jiebaR\_wordseg

iebaR包以RCpp接口用C++直接编写，无需rJava，并且算法优异，支持第三方词典。

还有就是它支持R 3.12。具体包信息可以参考：

<http://qinwenfeng.com/jiebaR/basic/README.html>

第三方词典可以先下载搜狗的细胞词库：<http://pinyin.sogou.com/dict/>

然后用深蓝词库转换工具进行转换（目前是2.0）：

<https://code.google.com/p/imewlconverter/downloads/list>

转换的目标词典类型为“无拼音纯汉字”方案，就是一个utf-8编码的txt文档。

jiebaR已经发布到CRAN上面，可以直接正常安装和调用了：

require(jiebaR)

## Loading required package: jiebaR  
## Loading required package: jiebaRD

#初始化分词器，默认的词典在jiebaRD包的安装路径下，只需要额外指定用户词典即可  
cutter <- worker(type = "mix",   
 user = "D:/download/finance.dict",  
 detect=T, symbol = F,  
 lines = 1e+05, output = NULL)  
  
test<-"应收账款周转率太低了，存货积压也太大，  
 这直接影响了公司的股价和市盈率，董事会需要立即作出行动，  
 改变市场战略，加大产品创新，争取对股东有个交代"  
#<=就是分词符号  
seg1<-cutter<=test  
  
seg1

## [1] "应收账款" "周转率" "太低" "了" "存货" "积压"   
## [7] "也" "太大" "这" "直接" "影响" "了"   
## [13] "公司" "的" "股价" "和" "市盈率" "董事会"   
## [19] "需要" "立即" "作出" "行动" "改变" "市场"   
## [25] "战略" "加大" "产品" "创新" "争取" "对"   
## [31] "股东" "有个" "交代"

cutter2<-worker()  
  
seg2<-cutter2<=test  
  
seg2

## [1] "应收" "账款" "周转率" "太低" "了" "存货" "积压"   
## [8] "也" "太大" "这" "直接" "影响" "了" "公司"   
## [15] "的" "股价" "和" "市盈率" "董事会" "需要" "立即"   
## [22] "作出" "行动" "改变" "市场" "战略" "加大" "产品"   
## [29] "创新" "争取" "对" "股东" "有个" "交代"