文件系统总名称为file，系统部署时，file文件夹存放在tomcat目录的bin文件夹中

在file文件夹下有以下子文件夹

analyzer\_data:

存放分词器的配置信息，如：词库，子文件夹名为分词器名。

lucene\_all\_index:

存放原始情报库所有内容的lucene索引，不包含任何子文件夹。

lucene\_product\_index:

存放所有不良产品内容的lucene索引，不包含任何子文件夹。

nlp\_all\_content:

敏感信息过滤结果总母体库，所有识别过滤得出的敏感信息以txt的形式存入该文件夹，不包含任何子文件夹。txt文件的文件名命名规范：mongodb中对应节点的nodeId\_所属站点collectionName(识别时间)，形式如下：

E459B32514D59A3F8CEFAF76C73EBD9E-0\_maoputietie(2014-08-29 00:01:00)

nlp\_classifier\_test:

文本分类待预测文档库，其中初始有一个子文件夹名为unlabeled，其中存放所有待预测文本txt文件。分类任务结束，立即清空。

nlp\_train\_content:

文本分类训练语料库，其中初始化有若干文件夹，文件夹名为经典分类类别名拼音全拼，当从nlp\_all\_content母体库中导入训练文档时，母体库也保存一份存档，两者文件名一样，用户自己上传的文件也会在两处都存一份。

nlp\_polarity\_test:

极性分类待测试文件夹，初始有子文件夹unlabeled，存放待测试文本文件，任务开始自动从其它地方导入，任务结束，自动清空。

nlp\_lingpipe:

lingpipe工具包支持库，初始有子文件夹model和train\_sentiment\_analysis，model存放模型，train\_sentiment\_analysis存放两个子文件夹neg和pos，分别存放消极训练预料和积极训练预料各1000篇。

nlp\_weka:

weka工具包支持库，初始有子文件夹analyzed\_classifying\_content，存放分词后的待测试文档，子文件夹analyzed\_train\_content，存放分词后训练文档，子文件夹classifying\_arff，存放待预测weka专用数据集，子文件夹train\_arff，存放训练weka专用数据集，子文件夹model，存放分类模型。