

N. 安全用語を対象とした Java による索引支援ブラウザの試作

山本晴彦*、河村 正一*、高木 伸司*、小幡 行雄*、後藤 智範*、藤原 鎮男**

* 神奈川大学理学部、平塚市土屋2946

** 東京大学名誉教授

1. はじめに

安全問題は環境問題とともに今日大きなテーマとなっており、様々な方向から研究が進められている。しかしながら、安全問題は学際的な要素が極めて強く、多分野にまたがる問題であるため、情報が複数の分野の多数のデータベースに分散している。例えば、労働環境に起因するアレルギー問題の情報は、労働安全衛生、医学、化学物質などに関するデータベースに分散している。したがって、安全問題のような学際的研究に関するテーマの情報収集には、複数分野の多数のデータベースについて検索をする必要がある。しかし、現在のオンライン情報検索の環境は次に挙げる問題を有し、学際的研究の情報収集は非常に困難な状況にある。

- (a) 個々のデータベース毎に、個別の分類体系、キーワード体系(シソーラス)を持っているため、その利用にあたってはそれらについての専門的知識を必要としている。
- (b) これらの分類体系およびキーワード体系についての資料は、大部な冊子形態での利用が多く、多様な専門知識と多くの時間・労力を必要とする。

筆者らは、上述の安全研究の情報アクセスの問題を解決することを意図し1992年度来下記に挙げる研究を行ってきた。

- (1) 安全用語の収集ための事例研究^[1]
- (2) 安全用語の出現頻度分析に基づく科学分野の80の書誌データベースの調査^{[2], [3]}
- (3) 化学、医学、原子力など5分野に関連する安全用語の用語データベース構築^{[4], [5]}
- (4) 安全用語データベースを用いる多重シソーラスシステムのモデル^[6]

上述のこれまでの研究は用語データベースの構築に重点を置いていたが、本研究は、文献データベース検索および作成に焦点を移し、インターネット環境で用語データの知識構造に沿った表示、および索引付けの支援を意図したものである。

本研究では、安全用語データベースを作成する過程において主要な作業である索引作業をインターネット上で効率的に行えるソフトウェアの試作を意図したものである。従来の印刷媒体のシソーラスを用いた方式と比較して、極めて効率的に行なうための要件、機能を提案し、これをjava言語およびその仕様の一部であるAWT(Abstract Windowing Toolkit)を用いて実装を行う。

2. 索引作業の現状

2. 1 索引作業環境

現在行われている索引作業は図 2.1 の上段に示される環境でなされている。すなわち、索引作業に必要な全ての情報、すなわち、対象資料、記入シート、シソーラスが印刷媒体で提供される。索引語の登録は記入シートに手書きで書き込まれる。

図 2.2 の下段は、索引作業のインターネット上での支援環境が整備された状況を示している。索引作業に必要な全ての情報が電子媒体で提供され、支援ツールを用いて索引作業が行われる。このような状況が実現すれば、書誌データベース作成において主要な位置を占める索引作業は迅速かつ効率的に行うことができよう。本ブラウザの試作は、このような環境の構築を意図したものである。

