

オンライン索引支援システム

日夏 健一

日本科学技術情報センター

JICSTが開発したワークステーションによるオンライン索引支援システムについて述べる。同システムは、文献データベース作成工程の一つである校閲・索引作業を支援するエキスパートシステムである。ワークステーションは、JICST科学技術シソーラス、分類コード表、FT-CT変換辞書等を搭載している。これにより、索引者はシソーラス等を意識しない自然語による索引が可能となる一方、均質なデータベースを効率的に作成できる。

Online Indexing Aid System

Ken-ichi Hinatsu

The Japan Information Center of Science and Technology

5-2, Nagatacho 2 chome, Chiyoda-Ku, Tokyo 100, Japan

This paper describes the online indexing aid system developed by JICST, which works under workstation environment. This system is an expert system for indexing bibliographic data base. This system has JICST thesaurus and JICST classification. Using this system, indexer can index by free term without considering thesaurus and classification.

1 はじめに

JICSTは、毎年60万件に及ぶ文献を処理しデータベース化している。

文献のデータベース化には、書誌事項の入力、抄録の作成、索引の付与の各工程が必要である。特に、抄録の作成、索引の付与は多くの専門家の手をわずらわしている。

一方、近年ワークステーション上で実現されるエキスパートシステム構築技術等の発達により、JICSTでもそれらに基づく索引付与と分散処理によるデータベースの作成が課題となっている。

上記を背景としてJICSTでは1988年から索引支援システムの開発に着手し、1990年に第1期システムの開発が終了したので、ここに紹介する。

2 全体システムの構成

2.1 ハードウェア構成

- (1) センタマシン
OKITAC 8300/38 (ミニコン)
センタシステム管理用
- (2) ノードマシン
OKITAC 8300/32 (ミニコン)
ノードシステム管理用
- (3) ワークステーション
if2000Σ
校閲、索引用

2.2 ネットワーク構成

- (1) センタシステム
センタマシンを中心に校閲、索引用のワークステーションがイーサネットによるネットワークを形成。
- (2) ノードシステム
ノードマシンを中心に校閲、索引用のワークステーションがイーサネットによるネットワークを形成。
- (3) センタ・ノード間
INS64 (ISDN) によるネットワークを形成。

