


# 简繁汉字转换系统（单机版） 使用说明

本系统向用户提供了单机版简繁汉字转换工具。该工具无需安装，可直接在 Windows 命令行界面下运行。

注意：由于简繁汉字转换系统需要大规模语言模型，因此本工具要求在 64 位操作系统下运行（建议采用 Microsoft Windows7 x64 或 Microsoft Windows Server 2008 R2 x64），而且机器必须装有 8G 以上内存。

## 1. 运行

打开 windows 命令提示符，切换到单机版简繁汉字转换工具所在目录，输入 jf，根据相关提示，进一步设置相关参数，即可使用简繁汉字智能转换系统的全部功能（如图 1 所示）。



```
命令提示符
G:\>cd jf
G:\jf>jf
jf v0.10 Copyright (C) Mandel Shi 2005-2014 (mandel@xmu.edu.cn)
简体汉字到繁体转换程序

jf [options] *txt
-a FreqFile  annotate the conversion table with frequency info
-ac convert simplified Chinese to ancient Chinese
-d print debug info
-c command line mode (j->f only)
-jf1 $ $ is a simplified string to be converted
-fj1 $ $ is a traditional string to be simplified
-jf File1 File2  convert simplified Chinese File1 to traditional File2
-fj File1 File2  convert traditional Chinese File1 to simplified File2
-jfx File1 File2 语委 xml 文件测试, 仅输出转换字结果
-jfx2 File1 File2 语委 xml 文件测试, 输出转换字和简繁对照句
-lm FILE 采用指定的语言模型
-2 BiGRAM_FILE use context bigram 翻译模型
-log FILE 把检查信息输出到这个文件
-sd DIR specifies the segtag directory
-w WORD_FILE do word conversion
-mk DIR move f/j files to DIR (k=j or f)
-ck check text files to see if they are simplified or traditional
-ckf FILE start checking from this file for a big directory

e.g. jf -mf f *.txt
如果不指明转换, 则是检查能否正确转换指定的繁体文件

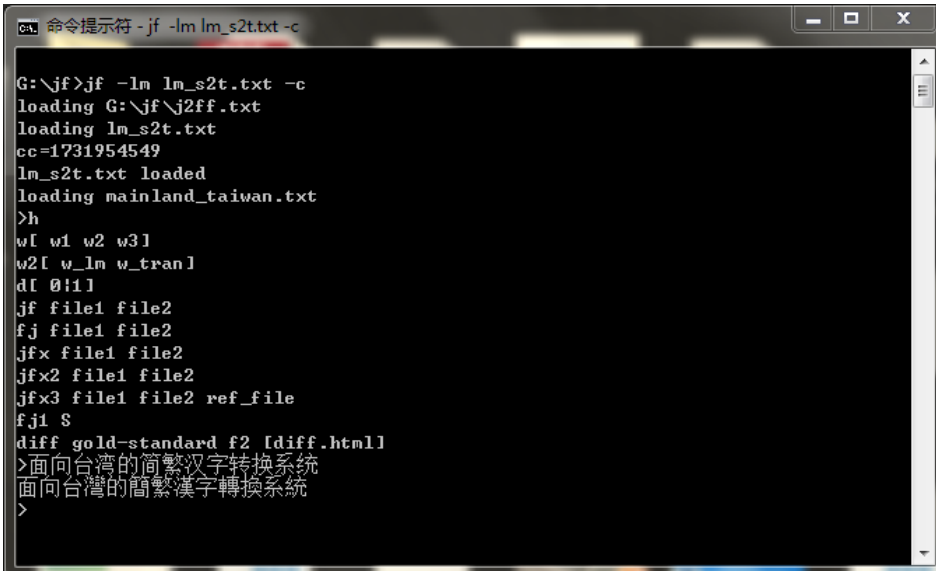
G:\jf>
```

图 1

## 2. 命令行模式

工具中提供了命令行模式供用户使用，目前仅支持简转繁。具体操作方法：

(1) 输入 `jf -lm lm_s2t.txt -c` 进入命令行模式（如图 2 所示，其中 `lm_s2t.txt` 为语言模型文件；由于语言模型规模较大，加载需要等待几分钟时间）。



```
命令提示符 - jf -lm lm_s2t.txt -c
G:\jf>jf -lm lm_s2t.txt -c
loading G:\jf\j2ff.txt
loading lm_s2t.txt
cc=1731954549
lm_s2t.txt loaded
loading mainland_taiwan.txt
>h
w[ w1 w2 w3]
w2[ w_lm w_tran]
d[ 0:1]
jf file1 file2
fj file1 file2
jfx file1 file2
jfx2 file1 file2
jfx3 file1 file2 ref_file
fj1 s
diff gold-standard f2 [diff.html]
>面向台湾的简繁体字转换系统
面向台湾的簡繁體字轉換系統
>
```

