本部分内容描述对于原始情报库的全文检索和建立自定义分类主题语料库功能模块的规划

——superhy

1. 建立情报库全文索引并管理索引（胡杨，刘卓然）

从MongoDB中原始情报库中取得一定量的内容：article和所有reply的组合起来成为字符串；取出MongoDB中的postUrlMD5字段和collectionName字段用于唯一标记到MongoDB中的一条记录。Ps：MongoDB里的数据读出之后需要经过处理加工，形成整块字符串，再加上之前的postUrlMD5和collectionName，组合成索引建立的原料。

建立索引库，每次将article和reply的组合字符串作为content域值，不做存储，但是需要分词，postUrlMD和collectionName作为同名的域值，不分词，但是需要存储在索引中（index的file文件夹必须严格事先设定好）。进行多线程建立索引，注意线程同步。

索引库的管理工作，索引库添加新文档，添加新索引库，从原来的索引库中删除文档，清空整个索引库，等操作。其中，索引需要建立相关mysql表进行管理，对索引的操作需要建立相关日志表记录操作日志。

1. 从索引当中检索文章内容（胡杨）

根据存储的索引信息和自身的搜索需求，选择对特定的索引进行检索，需要做到支持短语（分词后按布尔值或的关系检索）和多个关键词传入（布尔值或的关系检索）

检索结果需要获得索引域值存储的两个字段（collectionName和postUrlMD）以及对于检索结果的评分score，用map容器存储，返回list集合

1. 搜索结果的展示和主题分类语料库的建立（胡杨，戴丹）

主题分类实现创建，需要用户输入一个或多个关键词或者短语，进行检索，按照2的步骤检索出结果，根据返回的list<map>集合，级联查询出MongoDB中对应的记录进行展示，对勾选的记录，可以操作存入对应的主题分类语料库，将勾选记录的语料内容取出，存入本地磁盘文本文档（file目录严格事先设定好）。

对于语料库中的每一条语料都在mysql数据库中建立信息管理机制，语料库可以随时手动添加文档和从中删除文档（使用java操作文档的建立和删除），对应还需要建立分类主题的对应信息表，称为分类类别表。分类类别表在用户决定建立该主题的时候就建立了，如果取消建立，就删除已经建立的主题。

以上描述可能需要建立的数据库表如下：

1. 索引信息表
2. 主题语料表
3. 主题类别表
4. \*索引操作日志表
5. \*主题预料操作日志表