

全文検索システム『ひまわり』/利用者マニュアル/1_6/4. 詳細な検索を行う

[Top](#) / [全文検索システム『ひまわり』](#) / [利用者マニュアル](#) / [1_6](#) / 4. 詳細な検索を行う

言語を選択 | ▼

[Prev](#)

[全文検索システム『ひまわり』/利用者マニュアル/1_6](#)

[Next](#)

4. 詳細な条件で検索する ±

ここでは、詳細な検索条件を指定するために、次の事柄について扱います。はじめて『ひまわり』をお使いになる方は、まず、「[3. 検索してみよう](#)」をご覧ください。

- [検索文字列の指定](#)
- [フィルタ（絞込み検索）](#)
- [コーパスの選択](#)
- [検索オプション](#)

4.1 検索文字列の指定 ±

検索条件として指定できるのは、次の三つの項目です。

- 検索対象
- 検索文字列
- 絞込み条件(フィルタ)

検索対象 ±

検索対象は、「検索文字列」欄の左のプルダウンメニューで指定します。

検索文字列	フィルタ	コーパス	検索オプション
本文		あの	
本文			で終る ▼
本文(正規表現)			で始まる ▼
ルビ(rt)完全一致			
ルビ(rt)部分一致			

上の例は、[『青空文庫』（サンプル）](#)の例ですが、次に示すように、資料のどの部分を検索対象とするかを設定することができます。

- 本文：「青空文庫」の本文を検索対象とします。
- 本文（正規表現）：「青空文庫」の本文を検索対象とします。正規表現で検索文字列を指定することができます。ただし、通常の「本文」よりも、検索速度は低速です。
- ルビ(rt)完全一致：「青空文庫」のルビを検索対象とします。検索文字列との照合は、「完全一致」で行われます。
- ルビ(rt)部分一致：「青空文庫」のルビを検索対象とします。検索文字列との照合は、「部分一致」で行われます。

検索文字列 ±

検索文字列は、三つの部分から構成されており、『ひまわり』では、次の三つの欄から入力します。この三つの欄は、検索結果の「前文脈」、「キー」、「後文脈」列に対応しており、三つの欄で指定した文字列と一致するものだ

けが、検索結果に含まれることになります。ただし、「前文脈」、「後文脈」欄は、資料や「検索対象」メニューの選択対象によって、インターフェイスが変わる可能性があります。詳しくは、この後述べます。

例えば、下の例では、「キー」列が「です」で、「後文脈」列が「。」で「始まる」ものだけが検索されていることがわかります。なお、「始まる」や「終わる」といった条件は、検索結果の「前文脈」「後文脈」それぞれのセル内での位置であることに注意して下さい。

...	前文脈	キー	後文脈	Path
1	る事ができるなら満足	です	。三「私が両	/aozora...
2	は私で弁解を始めたの	です	。三十一「そ	/aozora...
3	にはそれができないの	です	。三十三「十	/aozora...
4	と私は見込んでいたの	です	。三十五「こ	/aozora...
5	掛ける事ができるから	です	。九「一口で	/aozora...
6	ったのは、それがため	です	。二「私はそ	/aozora...
7	語気で私は賛成したの	です	。二十「Kと	/aozora...

• 検索文字列（キー）：

- この欄は、検索時に必ず指定しなければなりません。
- 「キー」に指定できる文字列は、「です」などのように普通の文字列の他に、制限付きの文字クラスを含めることができます。「制限付きの文字クラス」とは、下の例の【国語】や【ばびぶべぼ】のように、文字の候補を列挙したものです。なお、[] は半角文字です。[] の中には、1個以上の文字を列挙することができます。
 - 【国語】語 は「国語」、もしくは、「國語」を意味します。
 - 遊【ばびぶべぼ】 は、「遊ば」、「遊び」、「遊ぶ」、「遊べ」、「遊ぼ」を表します。
- 検索対象のメニューで「本文（正規表現）」を選択している場合、正規表現が利用できます。
 - 『ひまわり』の正規表現は、[Java \(Pattern クラス\)](#)に基づきます。
 - 「本文（正規表現）」以外にも正規表現が利用できる場合があります。使用している言語資料のマニュアルを参照して下さい。

• 前文脈、後文脈：

- この二つの欄に対する文字列の指定は、任意です。
- ここで指定した文字列により、検索文字列（キー）の「前文脈」列と「後文脈」列の内容を制限します。
- 欄の右側の選択メニューでは、指定した文字列の照合方法を指定することができます。照合方法には、次の五つの方法と、それぞれの否定条件を合わせて、計10通りがあります。
 - 「で始まる」
 - 「で終る」
 - 「と一致する」
 - 「を含む」
 - 正規表現
- （上級者向け）指定した文字列は、内部的に正規表現として解釈されます。例えば、「で始まる」で「の」と指定した場合、内部的には「^の」に変換されます。

すでに述べたとおり、「前文脈」、「後文脈」欄が、次のようなインターフェイスになる資料（例：『分類語彙表』サンプル）があります。これは、辞書の見出しを検索する場合のように、前後の文脈を考慮しないで、見出し単独で検索する場合です。詳しい説明は、各資料のマニュアルを参照してください。