图 2

(2) 在命令行模式下，输入 `h` 可查看使用提示（如图 2 所示）。

(3) 在命令行模式下，输入简体文本，命令行界面返回转换后的繁体文本（如图 2 所示）。

## 3. 文件转换

工具中提供了文件转换功能供用户使用，支持简转繁和繁转简。具体操作方法：

(1) 文件简转繁：输入 `jf -jf file1.txt file2.txt -lm lm_s2t.txt`（其中 `file1.txt` 为简体原文文件，`file2.txt` 为输出转换结果的目标文件名，`lm_s2t.txt` 为语言模型文件）。

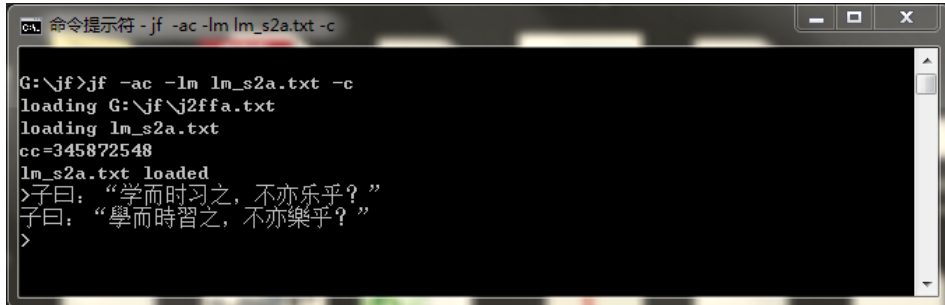
(2) 文件繁转简：输入 `jf -fj file1.txt file2.txt -lm lm_s2t.txt`（其中 `file1.txt` 为繁体原文文件，`file2.txt` 为输出转换结果的目标文件名，`lm_s2t.txt` 为语言模型文件）。

## 4. 面向古籍的简繁转换

工具中提供了面向古籍的简繁转换功能供用户使用。与普通转换（面向台湾

的简繁转换)使用方法相似,只需在命令中带上选项-ac。具体操作方法:

(1) 命令行模式:输入 `jf -lm lm_s2a.txt -c -ac` 进入面向古汉语的命令行模式(如图 3 所示,其中 `lm_s2a.txt` 为古汉语语言模型文件;由于语言模型规模较大,加载需要等待几分钟时间)。



```
ca. 命令提示符 - jf -ac -lm lm_s2a.txt -c
G:\jf>jf -ac -lm lm_s2a.txt -c
loading G:\jf\j2ffa.txt
loading lm_s2a.txt
cc=345872548
lm_s2a.txt loaded
>子曰:“学而时习之,不亦乐乎?”
子曰:“學而時習之,不亦樂乎?”
>
```

图 3

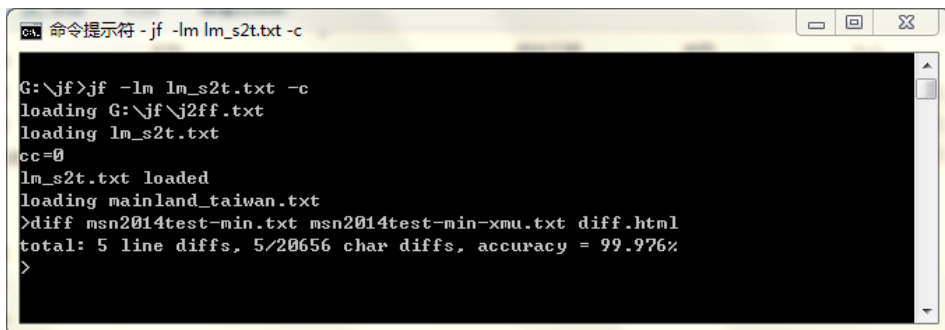
(2) 文件简转繁:输入 `jf -jf file1.txt file2.txt -lm lm_s2a.txt` (其中 `file1.txt` 为简体原文文件, `file2.txt` 为输出转换结果的目标文件名, `lm_s2a.txt` 为古汉语语言模型文件)。

(3) 文件繁转简:输入 `jf -fj file1.txt file2.txt -lm lm_s2a.txt` (其中 `file1.txt` 为繁体原文文件, `file2.txt` 为输出转换结果的目标文件名, `lm_s2a.txt` 为古汉语语言模型文件)。

## 5. 评测功能

本工具提供了对简繁转换结果进行评测的功能,具体操作方法如下:

(1) 按前文第 2 节介绍的方式进入命令行模式。  
(2) 在命令行模式下,输入 `diff gold.txt test.txt diff.html` 即可进行评测(如图 4 所示),其中 `gold.txt` 为标准答案, `test.txt` 为待评价的转换结果, `diff.html` 为指定的对比结果输出文件。



```
ca. 命令提示符 - jf -lm lm_s2t.txt -c
G:\jf>jf -lm lm_s2t.txt -c
loading G:\jf\j2ff.txt
loading lm_s2t.txt
cc=0
lm_s2t.txt loaded
loading mainland_taiwan.txt
>diff msn2014test-min.txt msn2014test-min-xmu.txt diff.html
total: 5 line diffs, 5/20656 char diffs, accuracy = 99.976%
>
```

