method for	→	3
new information	→	1
new method	→	3
patents advanced	→	4
search engine	→	2
search for	→	1
to make	→	2
your default	→	2

位置索引：

Term	Position
advanced	4: 3;
default	2: 6;
engine	2: 8;
for	1: 2; 3: 3;
Google	2: 4; 4: 1;
how	2: 1;
information	1: 4; 3: 4;
make	2: 3
method	3: 2;
new	1: 3; 3: 1;
patents	4: 2;
retrieval	3: 5;
search	1: 1; 2: 7; 4: 4;
to	2: 2;
your	2: 5;

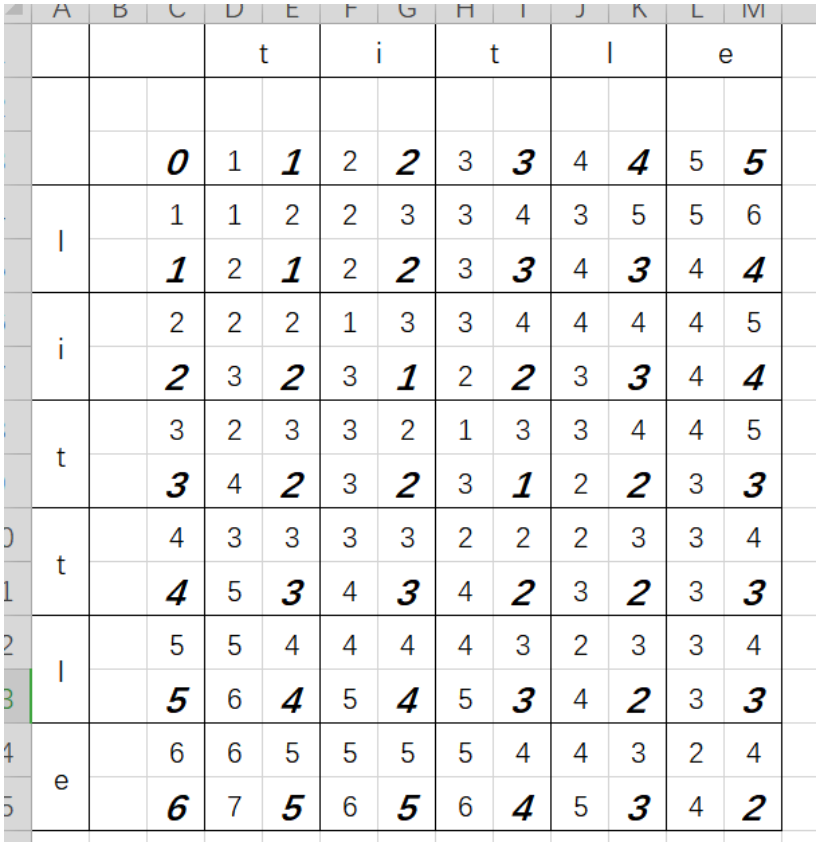
1.3 给出通配符查询 hy*er*sh 对应的2-gram索引转化而成的布尔查询，并给出一个错误解（即满足布尔查询却不满足通配符查询的解，不需要是正确的英文单词）

布尔查询： $h AND hy AND er AND sh AND h$

错误解：hhyeersh

1.4 计算单词little和title的编辑距离，并给出类似第四讲ppt第27页的计算过程。（要求严格参照）

编辑距离：2



		t		i		t		l		e	
		0	1	**1**	2	**2**	3	**3**	4	**4**	5	**5**
l		1	1	2	2	3	3	4	3	5	5	6
		1	2	**1**	2	**2**	3	**3**	4	**3**	4	**4**
i		2	2	2	1	3	3	4	4	4	4	5
		2	3	**2**	3	**1**	2	**2**	3	**3**	4	**4**
t		3	2	3	3	2	1	3	3	4	4	5
		3	4	**2**	3	**2**	3	**1**	2	**2**	3	**3**
t		4	3	3	3	3	2	2	2	3	3	4
		4	5	**3**	4	**3**	4	**2**	3	**2**	3	**3**
l		5	5	4	4	4	4	3	2	3	3	4
		5	6	**4**	5	**4**	5	**3**	4	**2**	3	**3**
e		6	6	5	5	5	5	4	4	3	2	4
		6	7	**5**	6	**5**	6	**4**	5	**3**	4	**2**

1.5 结合ppt上的两个倒排记录表合并算法伪代码，对于查询【x OR y】，给出一个合并算法。

倒排索引合并：

```
UNION(p1, p2):
answer <- <>
while p1 != NIL and p2 != NIL
do if docID(p1) < docID(p2)
    then ADD(answer, docID(p1))
         p1 <- next(p1)
    else if docID(p2) < docID(p1)
          then ADD(answer, docID(p2))
               p2 <- next(p2)
          else ADD(answer, docID(p2))
               p2 <- next(p2)
               p1 <- next(p1)

while p1 != NIL
do ADD(answer, docID(p1))
```

```
   p1 <- next(p1)

while p2 != NIL
do ADD(answer, docID(p2))
   p2 <- next(p2)

return answer
```