ネットワーク構成図を図1に示す。

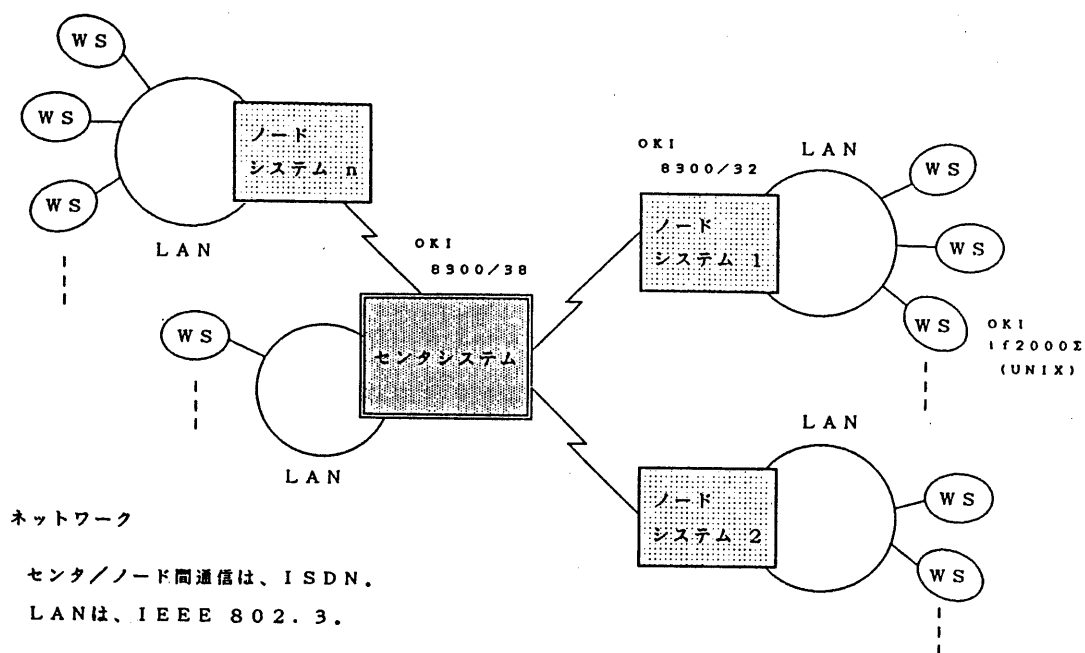


図1 ネットワーク構成図

3 ワークステーションOSの選定

マンマシンインターフェースに配慮したシステムを作るためには、ワークステーションOSの選定は重要である。

索引付与の作業は、標題や抄録を読みながらシソーラス等のツールを参照するなど、1画面上に複数のウィンドウを同時に開き、シソーラスウィンドウから最適語を索引する際など、ウィンドウ間のカットアンドペーストが自由に行える必要がある。

また、シソーラス検索を裏で行い、表では索引作業を継続できることが効率の点で望ましい。

こういった事情からOSとしては、マルチウィンドウ、マルチタスクが可能なUNIXを採用した。

4 文献の索引

4.1 文献データベースの入力項目

現在人手作業で使用している入力原稿用紙の概略を図2に示す。
主な入力項目は次の通りである。

- (1) 書誌事項
言語コード、ページ 等
- (2) 和文標題
- (3) 和文抄録
- (4) JICST分類コード
- (5) ディスクリプタ
シソーラス統制語
- (6) 準ディスクリプタ
自然語（統制なし）

なお、(5)と(6)をまとめる概念として、以後、索引語という言葉を用いる。

書誌事項 →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 33%;"></td> <td style="width: 33%;"></td> <td style="width: 33%;"></td> </tr> </table>			
和文標題 →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 33%; text-align: center;">.....</td> <td style="width: 33%; text-align: center;">.....</td> <td style="width: 33%; text-align: center;">.....</td> </tr> </table>
.....		
和文抄録 →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 33%; text-align: center;">.....</td> <td style="width: 33%; text-align: center;">.....</td> <td style="width: 33%; text-align: center;">.....</td> </tr> </table>
.....		
JICST 分類コード →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 33%;"></td> <td style="width: 33%;"></td> <td style="width: 33%;"></td> </tr> </table>			
ディスクリプタ →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%; text-align: center;">.....</td> <td style="width: 50%; text-align: center;">.....</td> </tr> </table>	
.....			
準ディスクリプタ →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%; text-align: center;">.....</td> <td style="width: 50%; text-align: center;">.....</td> </tr> </table>	
.....			
その他 →	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 100%;"></td> </tr> </table>			

