2022夏_第一次作业

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

b. 对于一下查询，给出返回结果

1). for AND (NOT method OR Google)

2). (search OR retrieval) AND information

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。

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）

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）

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。