

階層構造をもつ用語データのためのBrowsing Tool

○後藤貴信*、鈴木祐介*、後藤智範*

The Browser for Technical Terms with Hierarchical Structure

Thesauri have been widely used in bibliographic databases for 30 years. Recently, CD-ROMs of a variety of dictionaries with their GUIs are spreaded to current users. On the other hand, End users dose not use thesauri as for their bulky printed matter. The king of software browsing graphically and managing thesauri does not appear in PC environment..

The browsing tool for lexical database with hierarchical structure has been developed using Java.. This paper describes the functions, the components, and examples of its usage. The problems of the browser and the functions to be extended are discussed.

1. はじめに

シソーラスが書誌データベースに利用されてから約30年が経過している。最近では、自然言語処理研究およびその応用ソフトウェアにも利用され、その応用範囲が拡大しつつある。現在、多くの辞書がCD-ROM化され、そのGUIソフトによって広く利用されているのに対し、検索ツールとしてのシソーラスは、未だに冊子体での利用が主であり、一般利用者には有効に活用されていない。シソーラスのグラフィカル表示についての研究は、1960年代のL. B. Doyleの“semantic road map”の研究^[1]に溯ることができるが、汎用性の高いシソーラスブラウザの普及には至っていない。

本研究室では、このような状況を改善するためにシソーラスおよび階層構造をもつ用語データを多面的なアクセスを可能とし、特定のシソーラスに限定されない汎用性の高いBrowserとその応用ソフトウェアの開発を行ってきた^{[2][3]}。最初のブラウザはUNIXで標準的に使用されているXtoolkitを用いて試作を行った^[2]。一昨年度から、Javaで再記述を行い、年々機能強化を行ってきた。本稿では、当ブラウザの最新バージョンの機能、利用実例、検討課題について報告する。

2. Browser の機能

用語間に階層構造を有するデータ多角的に、言い換えればその構造からアクセスするためのブラウザは、下記に挙げる機能を有することが必要とされる。

- (1) 見出し語の迅速かつ簡便な検索・表示。
- (2) 特定された見出し語の上位語ー下位語のグラフィカル表示。
- (3) 特定された見出し語の関連語の表示。
- (4) 履歴表示
- (5) 複数シソーラスの同時使用
- (6) マウス操作主体

3. Browser の構成その利用事例

上記の要件・機能の満たすべく Java(JDK1.1.7)で実装を行った。動作の検証のためのシソーラスとして下記の2種類を用いた。

- (1) 1987年度版INSPECシソーラス 用語数: 約9000語(ディスクリプタ:約5500語)
- (2) 1987年度版JICSTシソーラス 用語数: 約50000語(ディスクリプタ:約40000語)

本ブラウザは、4つの window コンポーネントから構成される。

- (1) メイン window
- (2) 用語リスト window
- (3) 用語木 window
- (4) 関連語 window

(1) は最初に表示される window で、シソーラスの選択および、履歴ファイルの指定による過去の使用用語の履歴表示のために用いられる。

(2) は見出し語を表示するための window で、JICST シソーラスを用いた実例を図 3.1 に示す。利用者が(1)で JICST シソーラスを選択すると、用語木 window は空白状態で表示される。次に、利用者が探したい用語の先頭文字を含む行(「0-9」～「わ」)のボタンをマウスクリックすると、

