课前思考：

* 熵：在热力学中用来描述“能量退化”的物质状态参数之一，本质是揭示了系统内在的混乱程度。从信息论的角度来说，信息熵是信息量的期望，是把信息中排除了冗余后的平均信息量。从计算上看，将所有可能情况发生的概率乘以此种情况的不确定性再求和即可得到信息熵。
* KL散度：又称相对熵，是两个概率分布间差异的非对称性度量。在信息理论中，相对熵等价于两个概率分布的信息熵的差值。在信息理论中，相对熵是用来度量使用基于q的编码,来编码来自p的样本平均所需的额外的比特个数。
* 交叉熵：香农信息论中一个重要概念，主要用于度量两个概率分布间的差异性信息，衡量两个概率分布p，q之间的相似性，是计算语言学消岐的一种有效工具。  
  从计算的角度看，假设有两个概率分布p,q，其中p表示真实分布，q表示非真实分布，那么在相同的一组事件中，可以用非真实分布q来表示p发生所需要的平均编码长度。语言模型中，由于真实的分布p是未知的，可以利用通过训练集得到的模型的分布q来计算交叉熵，从而衡量这个模型在测试集上的正确率。
* 简而言之，  
  熵的意义是对A事件中的随机变量进行编码所需的最小字节数。  
  KL散度的意义是如果用B的编码来表示A“额外所需的编码长度”。  
  交叉熵指的是用B作为密码本来表示A时所需要的“平均的编码长度”

三者关系：KL散度 = 交叉熵 - 熵。为了让学到的模型分布更贴近真实数据分布，需要最小化 模型数据分布 与 训练数据之间的KL散度，而因为训练数据的分布是固定的，因此最小化KL散度等价于最小化交叉熵。一般交叉熵计算更为简便，故有如此代替的过程。

* 此外还有  
  KL散度和交叉熵的不同处：交叉熵中不包括“熵”的部分  
  KL散度和交叉熵的相同处：a. 都不具备对称性 b. 都是非负的

课堂思考：

1. 对于每个Query重新计算语言模型，是否会影响效率？

不是太影响。一方面，因为query较短，相对于相关文档的长度来说几乎“忽略不计”。另一方面，在更新MQ的过程中，只需要利用已知文档的MQ做简单的加法即可实现。（此外，系统中与文档集有关的计算往往更加影响效率，相比之下Query的影响可以忽略不计。）