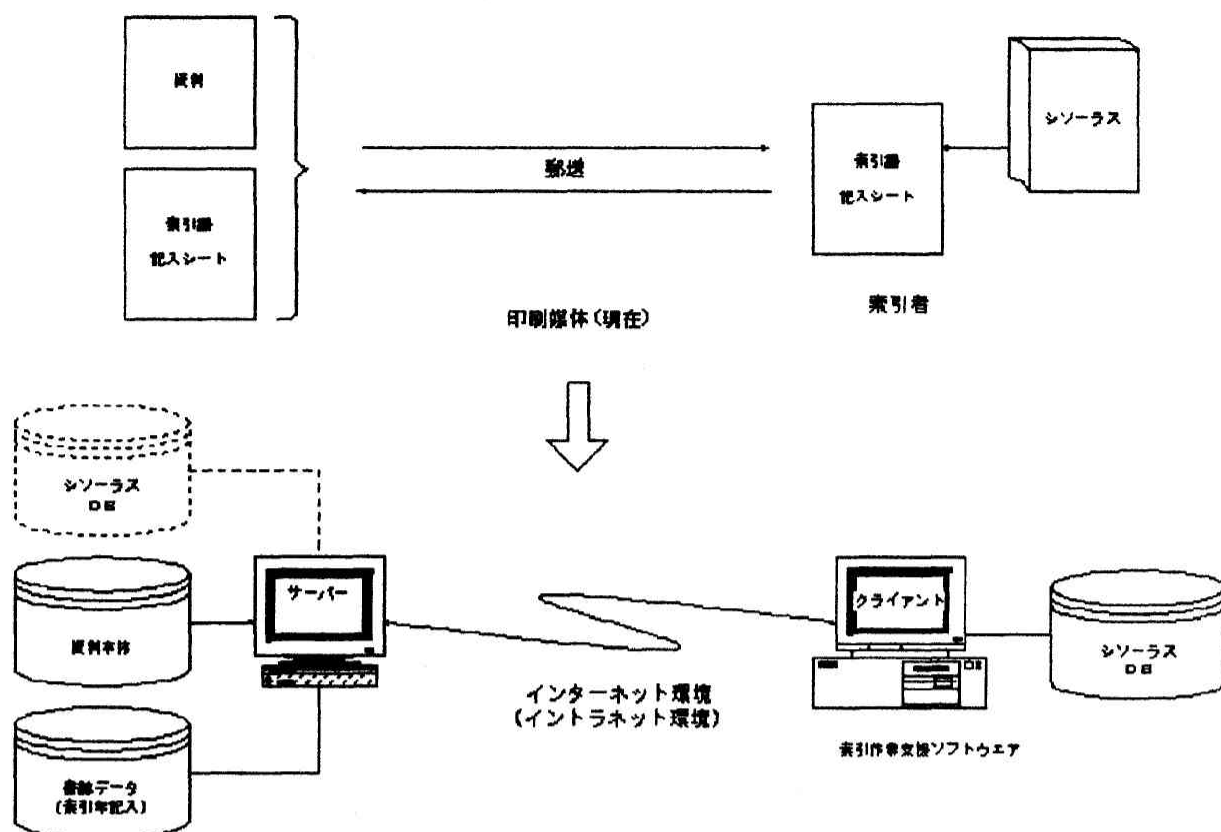


図 2.1 索引作業環境

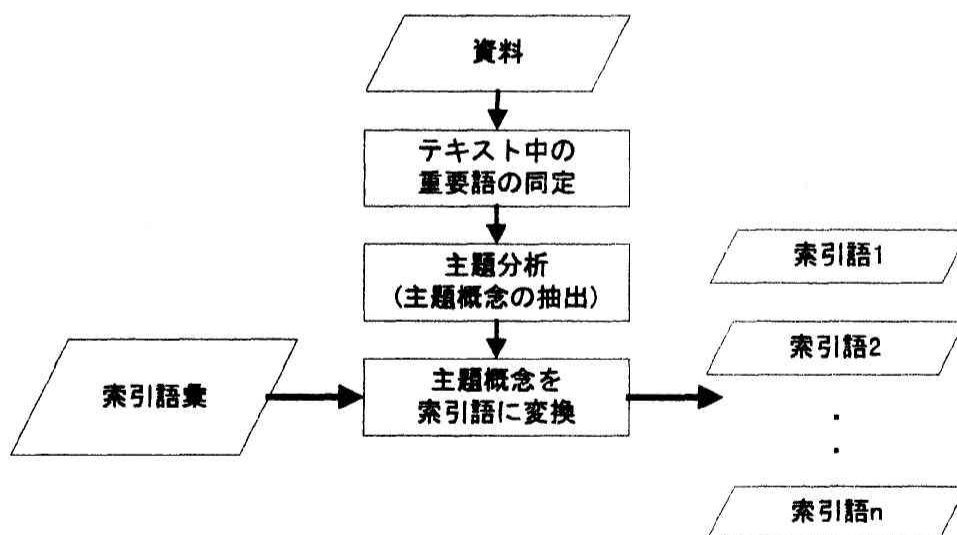


図 2.2 索引過程

2. 2 索引過程

索引過程(作業)は、一般に図 2.2 に示す手続きからなる。索引過程において、時間を要するのは同定された個々の主題概念に対し、それぞれ適合する索引語(ディスクリプタ)をシソーラスから見つけ出し、それらを所定のシートに記入する過程である。特に索引語彙、シソーラスが印刷媒体である以上、適合する索引語をシソーラスから見つけ出すためにはシソーラスの頁をめくるといった時間を要する行為が必要される。資料、シソーラス、記入用シート