図2 入力原稿用紙

4.2 索引語の付与

人手による索引語の付与では、標題、抄録文等をもとに、以下の手順で索引語の付与を行う。

- (1) 主題分析

(2) 索引語の選定

とりあえずシソーラス等を意識しないで主題分析に基づき自由に用語を選択する。

例： 選定された用語 A, B, C, D, E とする。

(3) 索引語の振り分け

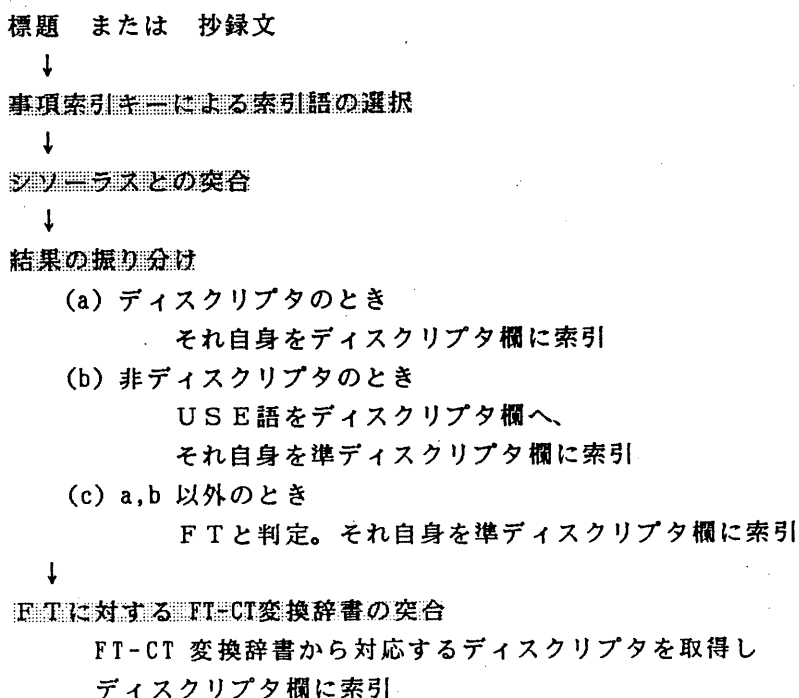
選定した語をディスクリプタとそれ以外とに振り分ける。

ディスクリプタ以外は、該当するディスクリプタに置き換える。

例： ディスクリプタ	A	説明：	AとBは選択された用語そのものがディスクリプタと一致した
	B		
	C'		
	D'		選定語Cに対しては、C'なる
	D''		ディスクリプタが対応した。
準ディスクリプタ	C		選定語Dに対しては、D', D''
	D		の2つのディスクリプタの組み
	E		合わせが必要であった。

これに対し、当システムでは、以下の基本処理仕様により(2)を実行すれば、(3)についてはシステムがある程度自動的に結果を導き出す。

索引付与作業時のワークステーション画面を図3に示す。



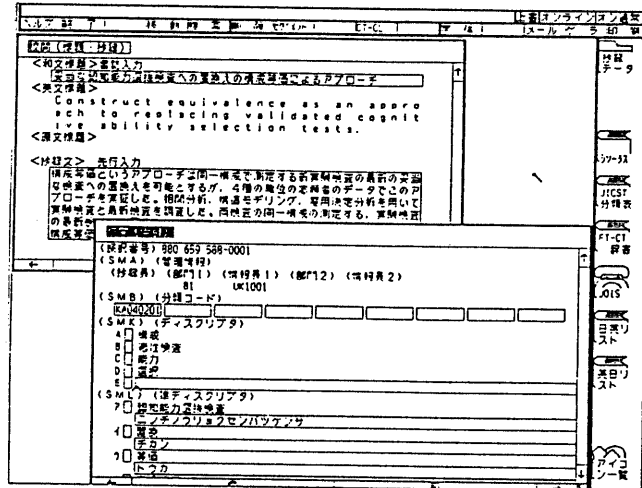


図3 索引付与作業時のワークステーション画面

4. 3 JICST分類コードの付与

索引語の付与を終えた後、再び中心主題を判定し、該当する分類コードを付与する。

当システムでは、各JICST分類に対しディスクリプタを予め付与し、CT-CLテーブルを形成している。したがって、中心主題に該当するディスクリプタを指示すれば、システムがテーブルをサーチして該当する分類コードが付与される。

