实验课作业具体事项

**注意：**

*我们将对提交的实验结果进行源代码查重，如果发现雷同，抄袭者与被抄者一律零分处理，请大家独立完成实验课内容。*

# 实验课内容与时间安排

|  |  |  |  |
| --- | --- | --- | --- |
| **实验内容** | **日期** | **开始时间** | **结束时间** |
|
| **分词、词频统计** | **2014.11.22** | **08:00** | **12:00** |
| **Bigram语言模型构建与存储** | **2014.11.29** | **08:00** | **12:00** |
| **数据平滑** | **2014.12.06** | **08:00** | **12:00** |
| **拼音汉字自动转换** | **2014.12.13** | **08:00** | **12:00** |

# 实验输入输出要求

## 一元词频统计：

输出一个词与频数对照列表，word\_freq.txt，词与频数之间用空格或者制表符切分，每个词及其频数占一行。

格式要求：

Word1 freq1

Word2 freq2

…

## Bigram频度统计

输出一个Bigram与频数对照列表，Bigram\_freq.txt， Bigram与频数之间用空格或者制表符切分，每个Bigram及其频数占一行。

格式要求：

Bigram1 freq1

Bigram2 freq2

…

## 拼音汉字自动转换：

该部分实验内容要求Demo展示，生成可执行文件。

在执行可执行文件的时候要求输入拼音串，系统输出是该拼音串对应的可能性最大的汉字串，如

输入： yi zhi mei li de xiao hua

输出： 一只美丽的小花

# 实验结果提交说明

实验课结果提交分两次，第一次提交前两次课的实验内容，第二次提交所有的实验内容。

## 第一次实验结果提交：

截止时间：2014.12.07

提交内容：一元词频统计表，Bigram频度统计表

## 第二次实验结果提交：

截止时间：2014.12.14

提交内容：实验源代码，一元词频统计表，Bigram频度统计表，音字转换的可执行文件

**请大家把以上内容打包，压缩包文件名格式为：学号-姓名**

## 实验提交方式

实验课共享百度云盘：

帐号： [statisticalnlp@163.com](mailto:statisticalnlp@163.com)

密码： zryycl

请大家在截止时间之前把实验结果提交到该云盘。

## TA联系方式

何彬

Email: [hebin\_hit@foxmail.com](mailto:hebin_hit@foxmail.com)

实验室： 综合楼803

大家对于实验课内容或者在实验过程中有疑问的地方，可以联系实验课TA