がディスプレイ上に表示され、さらにシソーラス上のディスクリプタの探索・記入が迅速に行われれば、索引作業の効率は著しく改善し、書誌データベースの作成コストの削減にも大いに貢献すると考えられる。

3. 索引支援ソフトウェアの要件

上述の索引作業環境を電子媒体に代替し、効率的に行うためのソフトウェアは、従来の印刷媒体のシソーラス、索引語記入シートを用いた方法と比較して、作業時間およびその利便性が著しく改善される必要がある。したがって、下記項目について迅速な操作が求められる。

(1) シソーラスから当該ディスクリプタの迅速なアクセス

索引者が当該文献に対して適合していると考えられるディスクリプタおよび関連するディスクリプタへのアクセスおよび表示が容易でかつ迅速に行われる。

(2) ディスクリプタの階層構造の直感的表示

例えば、図 4.4 に示されるような表示方法が考えられる。図 4.4 に示されるような表示形態では、索引語を付与する過程において、当該ディスクリプタを中心にその上位語、下位語の階層構造が直感的に理解できるため、当該文献の主題の深さに適合するディスクリプタが何かを容易に判定できる。

(3) ディスクリプタの迅速な入力

当該資料に対する主題概念を表わすディスクリプタを見つけ、これを入力する作業が迅速かつ容易に行える。すなわち、ディスクリプタの登録がキーボード入力ではなく、マウス操作で行えることが望ましいであろう。

(2)、(3)の操作が従来の方式において、シソーラスの頁をめくって当該主題を表わしているディスクリプタを探し、これを記入するという操作の電子的代替であ。

(4) 作業結果の迅速な伝達

また、作業結果、すなわち付与された索引語(ディスクリプタ、アイデンティファイア)が電子的に転送できるように、図 2.1 下段に示されるようにサーバークライアント環境で行われる、すなわち上述の機能を持つソフトウェアはクライアントソフトウェアであることが有望であろう。これは、科学技術振興事業団のように機関外に多くの索引者の支援を得ているような状況では、インターネット上で作業結果を転送できるため極めて有効であると考えられる。逆に、従来データベース作成機関において、機関内で行っていた索引作業を外部で行うことも可能になる。

本研究では、上述の要件を満たす索引支援ブラウザを開発するために、近年広く普及しつつあるJAVA[7][8][9]を用いて試作した。JAVAを用いたのは下記の理由による。

- (1) C言語およびXtoolkitによるGUIはUNIXワークステーションだけで稼動するもので、Windows95/NT環境では稼動せず、汎用性に欠ける。
- (2) Xtoolkitは非オブジェクト指向であり、再利用が困難である。また、その部品であるAthena Widgetは柔軟性が欠け、機能の変更が困難である。一方、JAVAのGUI KitであるAWT(Abstract Windowing Toolkit)はクラス化されており、オブジェクト指向の継承機構を用いて本ブラウザに適合するコンポーネントを容易に作ることができる。
- (3) オブジェクト指向プログラミング言語の特性を生かし、本ブラウザを構成するクラスは類似のアプリケーションプログラム(例えば、シソーラスブラウザ)に再利用することができる。
- (4) C言語+Xtoolkitでは、プログラム作成およびデバッグに非常に時間がかかる(本ブラウザの作成機関は約2ヶ月/2人)。

索引支援ブラウザは、昨年度筆者らがすでに試作した汎用シソーラスブラウザ[10]と機能上の類似が多いため、対象とするシソーラスおよび、そのアクセス方式については、汎用シソーラスブラウザと同等の論理構造を採用し、そのプログラムの一部をもとにJAVAを用いて再記述した。さらに、ブラウザだけではなくjava net classを用いてサーバーをも実装した。