5 データベースの検索

各ワークステーションにはシソーラスをはじめとして各種の辞書を搭載している。それらはデータベース化されていて、自由に検索することが可能である。

また、外部データベースとしてJOIS、JOIS-Fによってサービスされる各データベースを利用することができる。

5. 1 ワークステーション内のデータベース

以下の辞書をデータベースとして搭載している。

- (1) JICSTシソーラス
- (2) 日英対訳リスト
- (3) 英日対訳リスト
- (4) FT-CT変換辞書
- (5) JICST分類表

検索タグは、漢字タグ及びカナタグの両方可能。

サーチは、完全一致の他、前方、後方、中間一致が可能。

5. 2 外部データベースへのアクセス

索引作業中などに、JOIS、JOIS-Fにアクセスし検索することができる。

検索結果を索引に利用できる。

マルチタスクなので、JOIS と JOIS-Fとを同時に検索できる。

ウィンドウ間のカットアンドペースト機能により標題、抄録文中の文字列（カナモード）を検索タグに使用できる。

6 プライベートファイル

索引ノウハウ、索引履歴など個人的なメモファイルを作成し、必要時に参照することができる。

新しい概念をどのようなキーワードの組合せで索引したか、かつ、どのような分類コードを付与したかといった情報を書き留めておく。後に同一概念の文献が出現した際、このメモを参照すれば、新しい概念に対しても統一的な索引語の付与が行える。

7. ワークステーション校閲・索引作業の流れ

ワークステーションを使った校閲・索引作業も基本的には人手によるものと同様の流れになると考えられる。

流れ図を図4に示す。

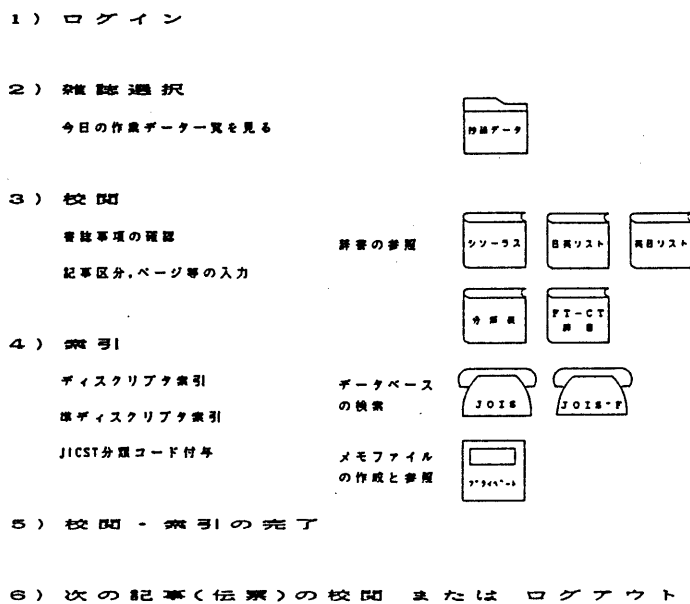


図4 ワークステーション校閲・索引作業の流れ

8. 課題

システムの技術的課題について述べる。

当システムのようなエキスパートシステムにとっては、マンマシンインターフェースへの配慮がシステムの評価に大きく影響する。従って、現バージョンでもその点に対しては十分配慮して開発したつもりである。しかし、今後現場での使用が進めば、ユーザから多くの注文が寄せられるものと思われる。

その他としては、プリンタ共用化のためのプリンタサーバシステムの開発、CD-ROMの接続、抄録作成支援機能の一つとしての英日機械翻訳システムの組み込み等が課題となっている。

9. 謝辞

本システムの開発に携わった方々、特に何事にも積極的に開発に取り組んでくださった沖電気工業（株）および関係者の皆様、また、有効な助言をお寄せくださった方々に対し心から感謝する次第である。

10. 参考文献

- 1) 高野文雄, 日夏健一: 索引支援システム, 情報管理, Vol.31, No9, 1988, pp.779-800
- 2) Martinez Clara, Lucey John, Linder Elliott. An Expert System for Machine-Aided Indexing. J.Chem.Inf.Comput.Sci. 27. pp.158-162 (1987)
- 3) Humphrey Susanne M. Indexing Aid Project: Research on Knowledge-Based Indexing of the Medical Literature. PB87-175790, 119p.
- 4) Humphrey Susanne M., Miller Nancy E. Knowledge-Based Indexing of the Medical Literature: The Indexing Aid Project. Journal of the American Society for Information Science. 38(3) pp.184-196(1987)
- 5) Vleduts-Stokolov Natasha. Concept Recognition in an Automatic Text-processing System for the Life Sciences. Journal of the American Society for Information Science, 38(4) pp.269-287 (1987)