**东北大学自然语言处理实验室**

**词性频次统计**

 

东北大学自然语言处理实验室

<http://www.nlplab.com>

NiuTrans团队

[niutrans@mail.neu.edu.cn](mailto:niutrans@mail.neu.edu.cn)

©2014 NiuTrans

词性频次统计

* **问题描述**

给定经过分词、词性标注之后的中文文本，对中文词性频次进行统计，词性频次统计是在整篇文档中进行统计。

* **示例数据**

中文： 新华社/NR 乌鲁木齐/NR $date/JJ 电/NN

新华社/NR 乌鲁/NR $date/JJ 电/NN

新华社/NN 乌鲁/NR $date/JJ 电/NN

新华社/NR 乌鲁木齐/NR $date/JJ 电/NN

* **任务目标**

1. 统计数据中出现不同词汇的个数；

如：新华社 出现4次，乌鲁出现2次……

1. 统计出每个中文词汇被标注成不同词性的次数

如 新华社 被标注为 NR 出现3次 ， 被标注为 NN 出现 1次……

* **测试数据**

当前文件夹sample-data\目录下

**祝好运!**