|  |
| --- |
| iflytek |
| 现网短信过滤工具 |
| 资源文件的格式和作用 |
|  |
| **cmlu** |
| **2012-3-22** |

# sms.txt

该文件存放应该识别的短信，通常运行该工具经过分析后，将现网应该识别但没有识别的短信放在该文件中。

每条短信占一行。

每行的格式：

Usercontent\tbizname\toperaName

# talking.txt

该文件存放talking问候语，作为拒识短信的一种分类。

每条短信占一行。

# smsrefuse.txt

其它类型的拒识短信，作为talking.txt的补充。

# hanzi.txt

存放汉字，每字占一行，每行其格式如下：

的\t1588562

汉字后面的数字越大，代表在文本中出现的概率越大。

不出现在此文件中的汉字将认为是乱码。

# Trad2Simp.txt

代表繁简体表，每行的格式如下:

亀\t龟

\t左边为繁体字，右边为简体字。