



DCC 2.0 动态流通语料库 —文心语料库检索平台的研制

朱君辉^{1,2,3} 刘鑫^{1,2,3} 杨麟儿^{1,2,3} 师佳璐^{1,2,3} 刘鹏远^{1,2} 杨尔弘^{1,3}

¹国家语言资源监测与研究平面媒体中心

²北京语言大学 信息科学学院

³北京语言大学 语言资源高精尖创新中心

2022年11月28日

研究背景

北京大学CCL语料库

国家语委现代汉语平衡语料库

北京语言大学BCC语料库

DCC 1.0 动态流通语料库



□ 贡献:

- 语料规模大
- 语料选材广
- 多领域语料
- 历时大数据

□ 局限性:

DCC 2.0 动态流通语料库

- 缺乏对深层的句法结构检索需求的支持

✓ 功能特色

- ✓ 支持远距离、深层次的句法结构的检索需求
- ✓ 不同检索模式兼顾检索功能的全面性与用户友好性

DCC 1.0 动态流通语料库

词义演化 新闻历时 小说历时 论文历时 多领域历时

改革



报纸频率关系
来源：国家语言资源监测与研究平面媒体中心



报纸频率

报纸频度

分类频率

分类频度



DCC 2.0 动态流通语料库

DCC 1.0 动态流通语料库



DCC 2.0 动态流通语料库

- 全领域、多覆盖的语料资源
- 语料深加工与多级标注
- 强大的检索功能

DCC 2.0 动态流通语料库

(一) 全领域、多覆盖的语料资源

- 目前已入库**新闻报刊**与**二语教材**语料两种领域的语料。
 - **新闻报刊**：主要来自DCC 1.0 动态流通语料库，时间跨度为1946年至2020年，长达70余年。
 - **二语教材**：收录500多册市面上广泛使用的国际汉语教材，包括《博雅汉语》《成功之路》等。

语料类型	字节数/字	句数/句
报刊语料	1,600,000,000	39,500,000
二语教材语料	5,376,330	243,071

- 丰富的元信息：所在文档ID、所在文档标题、所在文档来源、所在文档日期等。
- 文学、微博、口语、学术论文将陆续上线……

DCC 2.0 动态流通语料库

(二) 语料深加工与多级标注

词性标注

PN NT VV AS PN CD M P NR NN DEG NN PU
他 昨天 送 了 我 一 本 关于 中国 历史 的 书 。

命名实体识别

XXXX-XX-XX NUMBER COUNTRY
他 昨天 送了我 1 本关于 中国 历史的书。

依存句法标注

PN NT VV AS PN CD M P NR NN DEG NN PU
他 昨天 送 了 我 一 本 关于 中国 历史 的 书 。

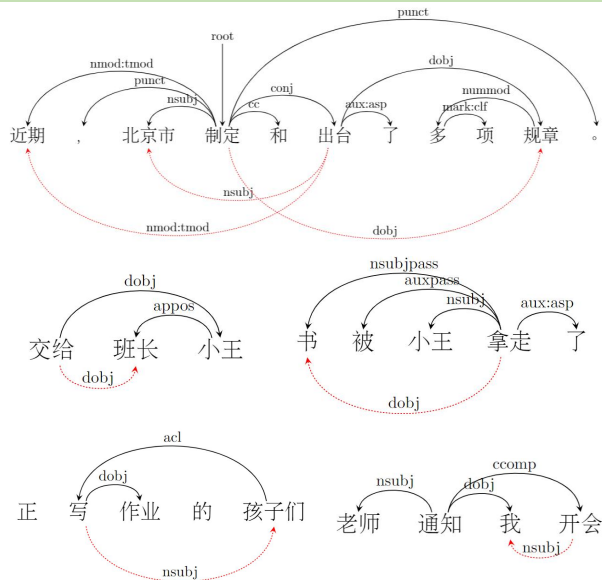
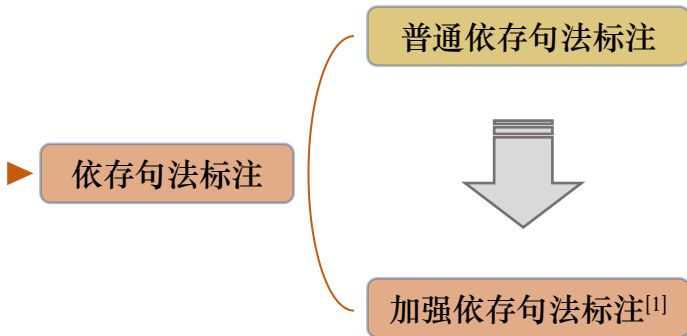
nsbj nmod:tmod aux:asp dobj punct dobj nummod case nmod:assmod nmod case

词汇等级标注

— — — — — — 四 — 四 — —
他 昨天 送 了 我 一 本 关于 中国 历史 的 书

DCC 2.0 动态流通语料库

(二) 语料深加工与多级标注



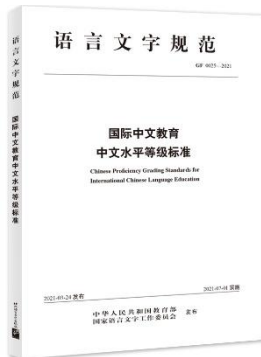
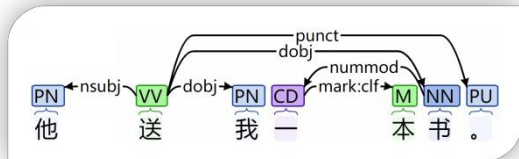
通过增加、修改词语之间的依存关系，使得依存关系可以同时体现句子的句法结构和语义关系。

[1]余婧思, 师佳璐, 杨麟儿, 肖丹, 杨尔弘. 汉语增强依存句法自动转换研究. 第二十一届中国计算语言学大会 (CCL 2022)

DCC 2.0 动态流通语料库

(三) 强大的检索功能

- 支持正则表达式；
- 支持复杂的检索表达式查询（如依存句法、命名实体、词汇难度查询等）；
- 支持按出现频次对规定的捕获内容进行统计；
- 支持通过模式检索，查询与给定例句相同句法结构的句子；
- 支持利用元信息继续检索；
- 支持从网页上下载查询结果（txt 文件）。



DCC 2.0 动态流通语料库



DCC 2.0 动态流通语料库

关于我们 帮助 旧版

普通检索

模式检索

请输入内容

Q 搜索

☒ 报刊 ☐ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

普通检索:

检索包含“学习”的句子: [word=学习]

检索包含“学习”且“学习”为动词的句子: [word=学习&tag=VV]

检索包含“学习”且“学习”为动词,“学习”后有宾语成分的句子: [word=学习&tag=VV]>dob[]

捕获“学习”后面接HSK二级词: [word=学习](?<学习后接的词>[level=二])

模式检索:

检索含有“越...越...”结构的句子,并捕获依存于“越”的词: 人\$越(多)\$越(热闹)

检索含有介词“对”后跟谓语的结构的句子,并捕获依存于“对”的词: 这\$对我(有)(帮助)

检索包含“没有.....就没有.....”句型的句子,并捕获“没有”后衔接的词: \$没有(水)\$就\$没有生命的存在。

捕获

元信息

使用帮助

帮助手册

ICALL
智能语言学习



检索模式

- **普通检索**：用于检索单个项以及有顺序和间隔规定的多个项。
- **模式检索**：用于检索与给定例句结构相同的句子。

普通检索

➤检索单个项

➤检索多个项

➤通过依存句法结构来检索

➤捕获

普通检索

• 单个项的检索



文心语料库检索平台

普通检索

模式检索

学习

Q 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

± 结果下载

① 最近几年,在一些国家,血型性格的关系已经受到了越来越多的人的关注,特别是一些年轻人,他们经常通过血型来判断一个人的思维方式、性格、脾气、工作、**学习**、生活甚至爱情上的特点,这已经成为年轻人休闲聊天的普遍话题。

① O型的人很乐观,在**学习**、工作、爱情上追求理想的结果。

① 你也**学习**汉字吗?

① 是的,老师告诉我如果要学好汉语,就一定要**学习**汉字。

① 所以我现在一边**学习**会话,一边学习汉字。

① 所以我现在一边学习会话,一边**学习**汉字。

① 1956年,20岁的丁肇中从台湾回到父母的母校美国密西根大学**学习**数学和物理学,他仅用了5年多时间就获得物理。

共有200条结果



文心语料库检索平台

普通检索

模式检索

[word=学习]

Q 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

± 结果下载

① 最近几年,在一些国家,血型性格的关系已经受到了越来越多的人的关注,特别是一些年轻人,他们经常通过血型来判断一个人的思维方式、性格、脾气、工作、**学习**、生活甚至爱情上的特点,这已经成为年轻人休闲聊天的普遍话题。

① O型的人很乐观,在**学习**、工作、爱情上追求理想的结果。

① 你也**学习**汉字吗?