試作した索引支援ブラウザーは下記に挙げるシソーラスおよび文献データを用いて動作テストを行った。

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

4. 索引支援ブラウザーの構造と機能

4. 1 Window 構成

本ブラウザーは下記の3つのWindowから構成されている。

- (1) 書誌データ Window (2) 用語リスト window (3) 木構造 window

図 4. 1A から 4. 1B に本ブラウザーの GUI コンポーネントの構成を示す。

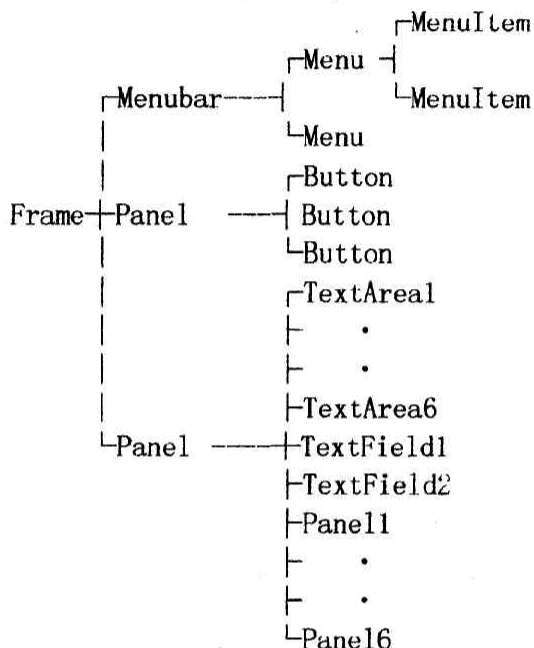


図 4. 1A GUI コンポーネント(書誌データ Window)

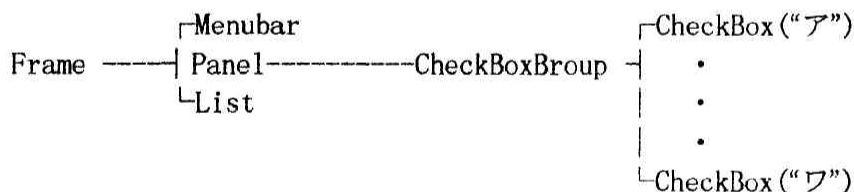


図 4. 1B GUI コンポーネント(用語リスト Window)

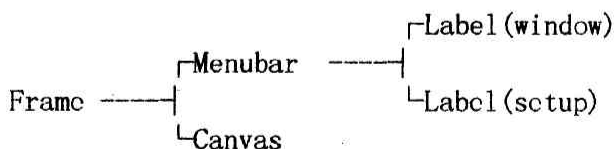


図 4. 1C GUI コンポーネント(木構造 Window)

window を構成する様々なコンポーネントは、AWT のレイアウト機能を用いて配置され、書誌データ Window は GridBagLayout、用語リスト window および木構造 window は共に BoaderLayout を用いている。汎用シソーラスブラウザーでは、シソーラスの木構造表示は、

Athena Widget で提供されている Tree Widget を用いたが、現在の AWT には木構造表示を実現するコンポーネントは提供されていない。このため、木構造表示を行う window を本ブラウザでは独自に作成した。ディスクリプタの木構造は図 4.1C の Canvas すなわち、グラフィックデータとして描画するように設計した。

次節でこれらの Window の機能について説明する。

4. 2 機能

(1) 書誌データ window

本ブラウザを稼働させる前にサーバーを起動させる。サーバーの IP アドレスを引数として本ブラウザを立ち上げる。起動すると対象シソーラスデータが読み込まれ、サーバーとの通信の準備が行われ、図 4.2 に示される書誌データ window が表示される[11]。

ここで next ボタンをマウスでクリックすることにより、サーバーから書誌データ(抄録付き)がクライアントであるブラウザに転送され、この書誌データ window に表示される。

(2) 用語リスト window

書誌データ window 上の Thesaurus ボタンをマウスクリックすると、図 4.3 に示す用語リスト window が表示される。この例では、JICST シソーラスの先頭の見出し語が表示されている。当ブラウザの開発途中で、対象シソーラスの用語数が多い場合、カーソルを移動しディスクリプタを見つけるのに手間取り、時間がかかることが判明したため、window 上部にアイウエオの小見出しを付加した。

