

# 东北大学自然语言处理实验室



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)

## 重排序

- 问题描述

在机器翻译中，重排序是指利用其他评价方式对翻译结果进行排序。在本题目中，我们要对给定的单语语料进行词频统计，根据词频对每句话进行排序、打分

问题 1：统计语料库中所有词的词频（包括标点符号）

问题 2：根据词频为每句话进行打分、排序（如果句子中的词都是频率很高的词）

- 示例

语料：

I am Chinese .

The weather is good today !

...

即：

第一句的打分： $f_I + f_{am} + f_{Chinese} + f_{.}$

第二句的打分： $f_{The} + f_{weather} + f_{is} + f_{good} + f_{today} + f_{!}$

$f$  表示每个词的频率

- 数据

当前文件夹 sample-data\目录下的英文单语语料 e。

祝好运！