**索引管理模块任务详细设计及内核API调用说明**

——胡杨 2014.6.24

1. **操作简述**

数据库中会有索引信息表，tbl\_index，自己根据pd图理解索引信息表的字段含义，file文件夹中，会有子文件夹lucene\_all\_index，这个存放的是索引的磁盘文件，tbl\_index表中已经存了一个字段（130），对应原始情报库总索引lucene\_all\_index。

索引信息表的展示：索引各字段信息以列表的形式展示，setId级联到经典分类主题表显示所属分类名称，indexPath不用显示。

限定一个时间只能对一条索引记录进行操作，所以列表中每个记录前面设置一个单选按钮，选择单选框，下方显示对应的索引控制面板。

控制面板需要设定如下按钮：

1. 新建：下拉菜单加载出tbl\_analyzer表中所有的analyzerName供选择，选择后，新建按钮会将对应目录下的索引文件全部从磁盘删除，然后完全新建一份索引文件在指定路径下，对应地，tbl\_index中相应的记录也要将修改时间更新为当前时间（无需另外编写），重建索引的内核方法为：rebuildIndexCompletely()【1】
2. 彻底移除索引文件：将索引目录中的索引文件全部从磁盘上删除，对应地，tbl\_index中的相应记录删除（无需另外编写），删除索引内核方法为：dropIndexFromDis()【2】
3. 清空索引中的所有文档：仅仅是清空索引中的文档，索引文件不删除，对应，tbl\_index中相应的记录需要修改（无需编写），清空索引中的所有文档的内核方法为：deleteAllDocFromIndex()【3】
4. 按照原始情报库中的站点删除站点对应文档：添加下拉菜单，可以选择mongodb中所有存在的collectionName，选择后，点击删除，删除对应内容在索引中的文档，mysql表修改（无需编写），对应内核方法：deleteSiteDocFromIndex()【4】
5. 回复索引回收站中的所有文档：恢复索引回收站中已经被删除的所有文档到索引当中，恢复后，mysql表修改（无需编写），对应方法：recoverAllDocInRecycle()【5】
6. 增量刷新新任务的内容进入索引：增量将新开展的采集任务所下载的内容更新入索引中，mysql表修改（无需编写），对应方法：refreshNewTaskDocIntoIndex()【6】
7. **容器说明**

（略）

1. **内核方法说明**

【1】：package ims.nlp.lucene.index class IndexDocManagment 传入：分词器名称，如果不传入，系统会自动加载默认分词器。

【2】：package ims.nlp.lucene.index class IndexDocManagment 传入：无

【3】：package ims.nlp.lucene.index class IndexDocManagment 传入：无

【4】：package ims.nlp.lucene.index class IndexDocManagment 传入：tbl\_site表中的nickName

【5】：package ims.nlp.lucene.index class IndexDocManagment 传入：无

【6】：package ims.nlp.lucene.index class IndexDocManagment 传入：无

PS：在新建class IndexDocManagment类的对象之后需要调用initIndexEntity(Index index)初始化这个类要操作的Index实体，传入index实体，才可以进行以上所有的操作，不然会报错。

1. **其他**

所有操作直接调用串行实现，无需多线程

1. **界面设计**

简单模型，界面自行调整

任务接收者：戴丹