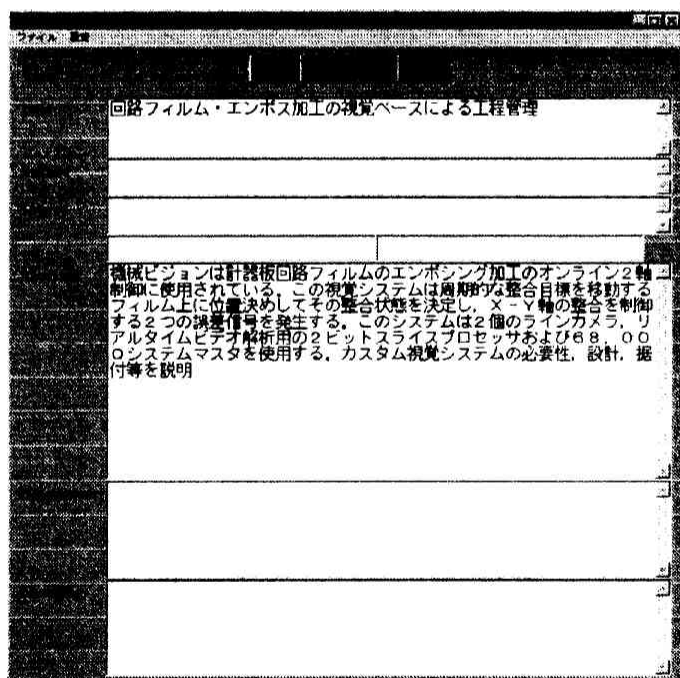


図 4.2 書誌データ window

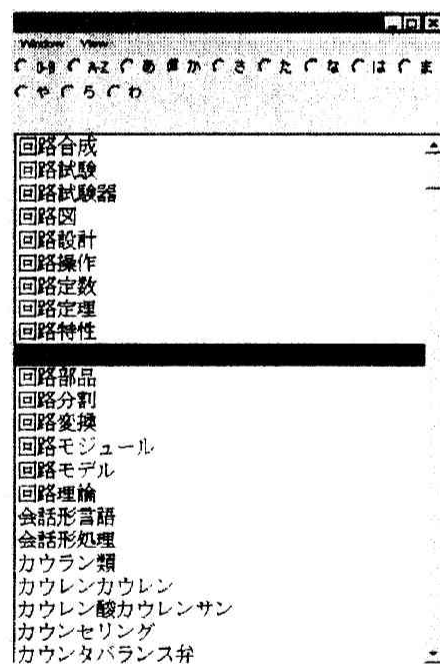


図 4.3 用語リスト window

(3) 木構造 window

用語リスト window のカーソルを移動して所定のディスクリプタ(この例では、「回路形成パターン」)を同定し、マウスで上部の View ボタンに触れ、tree ボタンを押すと、木構造 window が表われ、当該ディスクリプタの上位一下位語の木構造が表示される。図 4.4 はこの状態を示している。このとき当該ディスクリプタが window の左右中心の位置に、異なった色で表示されるように設計した。このような表示形態は、木が window に表示しきれない場合にも