当該行の文字(濁音などを含む)のポップアップメニューを表示する(この例では、「か」行)。利用者が先頭文字のボタンを押すと、先頭文字から始まる見出し語

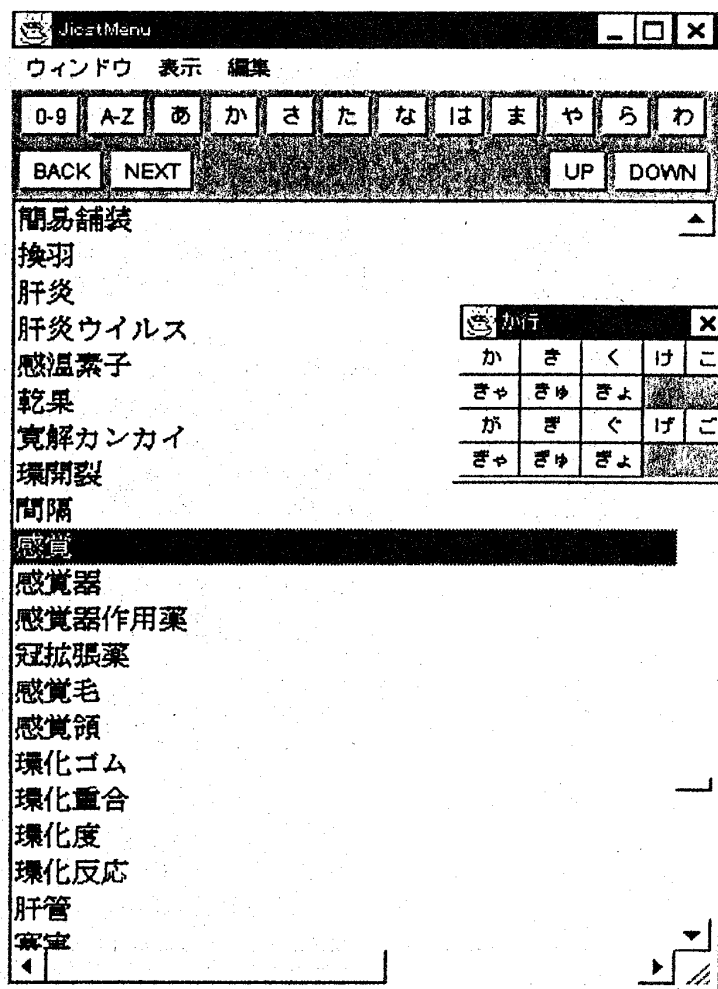


図 3.1 用語リストウインドウ

が表示される。図 3.1 はこの状態を示しているが、ポップアップメニューは見出し語が表示されると瞬時に消える。右上の「UP」「DOWN」ボタンは、右側のスクロールと同等の機能を有し、「NEXT」ボタンは、以前に指定された用語を迅速に表示させるためのボタンである。

対象用語の総数が数万語以上になると、当該用語を見つけるためのスクロールが必要とされ、これが利用者に多大な手間と時間を課することが最初の browser の試作時に判明した。ポップアップメニューの導入により、

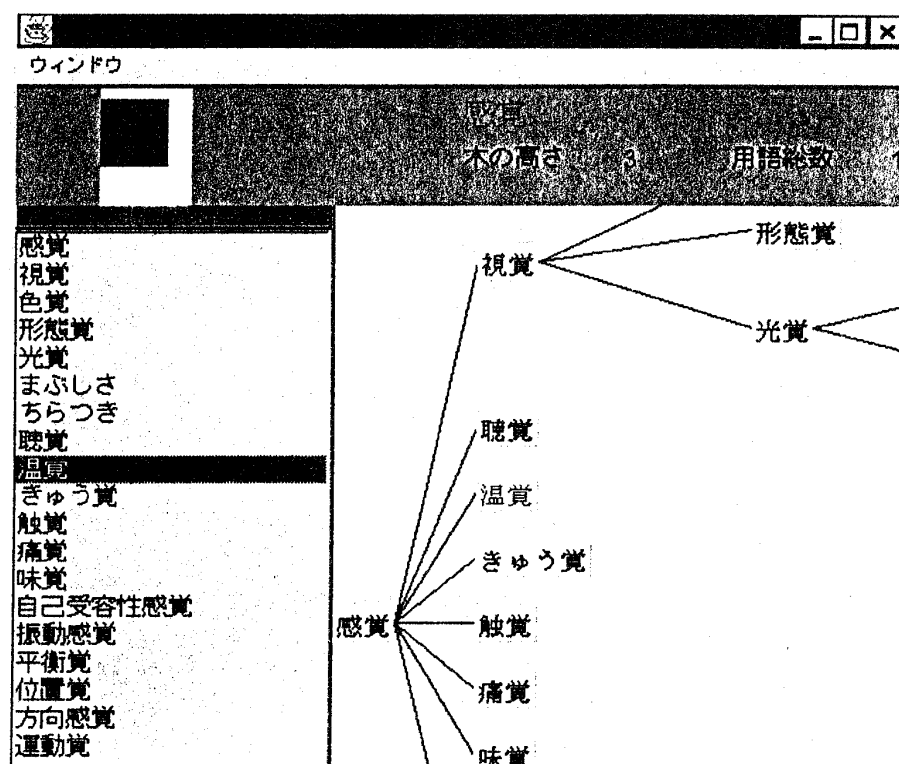


図3.2 用語木ウインドウ

見つけたい用語の範囲が限定されるためこの負担が軽減されようになった。

用語リストwindowに表示される用語を指定し、上の「表示」ボタンを押すか、または右クリックのポップアップメニューにより、「NT-BT」、「RT」ボタンが現れる(図3.1では「感覚」を指定)。「NT-BT」ボタンが押されると、用語木windowが表示される。図3.2にこの例を示す。