① 是的,老师告诉我如果要学好汉语,就一定要**学习**汉字。

① 所以我现在一边**学习**会话,一边学习汉字。

① 所以我现在一边学习会话,一边**学习**汉字。

① 1956年,20岁的丁肇中从台湾回到父母的母校美国密西根大学**学习**数学和物理学,他仅用了5年多时间就获得物理。

共有144条结果

普通检索

- “或”——可供选择的列表（“|”）
- “不过”或“可是”或“只是”



文心语料库检索平台



文心语料库检索平台

普通检索 模式检索

不过可是只是

Q 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

± 结果下载

① 可是,怎么送给他们呢?"

① 可是租金太贵了。

① 可是在浦东,太远了。

① 可是这套比那两套小,是一室一厅。

① 可是如果我们只看到眼前实际的生活,忽视了除了吃饱以外,还需要一些比较抽象的东西来满足我们的的好奇心,那么生活还有什么乐趣呢?

① 我们反正跑不过熊啊!"

① 狮子知道,如果它跑不过最慢的羚羊,它就会饿死。

共有200条结果

普通检索 模式检索

[word=/不过可是只是/]

Q 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

± 结果下载

① 可是,怎么送给他们呢?"

① 可是租金太贵了。

① 可是在浦东,太远了。

① 可是这套比那两套小,是一室一厅。

① 可是如果我们只看到眼前实际的生活,忽视了除了吃饱以外,还需要一些比较抽象的东西来满足我们的的好奇心,那么生活还有什么乐趣呢?

① 我们反正跑不过熊啊!"

① 狮子知道,如果它跑不过最慢的羚羊,它就会饿死。

共有200条结果

普通检索

- 按照词性进行检索（采用宾州中文树库CTB的分词词性标签集）
 - NN表示普通名词

普通检索

高级检索

Q 搜索

☒ 报刊
 ☐ 文学
 ☐ 微博
 ☐ 口语
 ☐ 教材
 ☐ 论文

结果下载

❶ 鸟虫篆,是滥觞于周秦时期仅见于**印章**及偶见于铜器、铜剑上的文字。

❶ 鸟虫篆,是滥觞于周秦时期仅见于印章及偶见于**铜器**、铜剑上的文字。

❶ 鸟虫篆,是滥觞于周秦时期仅见于印章及偶见于铜器、**铜剑**上的文字。

❶ 鸟虫篆,是滥觞于周秦时期仅见于印章及偶见于铜器、铜剑上的**文字**。

❶ **鸟虫篆**印的基调接近清丽静逸,而又别致得像花腔女高音,正是印人想换换“口味”而出现的一道新鲜佳肴。

❶ 鸟**虫**篆印的基调接近清丽静逸,而又别致得像花腔女高音,正是印人想换换“口味”而出现的一道新鲜佳肴。

❶ 鸟虫篆印的**基调**接近清丽静逸,而又别致得像花腔女高音,正是印人想换换“口味”而出现的一道新鲜佳肴。

CTB词性标签集

标记	解释	标记	解释	标记	解释
AD	副词	DEG	“的”	JJ	区别词
AS	助动词	DER	“得”	LB	“被”“给”等
BA	“把”、“将”	DEV	“地”	LC	方位词
CC	并列连词	DT	限定词	M	量词
CD	基数词	ETC	“等”、“等等”	MSP	VP前的“所”等
CS	从属连词	FW	其他语言的词	NN	普通名词
DEC	“的”词性标记	IJ	叹词	NR	专有名词

<http://hunter.blcuicall.org>

CTB词性标签集

标记	解释	标记	解释
NT	时间名词	SB	“被，给”
OD	序数词	SP	句尾小品词
ON	拟声词	VA	表语形容词
P	介词	VC	系动词
PN	代词	VE	“有”“无”“没有”
PU	标点	VV	其他动词

普通检索

- 如果一个句子当中有多个满足检索式的成分，会显示在不同行中。

The screenshot displays the 'Wenxin Language Library Search Platform' (文心语料库检索平台). The search bar contains the query '[tag=NN]'. Below the search bar, there are radio buttons for different document types: '报刊' (Newspapers), '教材' (Textbooks), '微博' (Weibo), '口语' (Oral), '文学' (Literature), and '论文' (Theses). The '教材' (Textbooks) option is selected. A green button labeled '结果下载' (Download Results) is visible. The search results are displayed in a list of sentences, each preceded by a small icon. The sentences are: '路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了？', '给你们打电话,总没有人接。', '给你们打电话,总没有人接。', '我们聊天去了,聊到半夜才回房间睡觉,真有意思。', '跟门口的那位,我听人家都叫他侃爷,我想他姓侃,是侃大爷。', '跟门口的那位,我听人家都叫他侃爷,我想他姓侃,是侃大爷。', and '跟门口的那位,我听人家都叫他侃爷,我想他姓侃,是侃大爷。'. The interface also includes a '检索结果' (Search Results) section on the right, which currently shows 'No Data' with the ICALL logo. The footer of the page indicates 'Copyright © BLCU-ICALL 2021'.

文心语料库检索平台

帮助 反馈

普通检索 模式检索

[tag=NN] 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

结果下载

检索结果

捕获 元信息

ICALL
No Data

共有200条结果

Copyright © BLCU-ICALL 2021

普通检索

- 词性上仍可使用“或”关系
 - NN表示普通名词、NR表示专有名词、NT表示时间名词
 - [tag=NN][tag=NR][tag=NT]
 - [tag=/NN|NR|NT/]



文心语料库检索平台



文心语料库检索平台

普通检索 模式检索

[tag=NN][tag=NR][tag=NT] 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

± 结果下载

1 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

2 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

3 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

4 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

5 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

6 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

7 给你们打电话,总没有人接。

8 给你们打电话,总没有人接。

<http://hunter.blcuicall.org>

共有200条结果

普通检索 模式检索

[tag=/NN|NR|NT/] 搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

± 结果下载

1 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

2 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

3 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

4 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

5 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

6 路易斯和张明在路易斯的房间里聊天昨天晚上你们去哪儿了?

7 给你们打电话,总没有人接。

8 给你们打电话,总没有人接。

共有200条结果

普通检索

- 支持正则表达式
 - 在CTB中，名词只有NN、NR、NT三种，因此可用/N.*/代替。

普通检索

高级检索

Q 搜索

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

结果下载

OPEC今天将在维也纳召开为期两天的会议,讨论是否将目前两千八百五十万桶的官方配额,每天提高五十万桶。

OPEC今天将在维也纳召开为期两天的会议,讨论是否将目前两千八百五十万桶的官方配额,每天提高五十万桶。

OPEC今天将在**维也纳**召开为期两天的会议,讨论是否将目前两千八百五十万桶的官方配额,每天提高五十万桶。

OPEC今天将在维也纳召开为期两天的**会议**,讨论是否将目前两千八百五十万桶的官方配额,每天提高五十万桶。

OPEC今天将在维也纳召开为期两天的会议,讨论是否将**目前**两千八百五十万桶的官方配额,每天提高五十万桶。

OPEC今天将在维也纳召开为期两天的会议,讨论是否将目前两千八百五十万桶的**官方**配额,每天提高五十万桶。

OPEC今天将在维也纳召开为期两天的会议,讨论是否将目前两千八百五十万桶的官方**配额**,每天提高五十万桶。

普通检索

- “且”——对词和词性都有限制（&）
 - 介词“对”

普通检索

高级检索

[word=对&tag=P]

搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

原能总署二日提议,会员国在三十天内释出六千万桶的战略储油,抑制因卡崔娜飓风重创美国墨西哥湾岸石油产地后,对油价产生的负面效应。

很明显,自然延伸原则与日本坚持的中间线建议不符,对日方不利,因此日本不会接受。

社论说,国际仲裁虽然对中国有利,但是中国也不愿意,因为东海仲裁虽然有利中国,但是中国另一处经常引起纠纷的南海领域,几乎所有大陆架都不在中国一方,如果仲裁的话,划界肯定对中方不利,不符合中国的整体利益。

社论说,国际仲裁虽然对中国有利,但是中国也不愿意,因为东海仲裁虽然有利中国,但是中国另一处经常引起纠纷的南海领域,几乎所有大陆架都不在中国一方,如果仲裁的话,划界肯定对中方不利,不符合中国的整体利益。

社论认为,中国海洋石油总公司在八月份公布公司业绩时,宣布东海油气田将在九月中旬投入生产,春晓油气田十月将会输送天然气到宁波,日本的资源能源厅和外务省曾经对此表示抗议,并要求中方作出金钱赔偿。

普通检索

- [(word=对|word=对于)&tag=P]
- [word=给&(tag=P|tag=VV)]

普通检索

高级检索

[(word=对|word=对于)&tag=P]

搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

1

以后的这几年,我们回家上坟,都要去那棵树下烧烧纸,对于那棵树,对于这个世界,我们心存感激。

1

以后的这几年,我们回家上坟,都要去那棵树下烧烧纸,对于那棵树,对于这个世界,我们心存感激。

1

农业部有关负责人表示,随着农村改革的推进和农民分工分业深化,农民对合作内容、合作领域、合作方式的需求呈现多元化,合作社的创新形式日益多样。

1

虽然反对派代表团29日并没有到达日内瓦,但叙利亚反对派高级谈判委员会当天宣布参加和谈,还是让人们对本轮和谈增加了些信心。

1

法国总统奥朗德对叙利亚反对派高级谈判委员会决定赴日内瓦参加谈判表示欢迎,并称必须尽一切努力为叙利亚问题的政治解决创造机会。

<http://hunter.blcuicall.org>

普通检索

➤检索单个项

➤检索多个项

➤通过依存句法结构来检索

➤捕获

普通检索

- 多个项直接相连——用空格分隔多个项

普通检索

高级检索

吃 苹果

Q 搜索

报刊

文学

微博

口语

教材

论文

↓ 结果下载

得知坑道里的部队想**吃苹果**,军后勤部门马上从平壤买来苹果,准备向坑道里送,从上甘岭到后方医院沿途几十里路上,朝鲜群众自发摆起茶水站、苹果站。

不过,理察最近在英国BBC特别节目中表示,当初唱片公司把他们塑造成「只喝牛奶,只**吃苹果**,每天晚上都洗澡的乖宝宝」,完全和真实的他相反。

中国移动也爱“苹果”,因为“**吃苹果**”的大部分都是有消费能力的高端客户。

饭后一起围坐着,主席抽烟,我们**吃苹果**。

对于剧情,网友表示,看到黄磊边**吃苹果**边哭的模样,好想进入电视安慰他;许多网友对秀禾两难的处境也深表同情,但有网友却说「可怜之人必有可恶之处」,既然报恩和自己的幸福只能择一,那么就该慎选,选了就是自己的责任,怪不得别人。

而金载沅在拍戏闲暇时,体贴地问她要不要**吃苹果**,立刻现场秀了一段「徒手把苹果分一半」的小伎俩,让JOLIN惊讶不已。

普通检索

- 多个项之间有顺序和间隔的约束
 - 顺序：按照检索式中的顺序，在检索式中位于前边的处于例句的前面，在检索式中位于后面的项处于例句的后面。
 - 间隔：利用正则表达式*、+、?、{m,n}

普通检索

- []: 匹配任意词
- *: 匹配零个或多个

普通检索

高级检索

吃 []* 苹果

Q 搜索

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

结果下载

i 脱水水果干标榜天然食材脱水制成,苏湘雯指出,这类零食制作过程通常经过油炸加工,以香蕉片来说,每100公克就有528大卡,热量相当于9根香蕉**吃下肚**;蔬菜片每100公克有510大卡,等于20碟蔬菜;苹果片每100公克也有503大卡,等于是8颗**苹果**。

i 为了替团团、圆圆欢度来台后的第一个生日,园方特地准备特制生日蛋糕,2层冰块中冰着团团、圆圆爱吃的**胡萝卜、竹笋、苹果**,里面还用胡萝卜排成“5”的字样,象征熊猫已5岁。

i 广州军区“塔山英雄部队”战争年代路过群众苹果园**不吃一个苹果**的故事名扬天下。

i **吃甜食或苹果**后,一定要漱口或刷牙。

i 不过,理察最近在英国BBC特别节目中表示,当初唱片公司把他们塑造成「只喝牛奶,只**吃苹果**,每天晚上都洗澡的乖宝贝」,完全和真实的他相反。

普通检索

- +: 匹配一个或多个

普通检索

高级检索

只要 []+ 就

Q 搜索

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

↓ 结果下载

① 总之,她们都在以不同形式、不同姿态、不同表情、不同动作出现在照相机镜头前,只要摄影师一按快门,这张“全家福”就可拍成。

① 根据国中及高中毕业生升学及就业调查结果,国中毕业生总人数升读高中者,仅占百分之二十二点一三,而高中应届毕业的升学率则达百分之六十,加上重考一年考上者的升学率,则高达百分之七十九点一四,显示只要考上高中,就大部分都可以考上大专。

① 阿沁不只在开唱前2周被严禁吃肉和淀粉类食物,到香港住宿时,经纪人还直接禁止饭店的客房服务为阿沁提供肉或淀粉的餐点;唱片公司还下令说,只要他被抓到偷吃,演唱会上就绝不提供技师为他调音,让向来不能没有技师的阿沁只好乖乖遵守。

① 军地携手“风雨同行”,所体现的不仅是气象预报的资源共享,同时也说明了这样一个道理:在共建共享和谐社会的伟大实践中,军地只要善于取长补短密切配合,就能互相促进、共同受益。

普通检索

- ?：匹配零个或一个

普通检索

高级检索

先进[]? 技术

Q 搜索

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

↓ 结果下载

i 主办单位有鉴于电脑动画、影像处理的技术发展迅速,需求量也逐年增加,特别加重电脑动画及多媒体制作的展出比例,交流最新、**最先进的技术**及相关资讯。

i 资策会表示,希望这些活动能促成在美华人与台湾软体业者建立沟通合作管道,为台湾延揽高级专业人才并引进**先进技术**,带动软体工业在品质方面的提升。

i 双方的合作可有效地集合更多制造经验与专业知识,可望加速**先进控制技术**的进步,将自动化晶圆制造带进新的纪元。

i 曾繁城认为,先进制程的应用量会越来越大,政府应取消目前**先进技术**赴中国设厂需经过重重审议相关设限,并建立某种制度,主动让先进技术能在中国生产,例如每2年自动解禁开放更进一步的技术,才能让市场更全面发展。

<http://hunter.blcuicall.org>

普通检索

- $\{m,n\}$: 匹配m个到n个

普通检索

高级检索

Q 搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

↓ 结果下载

i

广州军区“塔山英雄部队”战争年代路过群众苹果园**不吃一个苹果**的故事名扬天下。

i

吃甜食或苹果后,一定要漱口或刷牙。

i

在1960至1985年间,荷兰曾对哮喘等慢性病与饮食的关系进行调查,结论是,**吃一定量的苹果**和梨可减少呼吸道疾病及肺病的发生。

i

故事叙述蚂蚁昆虫**吃了彩虹苹果**,每天都在狂欢,蚂蚁主角「杰克」找到红苹果,解救其他昆虫朋友。

i

当局认为,当时这只黑熊距她只有23公尺,正**吃著树上的苹果**。

i

动物园提供的信息显示,从2日上午9点到下午3点45分,“团团”“圆圆”共进食竹子达33公斤,此外还**吃了窝窝头、苹果、胡萝卜**等。

普通检索

- 某些项的重复

- [tag=JJ]*
- [word=的]?
- [tag=CD] [tag=M] [tag=JJ]* [word=的]? [tag=/N.*]/

普通检索

高级检索

[tag=CD] [tag=M] [tag=JJ]* [word=的]? [tag=/N.*]/

搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

❶

鸟虫篆印的基调接近清丽静逸,而又别致得像花腔女高音,正是印人想换换“口味”而出现的一道**新鲜佳肴**。

❷

《中国书法大字典》中,一个**单字**能表现为几十种相貌乃至结构大异的书写形体,正是历来诸多书家自设自创而逐渐增添的体魄。

❸

《中国书法大字典》中,一个单字能表现为**几十种相貌**乃至结构大异的书写形体,正是历来诸多书家自设自创而逐渐增添的体魄。

❹

其三,以优雅且形变的鸟虫**一类繁饰**篆字时需知,繁饰并非挤兑、替换原字的形体。

❺

据了解,这是中方落实第三届连云港论坛成果的一**项重要举措**。

普通检索

- 限制词汇难度等级

- [word=只有] [level=三] [word=才]

普通检索

高级检索

[word=只有] [HSK=三] [word=才]

Q 搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

1

多少年来,有无数共产党员抛家离子,投身党和人民的事业,他们并非铁石心肠,心中也都秉持着一份真爱,这其中的隐忍痛苦也**只有自己才能**体会,而其因奉献于伟大事业、服务于人民利益所绽放出的光辉,使他们的人生更有价值,使他们的爱情分外浪漫。

1

只有周末才能抽出时间和他视频谈心,鼓励他认真学习。”

1

陈冬现在的家就在单位旁边,但是为了把更多的精力放在训练上,他选择平常住宿舍,只有周末才回到家里。

1

这种交流的唯一障碍是西方教条地坚持**只有自己才**明白民主。

1

尽管,那个声音**只有自己才能**听到,但彼此的眼神,大家都懂!