当該ディスクリプタは window の中心に表示されるため、カーソルを移動してディスクリプタを探す手間が省力できる利点がある。

図 4.2 書誌データ window

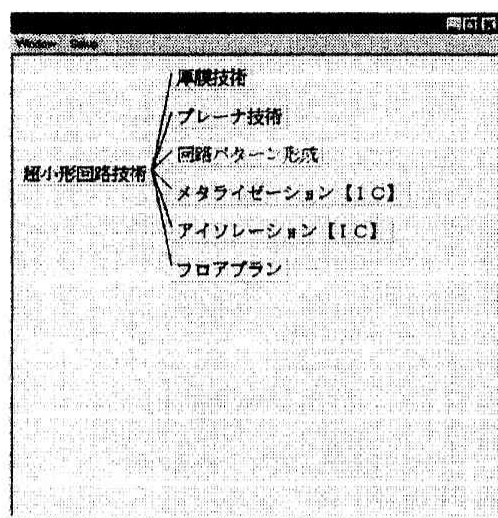


図 4.3 用語リスト window

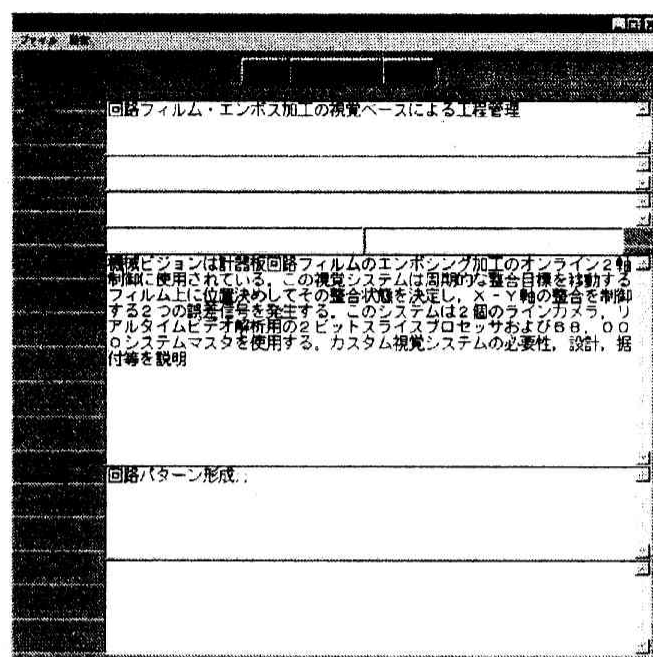


図 4.4 木構造 window

図 4.5 登録された状態

(4) ディスクリプタの入力

上述の操作の後で、書誌データ window の空欄になっているディスクリプタ項目にディスクリプタを入力する必要がある。本システムでは、第 2 章で述べたようにキーボード入力を全く必要とせず、木構造 window 上に表示されている任意のディスクリプタの上にマウスを移動し、マウスをクリックすると書誌データ window のディスクリプタ項目に瞬時に入力される機構を実現した。図 4.5 はこの状態を示している。この機構は、JDK1.1 で初めて提供された listener-adapter インターフェースを、それぞれの window のクラスに実装することにより実現している [12][13]。なお、ディスクリプタ項目にはキーボードからも入力でき、また誤った入力は消去することもできる。

(2) から (4) の操作を入力すべきディスクリプタの数だけ繰り返し、所定のディスクリプタを入力し終えた後で、書誌データ window 上部の send ボタンをマウスで押すと、ディスクリプタが付加された書誌データがサーバーに転送される。

5. 考察

本ブラウザーを用いた索引作業は、図 2.1 の上段に示される印刷媒体に基づく索引作業環境と比較すると、特に作業時間の側面において下記に挙げる利点がある。

(1) 当該資料の主題を表わすディスクリプタをシソーラスから見つける過程

(2) 同定されたディスクリプタを付与する(登録する)過程

本ブラウザーの使用は、これらの 2 項目の作業時間が短縮され、また全て電子的に電送されるので、データベース作成に要するコストも大幅に低下することが期待される。

しかしながら、現時点において、様々な文献データに対して本ブラウザーを試用した結果、JVM (JAVA 仮想マシン) で稼動することによる動作速度の低下とは別に下記に挙げる 2 点の問題点が判明した。

- (1) 本ブラウザの起動時間
- (2) 用語リスト window の切り替え時間

(1) はシソーラスの用語数が多い、具体的には数万語以上になる場合、昨年度試作した汎用シソーラスブラウザと同様に起動時間が 1 分余りかかる。(2) もシソーラスの用語数に起因する問題であり、「ア行」用語から例えば「サ行」の用語を表示するのに、数十秒の時間を要する。この問題は、次のような構造にすることにより解決可能であると考えられる。

- (1) シソーラスファイルの読み込み時に、「ア」「イ」「ウ」…の 50 音で始まる語に分割して格納する。
- (2) 「ア」～「ワ」の小見出しをプルダウンメニューにし、「ア」では、「ア」～「オ」を表示し、(1) で分割されたシソーラス用語を表示する。

上述の変更により、ディスクリプタが格納されている変数に約 50 に分割され、その用語総数は元の 1/50 の大きさになるので、表示時間が大幅に短縮され、さらにカーソルを移動してディスクリプタを探す手間も軽減されることが考えられる。

本ブラウザが実用化されるためには、下記の機能を追加する必要があると考えられる。

- (1) 原資料の表示
- (2) 分類記号の表示・登録機能
- (3) 非統制索引語の同定・登録支援
- (4) 作業環境の保存

CAS および日本化学会などの学術雑誌ですすでに行われているように、SGML (Standard Generalized Markup Language) の普及により、学術雑誌が電子的に編集され、SGML 形態でデータベース化される傾向にある。SGML から HTML への変換は容易に行えるため、サーバー側で HTTP の機能および、クライアント側では HTML ブラウザーの機能を実装することにより、(1) の機能が実現できる。

