# 词的分类

* 实词：名词、动词、形容词、状态词、区别词、数词、量词、代词
* 虚词：副词、介词、连词、助词、拟声词、叹词。

# ICTPOS3.0词性标记集

n 名词  
 nr 人名  
 nr1 汉语姓氏  
 nr2 汉语名字  
 nrj 日语人名  
 nrf 音译人名  
 ns 地名  
 　nsf 音译地名  
 nt 机构团体名  
 nz 其它专名  
 nl 名词性惯用语  
 ng 名词性语素  
  
t 时间词  
　　tg 时间词性语素  
  
s 处所词  
  
f 方位词  
  
v 动词  
 vd 副动词  
 vn 名动词  
 vshi 动词“是”  
 vyou 动词“有”  
 vf 趋向动词  
 vx 形式动词  
 vi 不及物动词（内动词）  
 vl 动词性惯用语  
 vg 动词性语素  
a 形容词  
 ad 副形词  
 an 名形词  
 ag 形容词性语素  
 al 形容词性惯用语  
b 区别词  
 bl 区别词性惯用语  
z 状态词  
r 代词  
 rr 人称代词  
 rz 指示代词  
 rzt 时间指示代词  
 rzs 处所指示代词  
 rzv 谓词性指示代词  
 ry 疑问代词  
 ryt 时间疑问代词  
 rys 处所疑问代词  
 ryv 谓词性疑问代词  
 rg 代词性语素  
m 数词  
 mq 数量词  
q 量词  
 qv 动量词  
 qt 时量词

# 虚词

d 副词  
p 介词  
 pba 介词“把”  
 pbei 介词“被”  
c 连词  
 cc 并列连词  
u 助词  
 uzhe 着  
 ule 了 喽  
 uguo 过  
 ude1 的 底  
 ude2 地  
 ude3 得  
 usuo 所  
 udeng 等 等等 云云  
 uyy 一样 一般 似的 般  
 udh 的话  
 uls 来讲 来说 而言 说来  
  
 uzhi 之  
 ulian 连 （“连小学生都会”）  
  
e 叹词  
y 语气词(delete yg)  
o 拟声词  
h 前缀  
k 后缀  
x 字符串  
 xx 非语素字  
 xu 网址URL  
w 标点符号  
 wkz 左括号，全角：（ 〔 ［ ｛ 《 【 〖 〈 半角：( [ { <  
 wky 右括号，全角：） 〕 ］ ｝ 》 】 〗 〉 半角： ) ] { >  
 wyz 左引号，全角：“ ‘ 『  
 wyy 右引号，全角：” ’ 』  
 wj 句号，全角：。  
 ww 问号，全角：？ 半角：?  
 wt 叹号，全角：！ 半角：!  
 wd 逗号，全角：， 半角：,  
 wf 分号，全角：； 半角： ;  
 wn 顿号，全角：、  
 wm 冒号，全角：： 半角： :  
 ws 省略号，全角：…… …  
 wp 破折号，全角：—— －－ ——－ 半角：--- ----  
 wb 百分号千分号，全角：％ ‰ 半角：%  
 wh 单位符号，全角：￥ ＄ ￡ ° ℃ 半角：$

# ICTCLAS 汉语词性标注集

代码|名称|帮助记忆的诠释 ----|----|-------------- Ag|形语素|形容词性语素。形容词代码为a，语素代码ｇ前面置以A。 a|形容词|取英语形容词adjective的第1个字母。 ad|副形词|直接作状语的形容词。形容词代码a和副词代码d并在一起。 an|名形词|具有名词功能的形容词。形容词代码a和名词代码n并在一起。 b|区别词|取汉字“别”的声母。 c|连词|取英语连词conjunction的第1个字母。 Dg|副语素|副词性语素。副词代码为d，语素代码ｇ前面置以D。 d|副词|取adverb的第2个字母，因其第1个字母已用于形容词。 e|叹词|取英语叹词exclamation的第1个字母。 f|方位词|取汉字“方” 的声母。 g|语素|绝大多数语素都能作为合成词的“词根”，取汉字“根”的声母。 h|前接成分|取英语head的第1个字母。 i|成语|取英语成语idiom的第1个字母。 j|简称略语|取汉字“简”的声母。 k|后接成分| l|习用语|习用语尚未成为成语，有点“临时性”，取“临”的声母。 m|数词|取英语numeral的第3个字母，n，u已有他用。 Ng|名语素|名词性语素。名词代码为n，语素代码ｇ前面置以N。 n|名词|取英语名词noun的第1个字母。 nr|人名|名词代码n和“人(ren)”的声母并在一起。 ns|地名|名词代码n和处所词代码s并在一起。 nt|机构团体|“团”的声母为t，名词代码n和t并在一起。 nz|其他专名|“专”的声母的第1个字母为z，名词代码n和z并在一起。| o|拟声词|取英语拟声词onomatopoeia的第1个字母。 p|介词|取英语介词prepositional的第1个字母。 q|量词|取英语quantity的第1个字母。 r|代词|取英语代词pronoun的第2个字母,因p已用于介词。 s|处所词|取英语space的第1个字母。 Tg|时语素|时间词性语素。时间词代码为t,在语素的代码g前面置以T。 t|时间词|取英语time的第1个字母。 u|助词|取英语助词auxiliary 的第2个字母,因a已用于形容词。 Vg|动语素|动词性语素。动词代码为v。在语素的代码g前面置以V。 v|动词|取英语动词verb的第一个字母。 vd|副动词|直接作状语的动词。动词和副词的代码并在一起。 vn|名动词|指具有名词功能的动词。动词和名词的代码并在一起。 w|标点符号| x|非语素字|非语素字只是一个符号，字母x通常用于代表未知数、符号。 y|语气词|取汉字“语”的声母。 z|状态词|取汉字“状”的声母的前一个字母。

# jieba 字典中出现的类型

a  
 ad  
 ag  
 an  
b  
c  
d  
 df  
 dg  
e  
f  
g  
h  
i  
j  
k  
l  
m  
 mg  
 mq  
n  
 ng  
 nr  
 nrfg  
 nrt  
 ns  
 nt  
 nz  
o  
p  
q  
r  
 rg  
 rr  
 rz  
s  
t  
 tg  
u  
 ud  
 ug  
 uj  
 ul  
 uv  
 uz  
v  
 vd  
 vg  
 vi  
 vn  
 vq  
x  
y  
z  
 zg

# simhash 过滤的掉的词汇

c  
e  
h  
k  
o  
p  
u  
 ud  
 ug  
 uj  
 ul  
 uv  
 uz  
x  
y