**BCC语料库介绍**

**网址**：<http://bcc.blcu.edu.cn>

**简介**：是以汉语为主、兼有其他语种的语言大数据，目标是为语言本体研究提供一个使用简便的在线检索系统和构建大数据的语言应用平台。BCC支持云服务，通过API调用方式开展知识抽取、模型构建等研究和应用工作提供便利。

**教程：**

（一）在线检索，即在浏览器内使用BCC，输入检索式，以页面形式返回结果。

1.统计：BCC检索式中可以包含词性或短语类型，也可以带通配符。



2.筛选：提供二次检索的功能，即在现有的返回结果中保留或者剔除符合检索式的语料实例，得到二次结果。

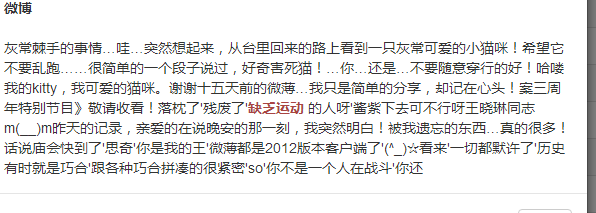
3.下载：下载检索或统计结果



4.高级：可以设置返回结果的显示形式；可以随机生成实例，如设定上下文显示字数，设定是否以句形式显示结果。



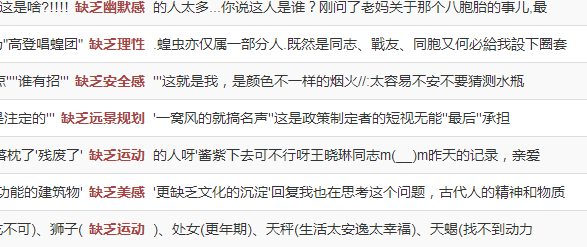
5.全文：点击该按钮可以查看检索实例更多的上下文。



云服务，通过编程使用BCC的Web API接口形式来调用BCC。

（二）BCC检索式

（1）汉字串（或者词串）



（2）属性符号

C:\Users\ADMINI~1\AppData\Local\Temp\WeChat Files\28d37ba770d291ce7d181f2444378dc.png

（3）通配符“.”、“@”、“~”





（4）集合符号“[]”



（5）离合符号“\*”

（6）属性约束符号“/”



（7）空格或者“+”

**相关论文：**

[1]荀恩东 饶高奇 肖晓悦 臧娇娇. 大数据背景下BCC语料库的研制[J].语料库语言学,2016.