用語木windowは、(a)階層表示、(b)二次元カーソル(左上のボックス)、(c)構成用語リスト、(d)用語木情報、の4つのコンポーネントから構成されている。従来は、(a)だけであった。シソーラスを構成する木の大きさ(用語数)、木の高さは様々であり、通常のディスプレイの解像度(1024×768～1600×1200)の範囲を超えるような木も数多く存在する。用語木を構成する個々の用語を閲覧するためには、縦横のスクロールバーによる操作が必須となるが、木が大きい(100語以上)となると、この操作は利用者に多大な負荷を課す。(b)および(c)はこの操作を軽減するために導入したコンポーネントである。二次元カーソルは、Xtoolkitにも含まれるコンポーネントであり、これにより縦横のスクロールバーは不必要となった。左に位置する用語windowの用語をクリックすると、階層構造コンポーネントはその用語が中央に位置するように再描画する。この2つのコンポーネントの導入により、100語以上からなる大きな用語木についても、自在にその構成用語およびその上位語一下位語を知ることが可能となる。

上位一下位語、および関連語の表示は、用語が示されるどのwindowからでも表示できる。これは、階層構造コンポーネントに表示されている用語ノードからでも可能である。例えば図3.2において、「聴覚」、「視覚」のノードでマウスの右クリックを行えば、ポップアップwindowが表われ、関連語windowが表示される。