全文検索システムひまわり - [『分類語彙表増補改訂版』(研究用データ) サンプル版] - config_bunrui_sample.xml

ファイル 編集 ツール ヘルプ

検索文字列 フィルタ コーパス 検索オプション

表記

正規表現(前) 正規表現

正規表現(後) 正規表現

上の図のインターフェイスの場合、「正規表現(前)」、「キー」、「正規表現(後)」欄の文字列をこの順序で連結したものが、(正規表現での)検索文字列になります。ただし、「キー」欄は(正規表現でない)通常の文字列しか指定できません。

上の例では、『分類語彙表』の「表記」(見出し)に対して、正規表現の「^国語\$」で検索を行っています。なお、^ は文字列の先頭を、\$ は文字列の末尾を表します。これにより、「国語」の見出しだけを検索し、「母国語」や「外国語」は検索されなくなります。

4.2 絞り込み条件(フィルタ) [↑](#)

「検索文字列」タブの隣の「フィルタ」タブをクリックすると、次のように、検索時のフィルタ用の欄が現れます。フィルタでは、下の例のように、検索結果を絞り込むための条件を指定することができます。

検索文字列 フィルタ コーパス 検索オプション

Path で始まる

タイトル で始まる

著者 と一致する

no 前文脈 キー 後文脈 Path

で始まる
で始まる
と一致する
で終る
を含む
正規表現
と一致しない

- 絞り込む対象は、各欄の左にある選択メニューで選択します。選択肢は、検索結果の列の名前と対応しています。
- 文字列の照合方法は、各欄の右側の選択メニューで選択することができます。この部分は、前文脈、後文脈欄と同様です。

4.3 コーパスの選択 [↑](#)

検索対象のコーパスが複数あるときは、次のように、「コーパス」タブに複数のコーパス名が表示されます。図の左側が現在、検索対象となっているコーパスで、右側が検索対象外になっているコーパスです。

「検索対象外」欄のコーパスを選択し、「追加」ボタンを押すと、検索対象となります。逆に、「検索対象」欄のコーパスを選択し、「除外」ボタンを押すと検索対象外になります。

検索文字列 フィルタ コーパス 検索オプション

検索対象

参議院/本会議(001-048回)
参議院/本会議(049-096回)
参議院/本会議(097-144回)
参議院/本会議(145回-)
衆議院/予算委員会(001-048回)

←追加

除外→

検索対象外

参議院/予算委員会(001-048回)
参議院/予算委員会(049-096回)
参議院/予算委員会(097-144回)
参議院/予算委員会(145回-)

4.4 検索オプション [↑]

「文脈」タブ [↑]

ここでは、検索時に文脈に関係するオプションを設定します。

- **前後文脈長:** 検索結果として表示される前後文脈の長さを指定します。
- **検索範囲:** 前後文脈を検索条件としたときに、絞込みの対象となる前後文脈の範囲を設定します。
- **キー範囲:** このオプションをチェックすると、前後文脈欄で指定した文字列にマッチした文字列を検索結果の「キー」欄に含めることができます。

「抽出」タブ [↑]

このタブでは、検索結果の出力方法を制御します。

- **全数:** 一致した結果をすべて「検索結果」欄に出力します。ただし、「抽出数上限」欄で数値を指定することにより、出力数の上限を設定することができます。「抽出数上限」欄が空欄の場合は、全部の結果が「検索結果」欄に出力されます。
- **ランダム:** すべての検索結果から、ランダムに検索結果を抽出します。抽出数は、「サンプル数」欄で指定します。
- **頻度計測のみ:** 検索結果を「検索結果」欄に出力しないで、検索結果の頻度計測結果のみ出力します。次の二つのオプションが使えます。
 - **一覧:** 選択した「検索結果」欄の列を対象に、値の異なりとその頻度を計測し、一覧表示します。例えば、「タイトル」「著者」列を選択した場合、特定の著者の作品ごとの検索総数が一覧表示されます。列の選択は、何か検索を行い、当該列の任意のセルを選択状態にして下さい。
 - **総計:** 検索総数のみ表示します。

「字体」タブ [↑]

このタブでは、字体変換におけるオプションを設定します。字体変換機能が提供されていない資料では、表示されません。どのような字体に変換されるかは、資料ごとに異なりますので、詳しい内容については、個々の資料のマニュアルを参照してください。

検索文字列	フィルタ	コーパス	検索オプション
文脈	抽出	字体	
<input checked="" type="checkbox"/> 等価字体		<input type="checkbox"/> 参考字体	

- 等価字体: このオプションをチェックすると、検索対象の資料において異体字として等価的に使用される字体を変換後の字体に含めます。
- 参考字体: このオプションをチェックすると、検索対象の資料において異体字として等価的に使用されることのない字体を変換後の字体に含めます。

なお、標準添付される字体辞書は、常用漢字の新旧字体を記述したものです。したがって、「等価字体」「参考字体」の区別はありません。さらに、単純に異体字の関係にある字体に変換するだけであり、検索対象の資料において、「等価字体」「参考字体」の関係にあるとは限りません。

[Prev](#)[全文検索システム『ひまわり』/利用者マニュアル/1_6/3. 検索してみよう](#)[Home](#)[Up](#)[Next](#)[全文検索システム『ひまわり』/利用者マニュアル/1_6/5. 検索結果を集計する](#)

Last-modified: 2018-07-18 (水) 12:24:43 (44d)

Site admin: [anonymous](#)

PukiWiki 1.4.7 Copyright © 2001-2006 [PukiWiki Developers Team](#). License is [GPL](#).
Based on "PukiWiki" 1.3 by [yu-ji](#). Powered by PHP 5.1.6. HTML convert time: 0.400 sec.