1

中国医师协会输血科医师分会副会长、中国人民解放军总医院输血科主任汪德清认为:“当前临床用血存在安全、血源、疗效三大问题,直接危及病患的生命健康,只有**解决才能**惠及输血各方,更好地保障用血患者的安全。”

1

“既然是背影,只有别人才能看得到。

普通检索

➤检索单个项

➤检索多个项

➤通过依存句法结构来检索

➤捕获

普通检索

- 依存标签采用 SD的依存句法标注规范。汉语中常见的依存关系见下表：

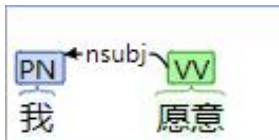
依存标签	依存关系	依存标签	依存关系	依存标签	依存关系
acl	名词修饰	compound	复合词	nsubj	名词性主语
advmod	副词修饰	conj	并列关系	nsubjpass	名词性被动主语
amod	形容词修饰	cop	系动词	nummod	数量词修饰
appos	同位语	det	限定词	parataxis	阐述关系
aux: ba	“把/将”字句	dobj	宾语	pcomp	介词补充
auxpass	助动词被动	neg	否定修饰符	pobj	介词宾语
auz	助动词	nmod	复合名词修饰	tmod	时间词修饰
case	case 标记	nn	复合名词修饰	root	根节点
ccomp	语义补充	auspass	“被”修饰		

普通检索

通过描写依存弧的方向以及依存标签，来检索满足条件的句子。

- 检索单条弧

- “[word=我] <nsbj [word=愿意]” 也可以写成 “[word=愿意] >nsbj [word=我]”



普通检索 高级检索

搜索

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

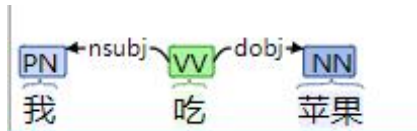
结果下载

- 他还坚定地说:“如果有一天我的生命之钟停摆了,我愿意把它定格在自己的岗位上。”
- 波斯纳是负责民主、人权和劳工议题的美国助理国务卿,他说:「我不愿意细谈每个案子,但是我们真的很诚心地花时间细谈具体案子和具体关切,说明我们最反对的意见。」
- 蔡勋雄警觉、笑笑回应表示「这个问题我不愿意讲」。
- 王志刚说,我愿意在亚太经济合作会议(APEC)架构下,建立区域金融体系监督系统、发展金融合作机制措施,与亚太社区各成员共同渡过金融危机。
- 如今在最幸福的时刻,殷剑又想到妈妈:“每当比赛不顺利或者遇到挫折的时候,我总是愿意和妈妈沟通,虽然她没有文化,也不懂帆船,但是她总能带给我力量,母亲的话陪伴着我成长。”
- 戴龙是美国云霄飞车爱好者俱乐部成员,曾经搭过全美一百四十六座云霄飞车,他说:「如果有必要,我愿意付出二千美元。」

普通检索

- 检索同一个节点上的两条弧

- [word=我] <nsubj [word=吃] >dobj [word=苹果]



普通检索 高级检索

搜索

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文 结果下载

- 「平等并不等于你吃一个苹果,我就必须吃一个苹果,平等是指你我都有吃一个苹果的权利」。
- 金森表示自己是“身”不由己,因为骨瘦如柴的身形是她的生财工具,“每次我决定增磅时,总是见到跟我一般瘦的模特儿.....为了保持瘦骨架,我每天只吃半个苹果”。
- 当时我还迷惑地问:难道我吃了你三个苹果,可以说一个也没吃吗?
- 每天吃东西,都想着宝宝是不是需要这样的营养,宝宝需要牛奶了,我赶紧喝杯牛奶,宝宝需要苹果了,我赶紧吃个苹果。

普通检索

- 检索不同节点上的弧

- [word=我] <nsubj [word=愿意] >ccomp [word=提供] >doobj[word=帮助]

普通检索高级检索

[word=我] <nsubj [word=愿意] >ccomp [word=提供] >doobj [word=帮助]

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

结果下载

PN

我

nsubj

VV

愿意

ccomp

VV

提供

doobj

NN

帮助

1

在这期间,如有需要三党领导人可以去北京见我,我愿意继续提供帮助,促使新政府尽快成立。”

1

萨马兰奇说:“如果中国提出申请,我愿意提供帮助。”

1

卡特强调,“我访问中东并不代表任何人,我只是要了解分歧各方的意图,如果美国总统布什、国务卿赖斯以及以色列政府有所需要,我愿意提供帮助。”

1

雷杰皮说:“尽管战争已经结束,我作为科索沃政府总理,但同时也是一名医生,如果需要,我愿意向科索沃各族公民提供帮助,也愿意同包括在媒体上反对过我的所有记者合作。”

1

福格茨说:“我愿意提供帮助。”

1

请与我联系,我愿意为你的孩子提供帮助。”

<http://hunter.blcuicall.org>

普通检索

- 弧的“或”结构

- [word=袭击] >/nsubj|nsubjpass/ []

普通检索高级检索

[word=袭击] >/nsubj|nsubjpass/ []

● 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

1

2014年12月16日,一伙装扮成军人的武装**分子袭击**了白沙瓦一所军人子弟学校,造成150人死亡,遇难者中绝大多数是学生。

1

未来,巴基斯坦安全形势仍不乐观,学校等人群密集地区要进一步加强安保措施,提高防范武装**分子袭击**的能力。

1

署理旺角警区指挥官丘绍萁表示,凌晨2时,因现场**警员**被严重**袭击**头部、生命安全遭严重威胁,警方为保护同事生命安全,在亚皆老街及上海街交界处开两枪示警。

1

据新华社香港2月9日电香港部分激进分子自8日晚开始与香港警方爆发严重冲突,有人主动**袭击**警员进而演变为暴乱。

1

1941年11月28日,日伪军3000余人**袭击**溧阳塘马村,为了掩护正在塘马村开会的苏南抗日根据地党、政、军干部安全转移,罗忠毅与第16旅政治委员廖海涛率部与敌人展开殊死搏斗,壮烈牺牲,时年34岁。

1

27日以来,强降雨**袭击**湖北多地,造成城区渍涝,道路交通受阻。

nsubjpass 被动主语

nsubj 主动主语

<http://hunter.bluicall.org>

普通检索

- 弧的通配符(“>>” 、 “<<”)
 - [tag=VV&word=表示]>> []

普通检索 高级检索

[tag=VV&word=表示] >> []

● 报刊 ○ 文学 ○ 微博 ○ 口语 ○ 教材 ○ 论文

成龙 在 领奖 时 激动 地 表示 , “ 作为 一 名 中国人 , 我 感 到 非 常 自 豪 ! ”

↓ 结果下载

① 成龙在领奖时激动地表示，“作为一名中国人，我感到非常自豪！”

■ 成龙在领奖时激动地表示,“作为一名中国人,我感到非常自豪!”

① 成龙在领奖时激动地表示,“作为一名中国人,我感到非常自豪!”

i 成龙在领奖时激动地表示,“作为一名中国人,我感到非常自豪!”

■ 成龙在领奖时激动地表示,“作为一名中国人,我感到非常自豪!”

i 成龙在领奖时激动地表示,“作为一名中国人,我感到非常自豪!”

<http://hunter.blcuicall.org>

成龙在领奖时激动地表示,“作为一名中国人,我感到非常自豪!”

普通检索

- 多层弧

- 表示 $\rightarrow \{1,2\}$ []

普通检索

高级检索

表示 $\rightarrow \{1,2\}$ []

搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

国际游联官员表示,他们周一之前不便发表任何评论,国际奥委会也同样会在周一工作后再作决定。

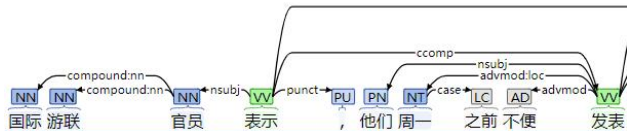
国际游联官员表示,他们周一之前不便发表任何评论,国际奥委会也同样会在周一工作后再作决定。

国际游联官员表示,他们周一之前不便发表任何评论,国际奥委会也同样会在周一工作后再作决定。

国际游联官员表示,他们周一之前不便发表任何评论,国际奥委会也同样会在周一工作后再作决定。

国际游联官员表示,他们周一之前不便发表任何评论,国际奥委会也同样会在周一工作后再作决定。

国际游联官员表示,他们周一之前不便发表任何评论,国际奥委会也同样会在周一工作后再作决定。



普通检索

- 词汇等级的限制

- [word=及时&tag=/VA|JJ/] <amod [level=三]

普通检索

高级检索

[word=及时&tag=/VA|JJ/] <amod [HSK=三]

Q 搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

❶

据山西医科大学第一医院医务处处长张春明介绍,安检测温门集成了人体安检、热成像测温、医闹人员识别报警等功能,“医院门诊和住院部将逐步安装,并与公安机关平台联网对接,以**及时发现**和防控安全隐患。”