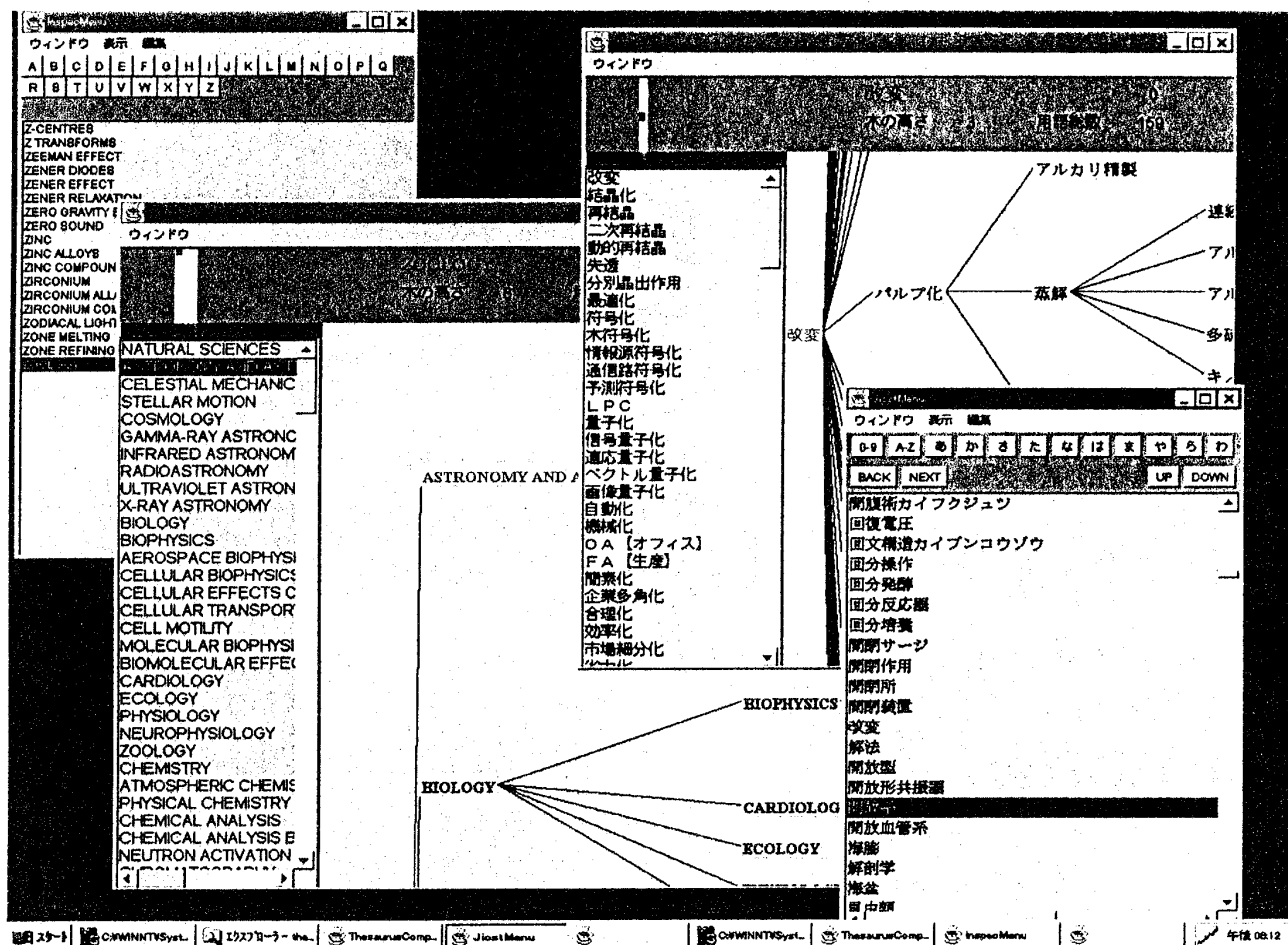


図3.3 複数シソーラスの利用例

本ブラウザは、ブラウザが稼動するPC(windows95/98/NT), WS(UNIX)のメモリーの容量が許す限り、同時に(1つのブラウザで)複数のシソーラスを利用することができる。図3.3は、INSPEC、JICSTの両シソーラスを同時に利用している例を示している。

複数のシソーラスの用語、木構造の比較などの目的に対しては、本ブラウザは極めて有効なツールとなる。

4. 検討項目と今後の課題

本ブラウザは、現在下記の機能拡張および性能向上のための改良を行っている。

- (1) SWING 対応
- (2) 見出し語を構成する単語(語基)からの検索
- (3) 構成用語リスト(用語木 window)のソート(用語およびその高さ)機能
- (4) 用語木に対するズーム機能
- (5) シソーラスファイルの読み込み速度の向上

さらに、新規機能として下記に挙げる編集機能を追加することを予定している。

- (1) 新規用語の登録・用語木の生成(シソーラスの新規作成)
- (2) 既存用語群の変更—用語木の編集(シソーラスの変更)

(a) ノードの変更、削除、(b) 部分木の削除、追加、(c) 部分木の合成

上記の機能を満たした時点において、本ブラウザを **freeware** として登録し、シソーラスの利用を広げることが計画している。

5. おわりに

近年、辞書・事典データの GUI^{[4][5]}、ハイパーテキスト化^{[6][7]}の研究が行われている。また、人工知能分野では、これらの研究に並行して、思考支援システム、ドメインオントロジー支援ツールの研究が行われている。いずれもその中心課題は用語—概念の表現、操作である。

本ブラウザは、可能性研究ではなく実用指向および汎用性を意図して開発したため、その機能をシソーラス、分類カテゴリーなどの階層構造をもつ用語データに限定しているが、本ブラウザのポップアップメニューに意味、訳語などの選択項目の追加、さらのオブジェクトの属性の拡張を行うことにより、汎用的な辞書 GUI に発展し、さらに推論機能を付加することにより思考支援システムとは親戚になるものと考えている。

参考文献

- [1] Doyle, L.B. "Indexing and abstracting by association" American Documentation. vol.13, no.4, pp.376-390(1962).
- [2] 下村 央人 他.『汎用シソーラスブラウザの試作』.第 33 回情報科学技術研究集会発表論文集,pp.99-105(1997).
- [3] 下村 央人 他.『JAVA による索引支援ブラウザの試作 2』.第 35 回情報科学技術研究集会予稿集,pp.133-137(1998).
- [4] 藤澤浩道 他. 概念ブラウザと個人情報環境. 情報処理学会研究報告. No.53, 情報メディア—7, pp.41-48(1992).
- [5] 都司達夫 他. 仮想電子辞書の設計. 情報処理学会研究報告. No.106, 情報学基礎—48, pp.27-34(1997).
- [6] 黒橋禎夫 他. 専門用語辞典の自動的ハイパーテキスト化の方法. 人工知能学会誌 Vol.7, No.2, pp.336-345(1992).
- [7] 内藤広志 他.オンライン辞書のハイパーテキスト化手法. 情報処理学会論文誌. vol.34, pp.320-330(1993).

* 神奈川大学理学部情報科学科
Department of Information and Computer Science, Kanagawa University