

CD-ROMによる文献データベースの検索及び管理

2W-1

林俊文 寺田欣泰 石川佳治 植村俊亮

奈良先端科学技術大学院大学 情報科学研究科

1. はじめに

本論文はマルチメディア開発において、普及し続けるCD-ROMを用いた文献データベースの管理及び検索について述べる。本研究科は1995年3月から、毎年125人の修了生を出す予定で、それら修了生の修士論文を一人当たり約1MBとすると、1年間あたり、合計125MB程度の論文が提出されることになる。この膨大な量の論文をCD-ROMにより管理する方式について述べる。

2. 文献データベースの設計

2.1 設計方針

提供される環境及び文献データの性質を考慮し、設計の基本方針として、次の点を採用する。

(1) 修士論文を格納したCD-ROMを作成し、頒布する方式を考える。

(2) 文献データベースの作成は、基本的に奈良先端大で利用できる計算機資源を使用する。

(3) 文献データベースの検索などは、CD-ROMの可搬性という特徴を活かし、奈良先端大の計算機環境のみならず、個人所有のマシン間の互換性も重視する。

(4) CD-ROMだけ入手すれば、その利用ソフトウェアも含めていて、自由に活用できるものとする。すなわち、メディアの独立性に配慮する。

(5) 修士論文はTeXを使って作成される。CD-ROMには、TeXファイル及びPSファイルの両方を格納するものとする。

2.2 文献データの構造

学術文献データの一般的な構造を一意に決めるのは困難であるが、奈良先端大では、修士論文の形式が決まっているため、次のような項目により、特徴をつけることができる。

文献：[題名、指導教官、修了年度、著者、学校(研究科、専攻)、論文(本文(TeX、PS)(英文、または、和文)、概要(英文及び和文)、キーワード)、付録(学会発表原稿)]

前節の設計方針と特徴付けによって、文献データと本学特有の環境情報を文書情報と管理情報に分けられることができる。

- 文書情報：概要、本文(TeX、PS)
- 管理情報：学籍番号、著者、所属講座(指導教官)、題名、目次、キーワード

主に検索はTeXファイルを利用し、結果をPSファイルで表示する。TeXファイルのもう一つの重要な役目は、異なるマシンで、PSファイルを何かの障害で表示できない場合に、TeXファイルを、再コンパイルし、PSファイルを再生成することである。

3. 文献データベースの管理

3.1 概要

本システムでは、すべての管理情報及び文書情報をCD-ROMに収めて記憶する。CD-ROMは

シーク時間が遅い（約0.5秒）という弱点を持つので、検索速度を保つために、システムを利用するたびに管理情報を主記憶にロードし、検索は主記憶で行なう方法を採用する。

3.2 CD-ROM の記憶管理

このシステムのCD-ROMの記憶領域は、CD-ROMの識別子であるTOC、管理情報、検索プログラム、変換テーブルとデータ部から構成されている。

3.2.1 TOC

CD-ROMの場合は、情報が追記されるため、記憶順という一意に自動識別ができる識別子が存在する。それらの識別子のそれぞれの情報はTOC (table of contents)に記録されており、TOCにはディスクのそれぞれのトラックについての時間情報が記録されている。TOCの中のデータ構造は図1に示す。

トラック番号	インデックス番号	分	秒	フレーム番号	0	分	秒	フレーム番号
		経過時間				絶対時間		
← 8bit	← 8bit	← 8bit	← 8bit	← 8bit	← 8bit	← 8bit	← 8bit	← 8bit →

図1 TOCのデータ構造

3.2.2 変換テーブル

CD-ROMは最大99個のトラックしか使用できないため、一つのトラックに複数の修士論文を格納せねばならない。一つのトラックには最大99個のインデックスが利用できる。管理情報など人間が識別できるキー（学籍番号など）を用いて、複数の修士論文が格納されるトラックから、検索したい修士論文のインデックス（物理アドレス）を取り出すための変換テーブルを作成する。そのインデックスに関する時間情報はTOCに格納される。

CD-ROMの記憶領域にある情報の関連を図2に表示する。

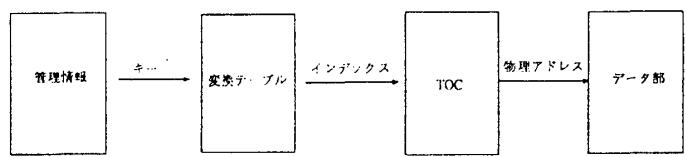


図2 CD-ROMの記憶管理

3.2.3 検索プログラム

桐やdBASEなどに代表される関係データベースアプリケーションによって、簡単に検索プログラムを作成することができるが、データベースのファイルの大きさで制限されることもあるし、使用するアプリケーションに依存してしまうこともあるので、独自の検索プログラムを作成する。

4. 課題

ほかに課題として以下のことがあげられる。

(1) 機種による検索プログラムの検討。

(2) CD-ROM上の最適なデータ配置。

参考文献

[1] 工藤礼子:コンパクトディスクによるマルチメディアデータベース, 東京農工大学大学院工学研究科修士論文, 1992年1月

[2] 小町裕史:電子出版技術入門, オーム社, 1994年4月

[3] Lorraine K.D:INVERTED SIGNATURE TREES AND TEXT SEARCHING ON CD-ROMS, Information Processing & Management Vol.25, No.2, pp. 161-169, 1989