❷

向社会公布水环境24小时社会监督电话,专人接听市民反映的各项水环境问题,督促各级责任人和管理部门对问题**及时**处置和**解决**。

❸

而中国共产党则在内部形成党内民主与党内集体领导制度,使得中国政治制度具有强大的政策动员能力,能够实现政策的**及时变化**。

❹

如果这一问题不能得到**及时的解决**,在全球化的背景之下,还会导致人才、资本“用脚投票”,大规模外流到具有较高保护水准的其他国家和地区。

❺

在党的十九大前**及时发现**并彻底查处孙政才严重违纪问题,充分表明了以习近平同志为核心的党中央坚持党要管党、全面从严治党的坚强意志、鲜明态度,充分体现了我们党将反腐败斗争和党风廉政建设进行到底的坚定决心、无畏气概。

<http://hunter.blcuicall.org>

普通检索

- 对单独字词的检索
 - 中文 [word=中文]
 - 学习 中文 [word=学习][word=中文]
- 词性/命名实体约束
 - [word=只&tag=/M|AD/]
 - [entity=ORGANIZATION]
- 通过依存关系约束
 - [word=刻画&tag=VV]>dobj[] ◀
 - [word=袭击] >/nsubj|nsubjpass/ []
- 通过词汇等级约束
 - [word=只有] [level=三] [word=才]
 - [word=及时&tag=/VA|JJ/] <amod [level=三]

普通检索-检索结果



DCC 2.0 动态流通语料库

普通检索

模式检索

[word=刻画&tag=VV]>dobj[]

Q 搜索

☒ 报刊 ☐ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

↓ 结果下载

- ❶ 客栈外的开阔地上,有一座1米高的石礅,上面竖立着一尊堂吉珂德手执矛盾的铁塑像,传神地刻画出了出征前踌躇满志的**神态**。
- ❷ 随着大众传媒的迅速发展,女性以瘦为美的概念和图像被迅速传播,逐渐在世界范围被接受,拥有纤瘦身材的女性被刻画为“富有吸引力、自信、快乐和成功”的**形象**。
- ❸ 人民日报任仲平文章《标注共产党人的精神坐标》,从涤荡“四风”到制度构建,从问题清扫到人心回归,于层层递进中标注出当代共产党人的精神坐标,以严、实、廉、清、敢、党6个字刻画了政治生态的**改变**。
- ❹ 我们专门的人物版和多个栏目,既生动描绘个体,也注重刻画新时代奋斗者**群像**。
- ❺ 其次,要坚持马克思主义哲学精神的指引,坚持刻画好党的形象的理想**性格**的“质的规定性、表现的丰富性、情致的始终如一性”的辩证统一。

普通检索

- 对单独字词的检索
 - 中文 [word=中文]
 - 学习 中文 [word=学习][word=中文]
- 词性/命名实体约束
 - [word=只&tag=/M|AD/]
 - [entity=ORGANIZATION]
- 通过依存关系约束
 - [word=刻画&tag=VV]>dobj[]
 - [word=袭击] >/nsubj|nsubjpass/ []
- 通过词汇等级约束
 - [word=只有] [level=三] [word=才]
 - [word=及时&tag=/VA|JJ/] <amod [level=三]

捕获

通过捕获可以直观地展示观察内容在语料中的频次。使用检索式(?<name> ...)来捕获和命名。

普通检索-捕获

- 对单独字词的检索
 - 中文 `(?<中文>[word=中文])`
 - 学习 中文 `[word=学习][word=中文]`
- 词性/命名实体约束
 - `(?<只> [word=只& tag=/M|AD/])`
 - `(?<机构名> [entity=ORGANIZATION])`
- 通过依存关系约束
 - `[word=刻画&tag=VV]>dobj(?<dobj>[])`
 - `[word=袭击] >/nsubj|nsubjpass/ (?<主语>[])`
- 通过词汇等级约束
 - `[word=只有] (?<三级词> [level=三]) [word=才]`
 - `[word=及时&tag=/VA|JJ/] <amod(?<三级词>[level=三])` ◀

捕获

通过捕获可以直观地展示观察内容在语料中的频次。使用检索式(`?<name> ...`)来捕获和命名。

普通检索-捕获结果

■ [word=及时&tag=/VA|JJ/] <amod(?<三级词>[level=三])



DCC 2.0 动态流通语料库

[关于我们](#) [帮助](#) [旧版](#)

普通检索

模式检索

搜索

报刊

教材

微博

口语

文学

论文

结果下载

1

司空见惯的新闻客户端的弹窗技术也是如此,只不过人们熟悉的是新闻事件的**及时传播**,甚至有时略有些信息过载,但当这种技术用到了刀刃上,就立刻发挥出传统模式所不具备的作用。

三级词

1

乌干达总理鲁贡达表示,乌干达政府感谢中国政府**及时**的人道主义**支持**,称这是非中友谊的象征。

三级词

1

乌干达总理鲁贡达表示,乌干达政府感谢中国政府**及时**的人道主义**支持**,称这是非中友谊的象征。

三级词

1

现如今,气象人则追求更准确、更精细,只为**及时预报**,贴心服务。

三级词

1

根据该方案确定的目标,到2022年,首都将健全完善方法科学、实施有效、**更新及时**的**标准**制定修订工作机制,基本建成体现北京特色、指标水平先进、系统构成完善的节能低碳和循环经济标准体系,努力打造全国节能低碳和循环经济标准创新中心、示范基地和辐射之源。

三级词

1

袁慧琴建议,加大戏曲优秀领军人物、各行当优秀演员的新剧目创作力度,实施“当代戏曲名家优秀代表作推广工程”,让优秀演员演出的传统戏、新创戏得到**及时记录**和广泛推广。

三级词

共有160条结果

捕获

元信息

三级词

--None--

三级词 (63 samples)

☐

(17) 处理

☐

(14) 支持

☐

(12) 支付

☐

(7) 情况

☐

(7) 解决

☐

(7) 调整

☐

(6) 指导

☐

(5) 应用

☐

(4) 保护

☐

(4) 推广

☐

(4) 能力

☐

(2) 人员



普通检索

- [word=/.*者/]

普通检索

高级检索

[word=/.“者”]

Q 搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

1

这是一支由76人组成的民间娱乐团,演奏者是来自陕西省西咸新区空港新城各行各业、各个年龄段的业余娱乐爱好者。

2

这是一支由76人组成的民间娱乐团,演奏者是来自陕西省西咸新区空港新城各行各业、各个年龄段的业余娱乐爱好者。

3

娱乐团艺术总监、今年75岁的退休音乐教师杨仲琪告诉记者,这两年乐团又加入了编钟、云锣、古筝等民族乐器,表现力更为丰富,器乐类一等奖。

4

尽管校方此前已在校园内进行了严格的安全部署,但袭击者仍然进入校园内。

5

我看见两名武装分子向他开火,我随即向里面逃跑,并跳出后墙逃生”,袭击事件亲历者艾哈迈德说。

6

印度总理莫迪当天也对此次袭击事件表示谴责,并向袭击受害者家属致以慰问。

7

2014年12月16日,一伙装扮成军人的武装分子袭击了白沙瓦一所军人子弟学校,造成150人死亡,遇难者中绝大多数是学生。

检索结果

元信息



普通检索

- 用来捕获前后缀
 - `(?<a>[word=/.*者/])`

普通检索

高级检索

(?<a>[word=/.*者/])

搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

1

这是一支由76人组成的民间坝乐团,演奏者是来自陕西省西咸新区空港新城各行各业、各个年龄段的业余坝乐爱好者。

a

1

这是一支由76人组成的民间坝乐团,演奏者是来自陕西省西咸新区空港新城各行各业、各个年龄段的业余坝乐爱好者。

a

1

坝乐团艺术总监、今年75岁的退休音乐教师杨仲琪告诉记者,这两年乐团又加入了编钟、云锣、古筝等民族乐器,表现力更为丰富。器乐类一等奖。

a

1

尽管校方此前已在校园内进行了严格的安全部署,但袭击者仍然进入校园内。

a

1

我看见两名武装分子向他开火,我随即向里面逃跑,并跳出后墙逃生”,袭击事件亲历者艾哈迈德说。

a

1

印度总理莫迪当天也对此次袭击事件表示谴责,并向袭击受害者家属致以慰问。

a

检索结果

捕获

元信息

a

--None--

a (125 samples)

☐ (418) 记者

☐ (85) 消费者

☐ (70) 或者

☐ (29) 作者

☐ (29) 学者

☐ (26) 投资者

☐ (24) 读者

☐ (23) 者

☐ (21) 工作者

☐ (20) 患者

☐ (17) 志愿者

普通检索

• 在检索结果中选择

普通检索

高级检索

(?<a>{word=/.,*者/})

