# 语料库检索系统使用说明

##### 注意：在检索与后台录入文档时所有标点符号都应该使用全角输入。

**普通查询**

1.1 查询表达式简介

查询表达式由操作符、基本项、简单项、复杂项、过滤项等构成。下面依 次介绍这些单元。

1.1.1 操作符

查询表达式中可以使用的特殊符号包括9个：SPACE     |     $       #      +     -     ~     !     :

这些符号分为四组：

Operator1:  SPACE    |

Operator2:  $    #    +    -    ~    
  
Operaotr3:  !

Operator4:  :

符号的含义如下：

（一） Operator1:  Operator1是二元操作符，它的两边可以出现“基本项”（关于“基本项”的定义见1.1.2）

（1） SPACE（空格）相当于逻辑中的“并”关系。

（2）  |   相当于逻辑中的“或” 关系。

在查询表达式中，Operator1可以连续多次使用，即 A Operator1 B Operator1 C Operator1 D ... 是合法的查询表达式（参见1.1.3）

（二） Operator2：Operator2是二元操作符，它的两边可以出现“简单项”（关于“简单项”的定义见1.1.3）

（3）  $   表示它两边的“简单项”按照左边 在前、右边在后的次序出现于同一句中。两个“简单项”之间相隔字数小 于或等于Number  
  
（4）  #   表示它两边的“简单项”出现于同一句中，不考虑前后次序。两个“简单项”之间相隔字数小于或等于Number  
  
（5）  +   表示它两边的“简单项”按 照左边在前、右边在后的次序出现于同一句中。两个“简单项”之间相隔字数刚 好等于Number

（6）  -   表示它左边的“简单项”出现于句子中，并且，在右边 相隔Number个字的范围内，-号右边的“简单项”不出现。

（7）  ~   表示它左边的“简单项”出现于句子中，并且，在左边 相隔Number个字的范围内，~号右边的“简单项”不出现。

除 $ 和 + 操作符外，Operator2不能连续多次使用，即只能用Operator2连接两项：A Operator2 B，形成查询表达式。

Operator2中的 $ 和 + 可以连续多次使用，且可以混合使用。见1.2小节的说明。

（三）Operator3：Operator3是一元操作符。

（8）  !   表示它后面的“简单项”是本次查询的主关键字符串，显 示查询结果时以该“简单项”作为中心来 进行定位。

（四）Operator4：西文冒号 : 是分隔符（delimiter）

（9） :  跟在 author，name，type，pattern 等关键字后面，用于分隔关键字和它们的取值。这样形成的查询式称之为“[过滤项](http://120.78.71.214/corpus/readme/" \l "2.5_%E8%BF%87%E6%BB%A4%E9%A1%B9_)”（见下面 1.1.5）

注意：上述操作符不能作为 基本项在语料库中进行检索。

1.1.2 基本项

指不 包含特殊符号和空格的连续字符串

1.1.3 简单项

简单项可以由以下三种形式的序列组成

（1）  基本项

（2）  基本项1 Operator1 基本项2 Operator1 ...

（3） (基本项1 Operator1 基本项2 Operator1 ...)

注意：在实际表达式中， Operator1 前后不能有空格

1.1.4 复杂项

复杂项可以由以下三种形式的序列组成

（1）  简单项

（2）  简单项1 Operator2 Number 简单项2

（3）  简单项1 Operator2 Number Operator3 简单项2

其中第二种形式，等价于   Operator3 简单项1 Operator2 Number 简单项2，换句话说，如果以第一个简单项作为查询结果的显示中心，！可以省略。

注意：Number为0和 正整数。Operator2，Operator3前后均不能有空格

Operator2后面的 Number是必须的，不能省略。Number=0表示相邻，Number=1表示间隔1个单位，其余依此类推。

1.1.5 过滤项

过滤项可以包含以下表达式：

（1）author:

（2）document:

（3）type:

（4）dynasty:

