技术需求说明

一、指标说明

1、比例完整句比例FULLSEN。每份年报完整句在文章中所有句子中的（完整句的定义是含有主谓结构的句子）

2、基础词汇所占比例WORDLIST。每份年报中基础词汇表的词语在所有词汇中的比例（基础词汇见 hsk123uniq三级基础词汇）  
3、平均笔画数STROKES。每份年报文章中汉字的平均笔画数  
4、笔画数为5的字符比例COUNT5。文章中笔画数为５的字符比例

5、笔画数为12的字符比例COUNT12

6、笔画数为22的字符比例COUNT22

7、笔画数为25的字符比例COUNT25

二、数据最后格式

三年深交所主板年报和汉语考试三级基础词汇已提供，请生成的数据最后按照excel附件字段格式（如附件excel文件字段显示）

三、其他

请把代码和应用的技术简介也发我一份（思路和其中应用的方法工具等，比如用的什么分词器等等），谢谢啦！

如果有什么需要沟通的请随时联系我

电话15800369519

微信 cyshuu