Q 搜索

● 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

± 结果下载

1

马险峰:国际金融市场对绿色债券的标准涉及三个方面:第一,资金是否投向绿色项目,定义较为严格;第二,发行企业对资金投向有非常清晰的监督标准;第三,投资者对绿色债券的透明度要求较高,要求有定期的信息披露报告,并通过报告信息可以跟踪资金的具体使用情况。

a

1

一个全新的商业模式在何方头脑海里逐步成型:运用“检测技术+大数据+电商”模式为消费者提供生鲜产品,一来解决菜农卖菜难,二来解决食品安全监管难。

a

1

有媒体指出,现在人们都习惯上网寻找维修、运输、律师等方面的服务信息,如果从业人员能够让搜索引擎删除自己获得的客户差评,消费者权益就无法保证。

a

1

一方面,的确是一些投资者的惰性,他们想走特别保险的套路,花最少的钱,挣最多的钱,于是不断地重复自己;另一方面,也与创作者自己的懒惰有关。”

a

1

我们也在采取措施减少对于机构投资者长期投融资的机制性与监管障碍,推出新的金融工具,同时确保金融稳定。

a

检索结果

捕获

元信息

a

--None--

a (125 samples)

☐ (418) 记者

☒ (85) 消费者

☐ (70) 或者

☐ (29) 作者

☐ (29) 学者

☒ (26) 投资者

☐ (24) 读者

☐ (23) 者

☐ (21) 工作者

☐ (20) 患者

☐ (17) 志愿者

☐ (15) 创作者

普通检索

- 通过词性捕获
 - (?<普通名词> [tag=NN])

普通检索

高级检索

(?<普通名词> [tag=NN])

搜索

● 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

①

华尔街股市周五在部份上市公司强劲盈余报告掩护下,总算没被疲软的债券市场拖累,捡便宜的投资人到处充斥,也使买气显得旺盛,收市时股价止跌回升,道琼工业平均指数上升了十三点多。

普通名词

①

华尔街股市周五在部份上市公司强劲盈余报告掩护下,总算没被疲软的债券市场拖累,捡便宜的投资人到处充斥,也使买气显得旺盛,收市时股价止跌回升,道琼工业平均指数上升了十三点多。

普通名词

①

华尔街股市周五在部份上市公司强劲盈余报告掩护下,总算没被疲软的债券市场拖累,捡便宜的投资人到处充斥,也使买气显得旺盛,收市时股价止跌回升,道琼工业平均指数上升了十三点多。

普通名词

①

华尔街股市周五在部份上市公司强劲盈余报告掩护下,总算没被疲软的债券市场拖累,捡便宜的投资人到处充斥,也使买气显得旺盛,收市时股价止跌回升,道琼工业平均指数上升了十三点多。

普通名词

检索结果

捕获

元信息

普通名词

--None--

普通名词 (472 samples)

☐ (51) 处分

☐ (48) 情节

☐ (34) 级

☐ (18) 记大过

☐ (17) 警告

☐ (17) 记过

☐ (17) 降职

☐ (17) 降衔

☐ (13) 健保

☐ (12) 国际

普通检索

- 或 (|)
- ([?<a>](#)[word=/.*者/][word=/.*家/])

普通检索

高级检索

(?<a>[word=/.*者/][word=/.*家/])

搜索

● 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

① 这一机制的**创造者**、上海银监局局长廖岷说,“在坚守区域性、系统性风险底线的同时,监管有必要加快制度创新,不断建立健全更

a

① 很快,中能财务获得了梦寐以求的合同能源管理融资创新贷款资格,从中国清洁能源机制基金获得合同能源创新贷款6000万元,期

a

① 采访中**记者**了解到,合同能源创新贷款的资金来源于政策性基金,资金运用扩展到了集团外企业,均在形式上突破了《企业集团财

a

① 交通银行首席**经济学家**连平这样评价。

a

① 前一天下午,他在二十国集团工商峰会开幕式上的主旨演讲,以“后花园”与“百花园”,说明中国不是现行国际机制的**挑战者**,而是建

a

走向世界,更要影响世界。

检索结果

捕获

元信息

a

--None--

a (110 samples)

☐ (278) 国家

☐ (185) 记者

☐ (117) 家

☐ (49) 或者

☐ (36) 专家

☐ (33) 大家

☐ (19) 消费者

☐ (18) 投资者

☐ (14) 作曲家

普通检索

- 且 (&)
- (?<前缀>[word=/老.*/&tag=NN])

普通检索

高级检索

(?<前缀>[word=/老.*/&tag=NN])

Q 搜索

结果下载

● 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

林奕华并说,这家厂商竟然还获得法国的ISO-9002认证;其次,协调期间,厂商却直接跟举办婚宴的男女双方联络,送了六箱汽水给男方,女方则送一箱;面对吴老板的索赔,已投保商品标明责任险五千万的制造商,竟然也想以赔两倍数量货品回应他,并说,保险要求要实际饮用后身体发生不适才有理赔。

前缀

报导说,老鼠身上的试验显示,疫苗能让老鼠产生对抗禽流感所需的抗体。

前缀

报导说,老鼠身上的试验显示,疫苗能让老鼠产生对抗禽流感所需的抗体。

前缀

有些双职工家庭,晚上有事出去了,便将孩子送到有老人的邻居家,非常方便。

前缀

检索结果

捕获

元信息

前缀

--None--

前缀 (137 samples)

☐ (202) 老人

☐ (178) 老师

☐ (79) 老百姓

☐ (41) 老板

☐ (32) 老将

☐ (22) 老年人

☐ (20) 老婆

☐ (20) 老鼠

普通检索

- 捕获多个项中任一项的内容

- 拿 (?<aa> [tag=/N.*/]) 来说

普通检索

高级检索

拿 (?<aa> [tag=/N.*/]) 来说

搜索

☒ 报刊

☐ 文学

☐ 微博

☐ 口语

☐ 教材

☐ 论文

结果下载

① 就拿蚌埠来说,因为外出打工人数众多,我们建立了留守儿童学校,配备专人看护,让那些父母不在身边的孩子也能享受幸福的童年。”

aa

① 就拿土耳其来说,虽然政府希望给予美军过境的方便,但有关提案却遭到了议会的否决,迫使美国不得不改变军事部署。

aa

① 拿男团来说,我们仍然有四个主要对手,即瑞典、法国、德国和亚洲的韩国。

aa

① 拿夏季来说,各地进入夏季的时间不同,如济南一般为5月15日、成都为5月21日、北京为5月28日、西安为5月31日、天津为6月1日、长春为6月24日、乌鲁木齐为6月30日、西宁和昆明则无夏。

aa

① 拿劳动来说,它已经离学生越来越远了。

aa

检索结果

捕获

元信息

aa

--None--

aa (312 samples)

☐ (15) 北京

☐ (9) 美国

☐ (6) 去年

☐ (6) 粮食

☐ (5) 上海

☐ (5) 中国

☐ (5) 电影

☐ (4) 教育

☐ (3) 交通

☐ (3) 哲学

普通检索

- 捕获多个项中任几项的内容

- 又 (?<a> [tag=/V.*/]) 又 (? [tag=/V.*/])

普通检索

高级检索

又 (?<a> [tag=/V.*/]) 又 (? [tag=/V.*/])

搜索

● 报刊

○ 文学

○ 微博

○ 口语

○ 教材

○ 论文

✦ 结果下载

❶

分析说,陈致中夫妻态度如此从容,背后原因可能有三:其一是真的问心无愧,两人真的“又傻又天真”,被母亲拿来当人头却毫不知情;其二是强作镇定,也就是“走夜路吹口哨”,装腔作势给支持者和反对者看,更给检调司法机关看。

a

b

❷

但一般预料,导致此一地区极易燃烧成灾的又热又干的圣安娜风可能在今晚到周末间加强风势,而使人口密集地区危险性升高。

a

b

❸

由于雨太大且又冷又饿,几人手牵着手决定冒险从那里过,没想到刚刚走出几米,走在前面的两个同伴因脚下踩空掉进水里,湍急的河水将两人冲走了。

a

b

❹

住在嘉义县竹崎山区,现年六十三岁的苏联生,从小看他父亲用毛笔替亲友、邻居写信而对书法产生兴趣,因他父亲没有受过什么教育,但书写均用毛笔,而且是悬腕运笔,又稳又好,人家说完他已写好,令他崇拜不已。

a

b

❺

新华社华盛顿2月8日电在童话故事中,那些又大又坏的狼往往都被描绘成有着黑色或者深色的皮毛。

a

b

❻

等迪安醒来时,山谷空无一人,又冷又饿的他只得找些老鼠、蛇和野果充饥。

a

b

共有1000条结果

检索结果

捕获

元信息

a

b

a (221 samples)

☐ (132) 好

☐ (77) 快

☐ (51) 大

☐ (37) 唱

☐ (33) 惊

...

b (274 samples)

☐ (142) 快

☐ (63) 跳

☐ (49) 好

☐ (33) 喜

☐ (24) 恨

...