图 4

(3) 用浏览器打开 `diff.html`,即可查看系统输出的对比结果(如图 5 所示),点击页面上方的链接“跳到与原文不一致的字表”,自动定位到页面下方的统计结果(如图 6 所示)。

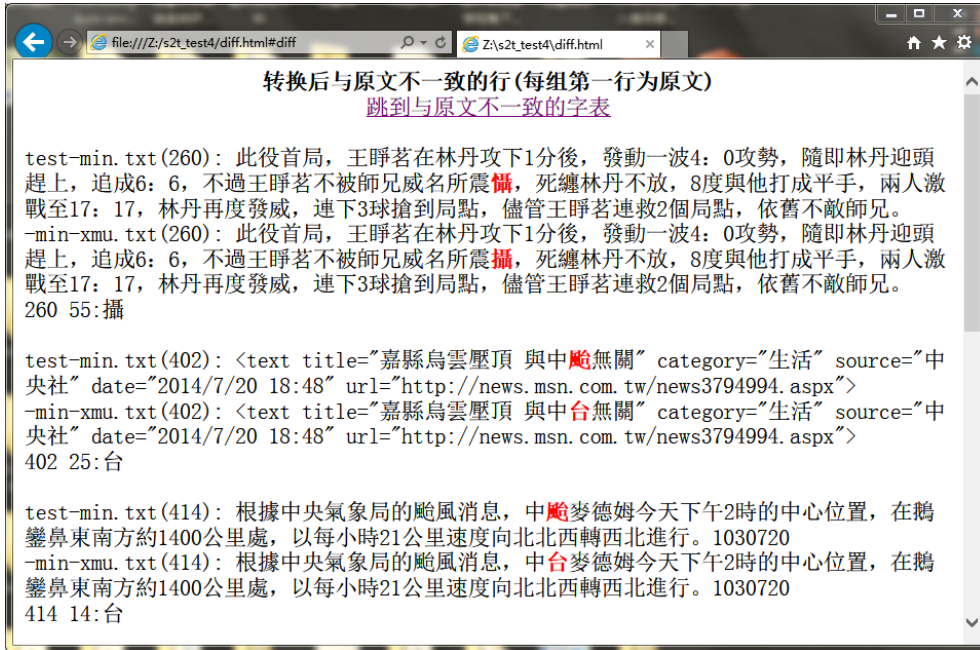


图 5

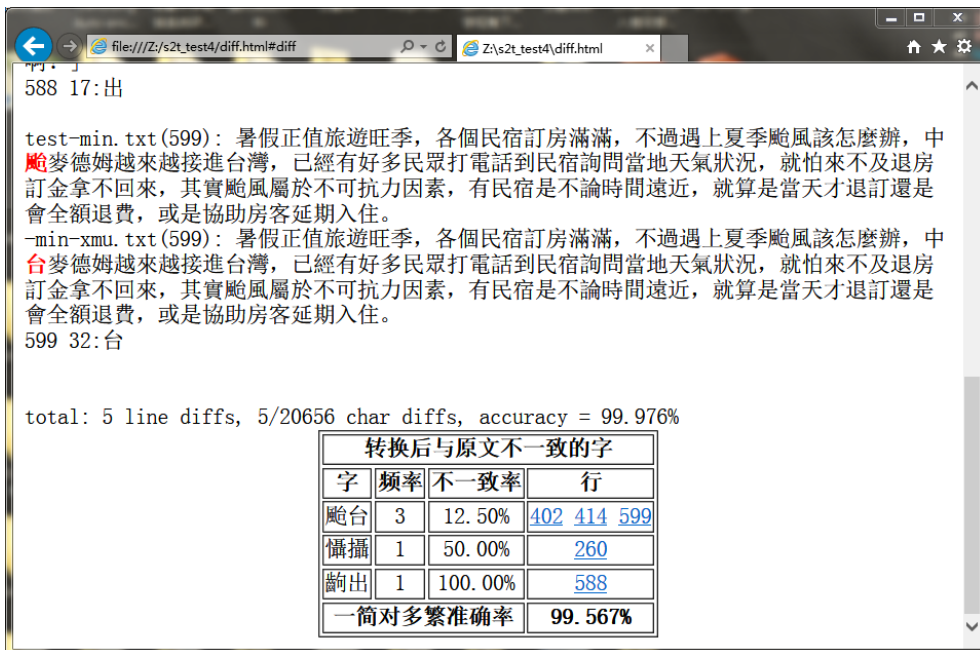


图 6