各个doc中没出现的词语：

1. D1: and, the, thing, ink, pink, links 5个
2. D2: this, one, I, think, called a, yink, wink, pink, links, 10个
3. D3: this, one, I, think, called, a, yink, thing, 8个

所有的词：this, one, I, think, is, called a, yink, he, likes, to, wink, drink, and, the, thing, ink, pink, links 共19个

He, likes, to, drink, and, the, thing is ink

1. Laplace

D1：

“drink”: ,

“wink drink”: ,

“pink ink”:

D2:

“drink”: ,

“wink drink”:

“pink ink”:

D3：

“drink”: ,

“wink drink”: ,

“pink ink”:

1. Lindstone:

D1：

“drink”: ,

“wink drink”: ,

“pink ink”:

D2:

“drink”: ,

“wink drink”: 0.016,

“pink ink”:

D3：

“drink”: ,

“wink drink”:

“pink ink”:

1. Lindstone:

D1：

“drink”: ,

“wink drink”: ,

“pink ink”:

D2:

“drink”: ,

“wink drink”: ,

“pink ink”:

D3：

“drink”: ,

“wink drink”:

“pink ink”:

两公式显然是不对称的，不同的形式计算结果不同。

从交叉熵的意义来看，交叉熵刻画的是一个查询被文档的编码所需的长度，即通过来表达的困难程度，代表正确答案，而代表的预测值。在检索的时候，给定具体的query，要找到更合适document，显然应该选择更小的对应的document，而不是去考察。