普通检索

- 捕获多个项中任几项的内容
 - 又 (?<a> [tag=/V.*/]) 又 (? [tag=/V.*/])
 - 在多个捕获中选择

普通检索

高级检索

又 (?<a> [tag=/V.*/]) 又 (? [tag=/V.*/])

搜索

报刊

文学

微博

口语

教材

论文

结果下载

新华社北京8月14日电:又^a好^b又快地发展生猪生产新华社记者董峻农业部、商务部、银监会和保监会有关负责人14日下午就当前生猪生产形势接受中国政府网专访,网民提问活跃,显示当前社会各界对生猪生产的高度关注。

军队建设向信息化聚焦,实现战略性转变,各项投入不断加大,推动部队建设进入又^a好^b又快的发展轨道,全军广大指战员以崭新的风貌展示在世人面前。

面对激烈的竞争,如何真正把医院做大做强,实现又^a好^b又快的科学发展,犹如一道难解的方程式,横亘在费丹兵的面前。

新世纪之初,人们更加关注,曾经引领了一个时期经济发展的“珠三角”,将如何保持又^a好^b又快的发展势头,续写又一个“春天的故事”?

检索结果

捕获

元信息

a

b

a (221 samples)

☒ (132) 好

☐ (77) 快

☐ (51) 大

☐ (37) 唱

☐ (33) 惊

b (274 samples)

☒ (142) 快

☐ (63) 跳

普通检索

• 通过依存结构捕获

- [word=刻画&tag=VV]>dobj []
- [word=刻画&tag=VV]>dobj (?<dobj> [])



文心语料库检索平台

帮助 反馈

快速检索

模式检索

[word=刻画&tag=VV]>dobj (?<dobj> [])

搜索

☒ 报刊

☐ 教材

☐ 微博

☐ 口语

☐ 文学

☐ 论文

± 结果下载

① 客栈外的开阔地上,有一座1米高的石墩,上面竖立着一尊壁古诗词手执矛盾的铁塑像,传神地刻画出了出征前踌躇满志的神态。

dobj

② 随着大众传媒的迅速发展,女性以瘦为美的概念和图像被迅速传播,逐渐在世界范围被接受,拥有纤瘦身材的女性被刻画为“富有吸引力、自信、快乐和成功”的形象。

dobj

③ 人民日报任仲平文章《标主共产党人的精神坐标》,从涤荡“四风”到制度构建,从问题清扫到人心回归,于层层递进中标注出当代共产党人的精神坐标,以严、实、廉、清、敢、宽6个字刻画了政治生态的改变。

dobj

④ 我们有人物版和多个栏目,既生动描绘个体,也注重刻画新时代奋斗者群像。

dobj

⑤ 其次,要坚持马克思主义哲学精神的指引,坚持刻画好党的形象的理想性格的“质的规定性、表现的丰富性、情致的始终如一性”的辩证统一。

dobj

⑥ 电影《生死抉择》、豫剧现代戏《全家福》等作品,就成功刻画了党内个别腐败分子的艺术形象,辩证法的度把握得很好,电视剧《海

dobj

共有200条结果

检索结果

摘要

元信息

dobj

--None--

dobj (117 samples)

☐ (27) 形象

☐ (26) 人物

☐ (8) 性格

☐ (4) 年轮

☐ (3) 群像

☐ (3) 角色

☐ (2) 丽质

☐ (2) 关系

☐ (2) 历程

☐ (2) 场景

☐ (2) 复杂

1/1 4-80

Copyright © BLCU-ICALL 2021

普通检索

• 通过依存结构捕获

- [word=连] <advmod [tag=/N.*/] <nsubj [tag=/V.*/] >advmod [word=也]
- [word=连] <advmod (?<a> [tag=/N.*/]) <nsubj (? [tag=/V.*/]) >advmod [word=也]

普通检索

高级检索

[word=连] <advmod (?<a> [tag=/N.*/]) <nsubj (? [tag=/V.*/]) >advmod [word=也]

☒ 报刊 ☐ 文学 ☐ 微博 ☐ 口语 ☐ 教材 ☐ 论文

AD 连

advmod

NN 学者

也

AD

advmod

VV 承认

↓ 结果下载

① 上述一系列显著成就,改变了“敦煌在中国,敦煌学在国外”的被动局面,现在连外国学者也公开承认中国是国际敦煌学的中心。

a b

② 他拉着我的手,是很有力量的那种,说话的气势也很是磅礴:“人家孔乙己说了,窃书不能算偷书,你连鲁迅也不信吗?”

a b

③ 奇怪的是顾客并没有因为涨价而减少,倒是不断增加,连外小区的**人**也来买。

a b

④ 起先,艾柯起了《修道院罪行》这样一个直白的侦探小说题目,中途又想以叙述者阿德索·德·梅尔克的名字作为题目,结果连出版社也不干了,认为这样一个题目会让作品毫无销路。

a b

检索结果

捕获

元信息

a

↓

a (485 samples)

☐ (7) 人

☐ (4) 儿女

☐ (4) 午饭

☐ (4) 口水

☐ (4) 学者

☐ (4) 水

☐ (4) 电商

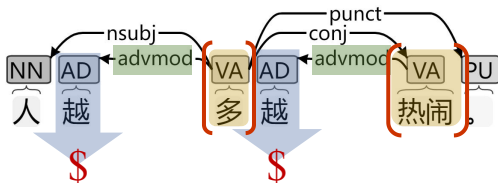
☐ (4) 苹果

检索模式

- 普通检索：用于检索单个项以及有顺序和间隔规定的多个项。
- 模式检索：用于检索与给定例句结构相同的句子。

模式检索

- 锚点：用于精准匹配**固定的成分**，用“\$”表示
- 捕获：**待观察的成分**捕获为变量，用“()”将内容括起来
- 未被标记为捕获或锚点的词，则被认为是帮助句子符合语法要求，使句子完整的成分。
- 在输入一个带标记符号的句子后，系统会返回例句的**依存句法图**呈现给用户，并依据该依存句法图匹配结果。



人\$越(多)\$越(热闹)

- ① 邻居们越听越糊涂。
多 热闹
- ① 我问问题,你回答,回答得越快越好。
多 热闹
- ① 爷爷一口气背到了底,我静静地听,越听越佩服,不禁喊道:“爷爷真棒!”
多 热闹
- ① 据说是越叫得随意,越不易丢失。
多 热闹
- ① 元宵是一点一点滚出来的,越滚越大,里面是馅,外面是江米粉。
多 热闹

模式检索

- 越.....越.....
 - 人\$越(多)\$越(热闹)



文心语料库检索平台

[帮助](#) [反馈](#)

普通检索

模式检索

人\$越(多)\$越(热闹)

检索

☐ 报刊

☒ 教材

☐ 微博

☐ 口语

☐ 文学

☐ 论文

± 结果下载

1

NN

人

2

AD

越

3

AD

多

4

VA

越

5

AD

越

6

VA

热闹

检索式

(?<越>[word=越]) <advmod (?<多>[?]) >conj (?<热闹>[?]) >advmod (?<越>[word=越])

1

邻居们越听越糊涂。

多

热闹

2

我问问题,你回答,回答得越快越好。

多

热闹

3

爷爷一口气背到了底,我静静地听,越听越佩服,不禁喊道“爷爷真棒!”

多

热闹

4

据说是越叫得随意,越不易丢失。

多

热闹

5

元宵是一点一点滚出来的,越滚越大,里面是馅,外面是江米粉。

多

热闹

6

我越学越觉得自己能够掌握这些知识,越学越觉得学习是非常有意思的事情。”

多

热闹

检索结果

捕获

元信息

多

--None--

多 (83 samples)

☐ (15) 多

☐ (14) 学

☐ (11) 大

☐ (10) 做

☐ (9) 长

☐ (7) 看

☐ (6) 吃

☐ (6) 快

☐ (5) 办

☐ (5) 练

☐ (5) 走

☐ 1/11 11

Copyright © BLCU-ICALL 2021

模式检索

- 非.....不可
 - 再说的话，我\$非(打)你\$不\$可。



文心语料库检索平台

帮助 反馈

普通检索

模式检索

再说的话，我\$非(打)你\$不\$可。

搜索

☒ 报刊

☐ 教材

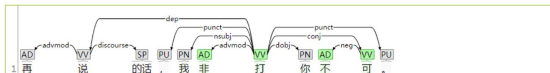
☐ 微博

☐ 口语

☐ 文字

☐ 论文

结果下载



检索式: (?<非>[word=非]) <advmod (?<打>[I]) >conj (?<可>[word=可]) >neg (?<不>[word=不])