说明：

**-- “author:”的含义是指“author”关键字后面加上指定作者的名字 。**

**-- 通过指定过滤项中author（作者），document（篇名），type（文章类型），dynasty（朝代），用户可以缩小查询语料的范围。**

**--**

**例1： 想查询“老舍”的语料，在查询表达式中输入“author:老舍”即可；**

**1.2 　使用 $  + 操作符的查询表达式**

**$ 符号表示间隔小于等于，如“把$10给”表示返回“把”与“给”之间少于10个字符的句子。**

**查询表达式支持多个“$”连用，如查询“被$10把$3给$2了”，表示 “被、把、给、了”四个关键字在一个句子中共现，并且相互之间有间隔字符的要求，“被”在“把”前出现，二者之间间隔小于10个字符。**

**+ 符号表示间隔等于，如“把+10给”表示返回“把”与“给”之间等于10个字符的结果。**

**查询表达式支持多个“$”或“+”连用，如支持查询“我$10你$3他$2 了” “你+3他+2了”。**

**此外系统也支持“$”与“+”的组合搜索，如“我$10你+3他$2了”， 该查询表示返回“我” 和“你”间隔小于等于10，“你”和“他”间隔等于3，“他”和“了”间隔小于等于2。**

**模式查询**

2.1 模式查询查询表达式简介

模式查询由基本项、操作符、匹配项、复杂项和过滤项构成。下面依次介绍这些项目。

2.1.1 基本项

指不包含特殊符号和空格的连续字符串。

2.1.2 操作符

查询表达式中可以使用的特殊符号包括3个：

① <

② =

③ -

这些符号可分为两组：

Operator1:  < =

Operator2: -

符号的含义如下：

下列项目中”Number”表示一个数字。

（1） <Number 表示匹配小于Number个字的文本。

（2） =Number 表示匹配等于Number个字的文本。

（3） Number1-Number2 表示匹配长度在Number1至Number2之间的文本（长度包括Number1与Number2）。

2.1.3 匹配项

复杂项由以下3种形式的序列组成：

下列项目中”Letter”表示一个或多个英文字母，大写与小写皆可。

1. (Letter Operator1 Number)
2. (Letter Operator2 Number Operator2)
3. (Letter)

其中，“Letter”表示某一段文字的匹配，(1)与(2)完成对一个Letter的声明，该段文字的长度满足声明时的操作符的要求。在完成声明后，Letter在在“复杂项”（见4.3.2.3）中第二次以上出现时，应使用(3)的形式，表示该位置匹配匹配的文段与第一次匹配到的文段相同。若Letter声明时（第一次出现）使用(3)的形式，则与(Letter<10)等效。

注意：在实际表达式中，匹配项内不能有空格。

2.1.4 复杂项

复杂项由基本项与匹配项相互组合而成。其中多个匹配项可连续出现。

2.1.5 过滤项

过滤项与普通查询的过滤项（4.3.4.5）一致。

**查询举例**

**查询式例子 1:  
  
计算机硬件  
  
意思是: 查出所有包含“计算机硬件”的句子。**

**查询式例子 2:     
  
把  被  
  
意思是: 查出所有包含“把”，同时也包含“被”的句子，即两个关键字之间无次序限制，无距离限制，只需要在一句范围内。**

**查询式例子 3:     
  
把|被  
  
意思是: 查出含有“把”或“被”的句子，两个关键字只需有一个在句中出现，就作为查询结果输出。  
  
查询式例子 4:     
  
把-4不  
  
意思是: 查出含有“把”，但在“把”右边4个字范围内不含“不”的句子。注 意：- 号属于opertaor2，其后必须有数字，且不能有空格。**

**查 询式例子 5:     
  
给~4把  
  
意思是: 查出含有“给”，但在“给”左边4个字范围内不含“把”的句子。注 意：~ 号属于operator2，其后必须有数字，且不能有空格。  
  
查询式例子 6:     
  
与其$10不如  
  
意思是: 查出同时含有“与其”和“不如”的句子，并且“与其”在先，“不如”在后出现，间隔10字以内。  
  
查询式例子 7:     
  
能力#3大  
  
意思是: 查出同时含有“能力”和“大”的句子，且“能力”和“大”之间的间隔在3个字之内，二者的先后次序不受限制。  
  
查询式例子 8:     
  
吃+3亏  
  
意思是: 查出同时含有“吃”和“亏”的句子，并且“吃”在先，“亏”在后出现，二者之间刚好间隔3个字。  
  
查询式例子 9:     
  
被$10!给  
  
意思是: 查出同时含有“被”和“给”的句子，并且“被”在先，“给”在后出现，二者之间间隔10个字以内。显示查询结果时，以“给”为“中心词”，即“给”居中对 齐。  
  
查询式例子 10:     
  
(把|被)$10给  
  
意思是: 查出同时含有“把”和“给”的句子，并且“把”在先，“给”在后出现，二者之间间隔10个字以内。  
或者，查出同时含有“被”和“给”的句子，并且“被”在先，“给”在后出现，二者之间间隔10个字以内。**

**查询式例子 11:**

**(把 |被)$10!给**

**意思是: 查出同时含有“把”和“给”的句子，并且“把”在先，“给”在后出现，二者之间间隔10个字以内。  
或者，查出同时含有“被”和“给”的句子，并且“被”在先，“给”在后出现，二者之间间隔10个字以内。  
显示查询结果时，以“给”为“中心词”，即“给”居中对齐。**

**查询式例子 12：**

**了$0(。 |？|，|！)**

**意思是：查出“了”与标点符号“。？，！”等紧邻出现的句子。这实际上就部 分地达到了查询“句尾了”（汉语学界一般所说的“了2”） 的目的。**

**查询式例子 13：**

**所以 author:老舍**

**意思是：在现代汉语语料中查作家老舍的文章中“所以”的用例。**

**查询式例子14：**

**模式查询：有一种(X,=2)叫(Y,=2)**

**X和Y是两个不同的字符串，且长度都为两个汉字字符。（返回结果：有一种力 量叫感动；有一种放弃叫成全；……）**

**查询式例子15：**

**模式查询：爱(V,=1)不(V)**

**V字符串的长度为1个汉字字符； （返回结果：爱借不借；爱理不理；爱管不管……）**

**模式查询：爱(V,<5)不(V)**

**V字符串的长度不超过5个汉字字符；**

**模式查询：爱(V,1-5)不(V)**

**V字符串的长度介于1-5个汉字字符之间。**

**－－－　正文完　－－－**