

比较布尔模型、向量空间模型和概率模型，描述这三者的异同之处。

● 相同点：

向量空间模型、概率模型，都是布尔模型的扩展。主要体现在：在表示文档时，三种模型都使用到了“关联矩阵”，都可以将每一篇文档（document）和询问（query），表示成若干个词项（term）的向量，并以此进行匹配、比较、排序、检索

● 不同点：

1. 向量的内容不同：
 - a) 布尔模型中，每个文档向量的各个值写入的是 0 或 1，表示这篇文档中是否有某个词项；
 - b) 向量空间模型中，各项的取值有多种取法，如，与布尔模型相同的二值取法、词频（tf）、词频、文档频率倒数之积（tf-idf）等，以及他们的变形（如 Lucene 等）；
 - c) 概率模型中的取值与布尔模型相同，也可选择使用 BM25。
2. 文档向量各项（维度）的选取也是不一样的：
 - a) 布尔模型、概率模型选取各个词项作为各个维度；
 - b) 向量空间模型中，除此之外还有多种选法，如，选取“核心语义”等。
3. 向量在检索中起到的作用不同：
 - a) 布尔模型中，直接使用各文档的向量与 query 进行精确匹配，返回符合要求的文档；
 - b) 向量空间模型中，往往需要各个文档向量与 query 向量分别计算余弦相似度（或类似的相似度度量），并在排序返回更加匹配的文档；
 - c) 概率模型中，向量仅仅是对文档的基本表示，在 BIR 检索中，需要针对不同的 query，对 i 个此项词项得到全局概论 p_i , q_i ，利用贝叶斯公式、对数运算等得到各个文档的判别函数，经过若干次迭代后取得返回结果。
4. 关注点不同：
 - a) 布尔模型、向量空间模型，都使用了一个假设，即“用相似代替相关”，立足点是文本的相似，并基于此进行匹配、检索
 - b) 概率模型的关注点是“相关性”，因而使用概率的工具进行建模和检索
5. 特点：
 - a) 布尔模型，匹配的原理简单关注的是精确匹配；但是返回的结果没有顺序的先后之分。
 - b) 向量空间模型，最大的特点在于其灵活多变，对于向量的具体形式、以及检索的实现没有严格的规定，可以进行充分创造，如 LSI 模型
 - c) 概率模型，最显眼的地方在于使用了概率的工具，逐步推导建模，并且考虑了错误检索的代价。