分類記号は階層構造をとっているため tree viewer はそのまま利用できる。このため (2) については容易に実装可能である。

非統制索引語 (Identifier) は、もともとシソーラス用語ではないため、ディスクリプタのような表示は不可能である。一般に、非統制語項目に実際に付与されている用語は、原資料または抄録中に出現する語が多い。したがって、(3) については、重要語抽出機構を実装することにより、候補語を表示しマウスクリックにより登録することは不可能でない。

個々の索引者の主題範囲は限られているので、索引作業にはシソーラス中の一部の主題分野の用語が用いられる。ある索引対象文献に用いられたいくつかの用語の木は、他の索引対象文献にも用いられる可能性が非常に高い。言い換えれば、個々の索引者にとって、文献毎に全くことなる用語の木を用いることは希で、いくつかの木はほとんど常に共通して用いられる可能性が非常に高い。したがって、(4) の機能により、索引作業が終了した時点で、その際に用いられた用語の木を保存しておけば、次回には用語 window から探さずに済み、索引作業を迅速に行うことができる。

6. おわりに

文献データベースの利用拡大を阻害する要因として、その利用コストの高さが挙げられる。ほんの 20 分間使用し、十数件を出力しただけで数千円から一万円以上を要する。文献データベースの利用コストの高さの主な要因は、その作成コストであり、特に索引作業には人件費が要する。

本ブラウザが実用レベルに達し、図 2.1 の下段に示されるようにインターネット上で迅速かつ効率的な索引作業が実現できれば、データベース作成コストの削減に寄与するものと期待される。

参考文献

- [1] 後藤智範、山本晴彦、他. "安全研究の関する重要語の収集・評価のための事例研究". 情報の科学と技術. vol.42, no.11, p.1065-1074(1992).
- [2] Gotoh,T., et al. "Survey of Effectiveness of Databases for Safety Investigation based on Quantitative Analysis to the Important Terms by Cross-file Searching". The 46th Conf. of IFID. Madrid. Sept. 1992.
- [3] 後藤智範、山本晴彦、他. "データベース構築促進および技術開発に関する報告書 安全研究における多重シソーラス・システム構築のための基本安全用語データベースの開発". データベース振興センター. 1993年3月.
- [4] 後藤智範、山本晴彦、他. "データベース構築促進および技術開発に関する報告書 安全研究における多重シソーラス・システム構築のための基本安全用語データベースの開発". データベース振興センター. 1994年3月.
- [5] Gotoh,T., Yamamoto, H., et al. "Construction of Terminology Database in Safety Research: For efficient accessing safety information from numerous databases". 47th IFID(International Federation for Information and Documentation) Conference and Congress, Omiya, Japan, October 5-8, 1994. p.73-77(1994).
- [6] Gotoh,T. and Fujiwara,S. "Toward the Gateway Thesaurus System for Safety Research:For efficient accessing to safety information from a large number of databases". 情報知識学会誌. vol.4, no.1, p.57-62(1994).
- [7] ローラ・リメイ他著. 武舎広幸 他訳. Java 言語入門. 東京. トッパン. 653p. 1996 年.
- [8] 青柳龍也著. Java API.プログラミングガイド. 工学図書. 452p. 1996 年.
- [9] D. Flanagan 著. 日本サン・マイクロシステムズ株式会社監訳. JAVA クイックリファレンス JAVA IN A NUTSHELL, オーム社, 567p. 1996 年.
- [10] 下村央人、笹川泰秀、後藤智範. 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. p.99-105(1997)
- [11] この節の図 4.2 から図 4.5 の説明は、下記のような環境での実際の使用例に基づいている。
サーバー: OS Solaris2.5 または Windows NT4.0
クライアント: OS Windows NT4.0
- [12] 丸山不二夫. 第 1 回 JDK1.1 の新しいイベントモデル. 月刊サンワールド. vol, 7, no.5, p.96-105(1997).
- [13] 丸山不二夫. 第 2 回 JavaBeans のイベント処理. 月刊サンワールド. vol, 7, no.6, p.100-105(1997).