① 即便有些证明非开不可,也要让数据多跑路、民众少跑腿。

打

① 但企业要做大做强,这条路非走不可,不进则退。*

打

① 在国外打中国品牌,资金、营销、人才面临重重挑战,有可能几百万资金投下去都“打了水漂”,但这是一道非跨越不可的坎。

打

① 作为小镇镇长,何永智有信心将小镇打造成贵阳“非去不可的目的地”。

打

① 今日之中国,早已不是那个任人欺凌的旧中国,但要更上层楼,非得脚踏实地、一步一个脚印不可。

打

① 周围的人都说好,大家都在谈论,大有“非读不可”之势。

检索结果

捕获

元信息

打

--None--

打 (40 samples)

☐ (7) 走

☐ (4) 去

☐ (3) 解决

☐ (2) 做好

☐ (2) 开

☐ (2) 烧

☐ (2) 管

☐ (2) 行

☐ (2) 触动

☐ (1) 买

☐ (1) 保护

☐ (1) 非

模式检索

■ 介词“对”后可以跟的谓词性成分

- 这\$对我(有利)



DCC 2.0 动态流通语料库

[关于我们](#) [帮助](#) [旧版](#)

普通检索

模式检索

这\$对我(有利)

搜索

☐ 报刊 ☒ 教材 ☐ 微博 ☐ 口语 ☐ 文学 ☐ 论文

结果下载

graph TD; nsubj --- case; case --- nmod; nmod --- prep; prep --- WA; WA --- 有利; 这 --- PN; 对 --- P; 我 --- PN

检索式: (?<有利>[I])>nmod prep (?<我>[I])>case (?<对>[word=对])

① #一件好事星期天早上,姐姐对我 说:“今天我们来做一些有意义的事情吧。”
有利

② 另一个人对他 说:“你把运动鞋穿上有什么用?”
有利

③ 对羚羊来说,如果它跑不过最快的狮子的话,它就会被吃掉。
有利

④ 你对历史也 有 兴趣?
有利

⑤ 您对中国历史一定很 有 研究嘛。
有利

⑥ 今天参观这个乾陵博物馆对您 研究唐代历史一定会有很大帮助。

捕获

元信息

有利

--None--

有利 (118 samples)

☐ (38) 说

☐ (15) 有

☐ (12) 进行

☐ (6) 充满

☐ (3) 没有

☐ (3) 重要

☐ (2) 习惯

☐ (2) 关心

☐ (2) 加以

☐ (2) 好

☐ (2) 干涉

☐ (2) 吃

© BLCU-ICALL 2022 版权所有

60

检索结果的过滤与下载

- 点击语料前的  标志，可以查看语料详情，包括语料id、文章类型、题目、日期、来源、文章段落等。
- 在界面右侧的“元信息”下，可对检索结果的元信息进行选择。
- 检索结果页面右上角位置有“结果下载”按钮，用户可指定下载的检索结果条数（默认为 50 条）与文件名，点击“结果下载”按钮，可将查询结果以本文文件 (*.txt) 格式保存至本地电脑。

语料详情

所在段落

西安是很值得参观的。你对历史也有兴趣？我也是搞历史的。

语料元信息

类型	
doc_type	textbook
textbook	今日汉语(一)(第二版)
title	31.我们是同行

捕获

元信息

date

date (825 samples)

- ☐ (4) 20091021
- ☐ (4) 20101220
- ☐ (3) 19920720
- ☐ (3) 19970416
- ☐ (3) 19970531
- ☐ (3) 19980213
- ☐ (3) 20001022



DCC 2.0 动态流通语料库的应用

■ 国际中文教育语法点自动抽取^[2]

- 对象：《国际中文教育中文水平等级标准》(GF0025 - 2021) 中《附录（规范性）语法等级大纲》中的语法点。

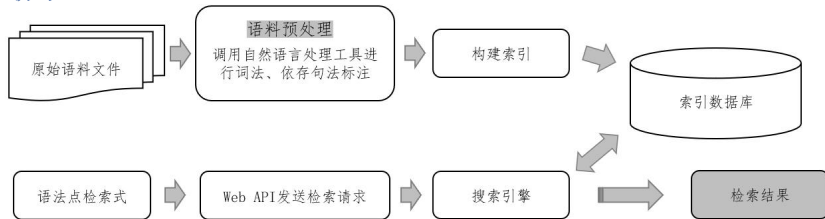
■ 语法点检索式实例

id	级别	语法项目	类别	细目	语法内容	
93	二级	固定格式			还是……吧	[word=还是&tag=AD][]*[word=吧&tag=SP]
94	二级	固定格式			又……又……	[word=又&tag=AD]<advmod[tag=/V.*/][word=又&tag=AD]<advmod[tag=/V.*/]
95	二级	固定格式			(在)……以前/以后/前/后	[word=在]?[word=/以前 以后/&tag=NT] <nmod.tmod [tag=/V.*/]
168	三级	固定格式			除了……(以外), ……还/也/都……	[word=除了 &tag=P]<case[]>case[word=外 word=以外&tag=LC][word=','?][word!=','?]*[word=/也 还 都/&tag=AD]
169	三级	固定格式			从……起	[word=从&tag=P]<case[] </nmod.*/?[]? >ccomp[]? >/advmod.*/?[]? >case[word=起&tag=LC]
170	三级	固定格式			对……来说	[word=对&tag=P]<case[]?</nmod.*/?[]?>/advmod.*/?[]?>case[word=来说&tag=LC]
171	三级	固定格式			一……也/都+不/没……	[word=/一 一点 一点儿/<nummod[]<nsubj[][]>advmod[word=也 word=都]
172	三级	固定格式			越……越……	[word=越&tag=AD]<advmod[]>>[1,4][]>advmod[word=越&tag=AD]
240	四级	固定格式			一+量词+比+一+量词	
241	四级	固定格式			(自)……以来	[word=自&tag=P]? <case [] >case [word=以来&tag=LC]
242	四级	固定格式			由……组成	[word=由&tag=P]<case[]<nmod.prep[word=组成&tag=VV]
243	四级	固定格式			在……方面	[word=在&tag=P]<case[word=方面&tag=NN]
244	四级	固定格式			在……上/下/中	[word=在&tag=P]<case[]*>case[word=/上 中 下/&tag=LC]
313	五级	固定格式			从……来看	[word=从&tag=P]<case[]>case[word=来看&tag=LC]
314	五级	固定格式			到……为止	[word=到&tag=P]<case[]>case[word=为止&tag=LC] [word=到此为止]
315	五级	固定格式			够……的	[word!=不][word=够&tag=VV][tag!=PU]*[word=的][tag=PU]
316	五级	固定格式			拿……来说	[word=拿&tag=P]<case[]<case[word=来&tag=MSP]<aux.prtmod[word=说&tag=VV]

[2]朱君辉,刘鑫,杨麟儿,王鸿滨,杨尔弘,汉语语法点特征及其在二语文本难度自动分级研究中的应用[J],语言文字应用,2022(03):87-99.

检索式的应用

■ 语法点抽取流程



■ 语法点在文本中的抽取结果实例

文本

北京有座美丽的中山公园，公园里有个用五色土砌成的社稷坛。

(《成功之路·成功篇》第2册《古迹今日》首句)

语法点

一级语法点及频数：“结构助词”的(2次)，方位名词“里”(1次)，量词“个”(2次)，
“有字句”(2次)，陈述句(1次)；

二级语法点及频数：偏正短语(3次)，名词性短语(3次)，方位短语(1次)。

参考文献

- [1] 郭慧志, 王强军, 刘华, 张普. 大规模动态流通语料库的构建[J]. 第二届全国学生计算语言学研讨会论文集. 2004
- [2] 靳光瑾, 肖航, 富丽, 章云帆. 现代汉语语料库建设及深加工[J]. 语言文字应用, 2005,(2).
- [3] 李正华. 汉语依存句法分析关键技术研究[D]. 哈尔滨工业大学, 2013.
- [4] 谭晓平. 现代汉语文本语料库建设及应用现状研究[J]. 对外汉语研究, 2018,(2).
- [5] 荀恩东, 饶高琦, 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,(1).
- [6] 詹卫东. 近30年来中文语言知识资源发展及应用[J]. 语言战略研究, 2018,(4).
- [7] 詹卫东. 北京大学CCL语料库的研制[J]. 语料库语言学, 2019.
- [8] 张永伟, 刘婷, 刘畅, 吴冰欣, 俞敬松. 融合句法信息的文本语料库检索方法研究[J]. 数据分析与知识发现, 2022
- [9] De Marnee M C, Manning C D. Stanford typed dependencies manual[C], 2010.
- [10] Kilgarrieff A, Baisa V, Bušta J. The Sketch Engine: ten years on[J], 2014,(1).
- [11] Luotolahti J, Kanerva J, Ginter F. Dep_search: Efficient Search Tool for Large Dependency Parsebanks[C]//Proceedings of the 21st Nordic Conference on Computational Linguistics. Gothenburg, Sweden: Association for Computational Linguistics, 2017.



谢谢大家！