

国立国語研究所学術情報リポジトリ

全文検索システム『ひまわり』を用いた分類語彙表の検索と閲覧

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 山口, 昌也 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002960

全文検索システム『ひまわり』を用いた分類語彙表の検索と閲覧

独立行政法人 国立国語研究所 研究開発部門

山口昌也

1. はじめに

国語研究所では、分類語彙表を含めて、言葉に関するさまざまな資料を電子的なテキストの形で一般に公開しています。分類語彙表以外にも、20世紀初頭の総合雑誌「太陽」を収録した『太陽コーパス』や、大量の自発音声を収録した『日本語話し言葉コーパス』などがあります(「コーパス」とは言語研究用の資料のことです)。

全文検索システム『ひまわり』は、このような言葉に関する資料から、指定された言葉を簡単に検索して、言葉に関する分析をするためのツールです。今回は、『ひまわり』を使って、分類語彙表を検索、閲覧するデモンストレーションを行います。

2. 検索と閲覧の例

まず、『ひまわり』を使った簡単な検索例を紹介します。下の図は、「科学」が含まれる見出しを『分類語彙表』から検索した結果です。当該語の「読み」、「分類番号」などとともに、「表記」欄の左右の欄に、意味的に類似する語が数語表示されます。『分類語彙表』の中で「科学」が属している分類項目「学科・学問」に含まれる語全体も Web ブラウザで表示できます。

類似する語(前)	類似する語(後)	読み	分類番号	記号
ライフサイエンス	ライフサイエンス	かいはく	1.3074	04
運動科学	運動科学	かいはく	1.3074	04
サイエンス	サイエンス	かいはく	1.3074	04
科学社会学	科学社会学	かいはく	1.3074	04

体/活動/心/学問・学科 (1.3074)

01

- 学(04) 学術(0401) 学芸(0401) 学問(0401) 学業(0401)
- 文武(0401)
- 科学(0401) サイエンス(0401)
- 俗学(0401) 雑学(0401)
- 実学(0401) 耳学問(0401) 死に学問(0401)
- 先学(0401) 後学(0401)
- 百学(0401) 異学(0401) 家学(0401)

注:

「類似する語(前)」「類似する語(後)」欄:『分類語彙表増補改訂版』において、当該語の前後にある語
「概念パス」欄:『分類語彙表増補改訂版』の最上位階層から当該語までの分類項目を列挙したもの

3. 『ひまわり』で検索するには

実際に検索できるように数台のPCをご用意しています。簡単な検索方法を下に示しますので、機会があれば、実際にお試してください。

(1) 検索方法の選択

表記: 『分類語彙表』の「表記」項目をキーとして検索

読み: 「読み」項目をキーとして検索

(例: 「かがく」、「ばあそなるこんびゅうたあ」)

分類番号: 「分類番号」項目をキーとして検索 (例: 1.3074)

(2) 検索したい文字列の入力

検索文字列欄に検索したい文字列を入力し、「検索」ボタンを押してください。入力した文字列を含む語が検索されます。いろいろな検索例を下に示します。

「人」を含む語を検索

「人」で始まる語を検索

「人」で終わる語を検索

その他の機能

- ◆ 検索結果をダブルクリックすると、当該語と同じ分類番号を持つ語すべてをブラウザで閲覧することができます。
- ◆ 検索結果の列名の部分をクリックすると、クリックした列をキーにして並べ替えが実行されます。
- ◆ [ツール]→[分類一覧]で全分類を一覧表示することができます。各項目をダブルクリックすると、当該の分類番号を持つ語すべてをブラウザで閲覧することができます。

4. 『分類語彙表増補版』, 『ひまわり』の入手について

『ひまわり』は当研究所の Web ページ(<http://www.kokken.go.jp/lrc>)から無料でダウンロードすることができます。『分類語彙表』のサンプル版も『ひまわり』に同梱されています。ぜひ、試してみてください。なお、本デモンストレーションと同内容の検索を行うには、『分類語彙表増補改訂版』(研究用データ)が必要です。入手方法などの詳細は、上記の Web ページをご参照ください。