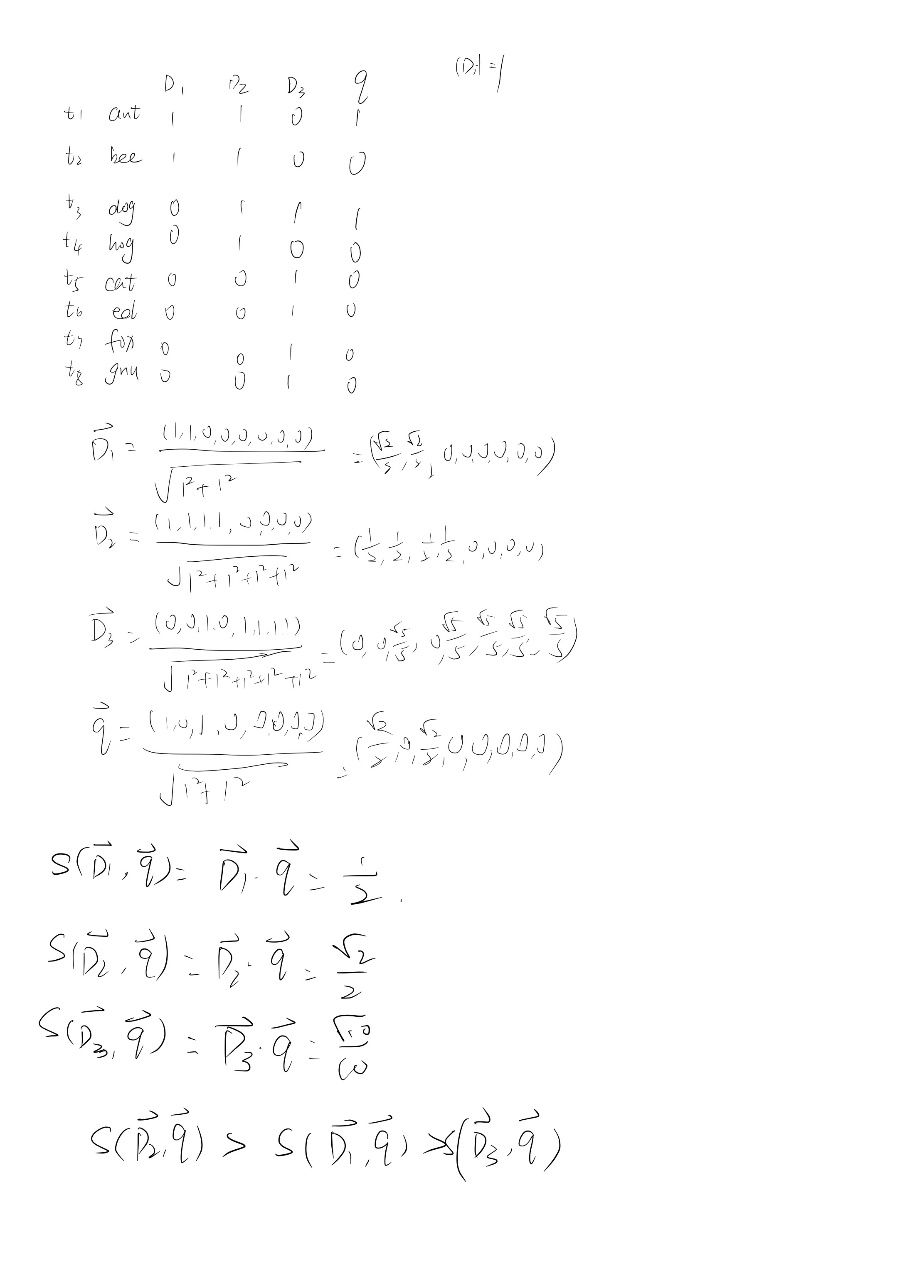
1. 为什么要求基向量线性独立，有什么好处？否则有什么坏处？

线性独立说明任何一个基向量不能被其余的基向量所“表示”。这样的选取办法，可以减少基向量的数量，减少向量的维度，继而减少计算量。否则，若不“独立”，那么说一些向量可由其他基向量表示，那么这样的基向量没有存在的意义，并且每个向量的表示方式将“不唯一”。->信息冗余

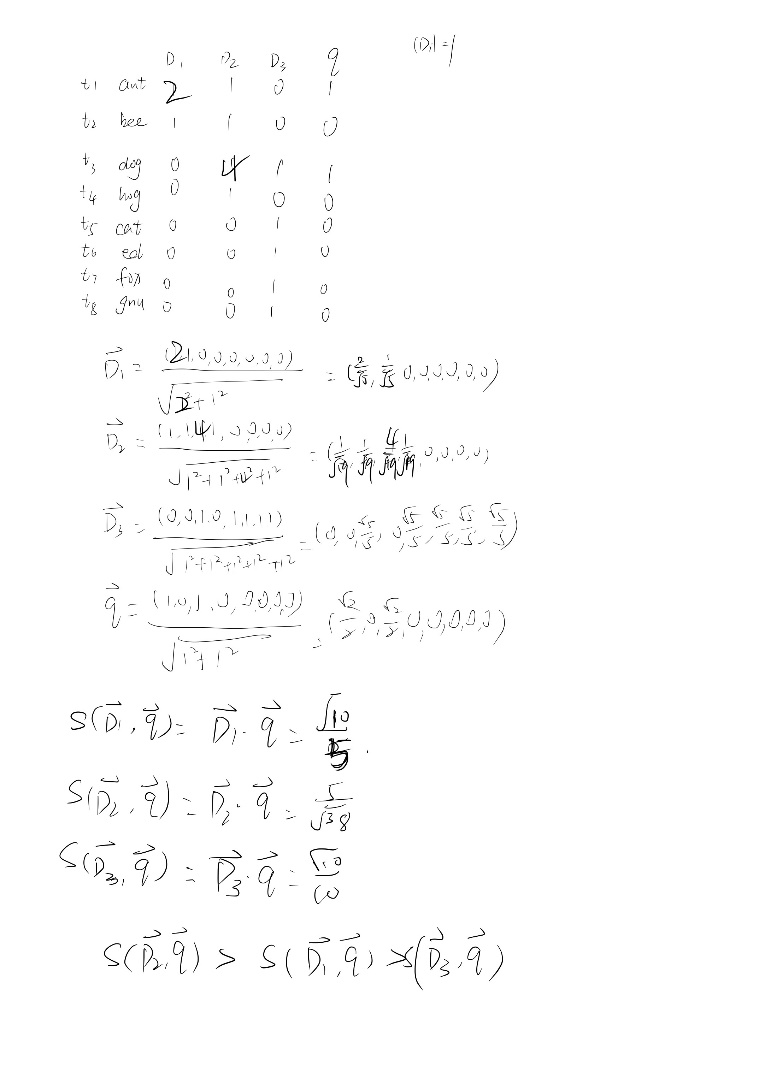
1. +1的作用？

若每个文档都出现了w，那么N/df=1。此时，若不+1，那么log（N/df）= 0，这个权重显然不合适。

1. 利用VSM计算相似度。



**Boolean**



tf