

# 自动化计算文本文件TTR的bash脚本

## 简介

这是一个可以计算文本文件 TTR 的bash脚本，文件名为：`calculate_TTR.sh`。它会接收一个文件名作为参数，并输出总单词数、特异单词数和 TTR。

## 脚本代码

```
#!/bin/bash

# 读取文件
filename="$1"
if [ ! -f "$filename" ]; then
    echo "文件不存在或无法读取！"
    exit 1
fi

# 去除标点符号和空格，只保留单词
words=$(cat "$filename" | tr -cs '[:alnum:]' '\n*' | tr '[:upper:]' '[:lower:]')

# 计算总单词数量和不同单词数量
total_words=$(echo "$words" | wc -w)
unique_words=$(echo "$words" | sort | uniq | wc -w)
# -w或--words 只显示词数。
# -l或--lines 显示行数。

# 计算TTR
ttr=$(echo "scale=4; $unique_words/$total_words" | bc)
# -c或--bytes或--chars 只显示Bytes数。

# 输出结果
echo "Total: $total_words"
echo "Unique: $unique_words"
echo "TTR: $ttr"
```

## 计算代码

```
chmod 777 calculate_TTR.sh
./calculate_TTR.sh BROWN_A.txt
./calculate_TTR.sh pubmed-a.txt
```

## 结果

```
15:24:56 || TTR.Comparison > ./calculate_TTR.sh BROWN_A.txt
Total: 91063
Unique: 11927
TTR: .1309
15:28:45 || TTR.Comparison > ./calculate_TTR.sh pubmed-a.txt
Total: 1032476
Unique: 31436
TTR: .0304
```