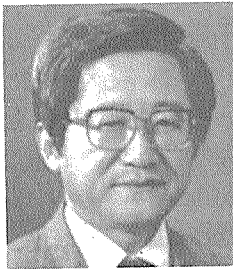


● 振興컬럼

포켓 데이터 베이스와 電子북



金 貞 欽
高麗大 教授 / 理博

8 cm CD-ROM의 威力

어제부터 筆者의 冊床 위에는 또 하나의 새로운 OA機器가 놓여있게 되었다. 크기 10.76 × 15.92 × 4.21cm, 무게가 550g 되는 손바닥 크기의 CD-ROM Player가 그것이다. 商品名으로는 Data Diskman, 8cm 直徑의 CD-ROM (Compact Disk-Read Only Memory)을 읽어내는 機械이다. 256 × 200 Dot의 液晶表示板(LCD)이 달려 있어 漢字가 섞인 글로 1회에 15字 × 10行씩 表示해 줄 수가 있다. 動力은 單3乾電池(1.5V) 6個 보통의 CD Player로서도 쓸 수 있어 8cm 直徑의 CD를 꽂으면 CD音樂도 가능하다.

그러나 이 기계의 目的은 音樂鑑賞에 있기보다는 CD-ROM에 內藏된 데이터(data)를 檢索해내는데 있다. 直徑이 不過 8cm이고 두께가 1.2mm밖에 안되는 얇고 가벼운(6.5g) 폴리카보네이트(플라스틱의 一種)製의 円盤인 이 CD-

ROM속에는 엄청난 分量의 情報가 들어 있다. 그 情報量은 200MB (2億Byte)이다. 英字로 라면 2億字, 國漢英 混合文이라면 1億字를 넣어 둘 수가 있다. (약 52萬字로 쓰여진 8萬大藏經이라면 2帙을 새겨둘 수가 있다).

우리나라에서 제일큰 百科大辭典인 東亞世界大百科辭典 30卷안에 들어 있는 글자數가 모두 약 9,000萬字이니, 이 얇고 작은 CD-ROM 한 장속에 30卷으로 된 東亞世界百科大辭典이 몽땅 들어가고도 10%정도의 餘裕空間이 남을 정도이다. 따라서 普通の 小說冊이라면 약 200~300卷 정도의 內容을 收錄해 둘 수가 있다. 그 收錄된 內容이 단추 몇개의 조작으로 LCD 表示板에 차례로 나타나는 것이다. 그러니 이것 하나만 있으면 大百科辭典 한 帙을 갖고 있는거나 마찬가지다.

한쪽 호주머니에 萬卷의 冊을 넣고다니는 時代

그뿐만 아니다. 그 CD-ROM은 얼마든지 다른 것과 바꾸어 질 수가 있게 되어있다. 그 CD-ROM은 摩耗나 損傷을 막기 위해 캐디(caddie)라는 5mm 높이의 얇은 알미늄 용기에 담겨져 있다. 따라서 그런 용기속에 넣어 保護되어 있는 상태라해도 그 크기는 9cm × 9.4cm × 0.5cm에 불과하다. 따라서 캐디에 들어있는 이런 CD-ROM 10개를 겹쳐놓아도 그 높이는 약 5cm가 되어 10장 全體의 크기는 9cm × 9.4cm × 5cm에 불과하다. 따라서 上衣의 한쪽 호주머니에 10장을 모두 넣어둘 수 있다. 그런데 그 안에는 理論上 약 10億字의 글을 담아둘 수가 있었던 것이다. 쉽게 말해 東亞百科大辭典 11帙(330卷) 또는 보통크기의 小說冊 2,000~3,000卷의 內容을 한쪽 호주머니에 넣고 다닐 수가 있다는 말이 된다. (약 52萬字로 구성된 8萬大藏經이라면 19帙에 해당). 얼마나 엄청난 것인가를 알 수가 있다.

더구나 캐디라는 保護막없이 보통의 CD 盤

을 쌓아올리듯이 CD-ROM을 알몸으로 5cm 높이로 쌓아올리면 약 41枚 (1.2mm×41=4.92cm)가 됨으로 收録할 수 있는 내용은 위 경우의 약 4배가 된다. 쉽게 말해 이제 우리는 한쪽 호주머니속에 1萬卷의 冊을 넣고 다닐 수 있다는 것이 된다. 兩쪽 호주머니라면 2萬卷의 冊이 된다. 웬만한 中小都市의 市立圖書館 藏書만한 内容이다.

갓가지 檢索소프트

더구나 데이터 디스크맨에는 5가지의 檢索소프트(software)가 IC化되어 內藏되어 있다. 單語檢索·條件檢索·메뉴檢索·複合檢索·參照檢索이 그것이다. 또 單語檢索은 다시 「前方一致」와 「後方一致」라는 두가지 方法으로 檢索하게 되어 있다. 前方一致란 「情報社會」·「情報産業」처럼 用語의 첫部分에 「情報」라는 共通接頭詞가 있는 말을 모두 찾아보고자 할 때 쓰는 檢索法이고, 「後方一致」는 「春秋時代」·「戰國時代」등 말의 뒷部分에 「時代」란 말이 붙는 用語를 찾아내고자 할 때 쓰는 檢索法이다.

예컨대 「後方一致」로 「時代」란 말을 入力시키고 단추를 누르면 表示板에는

古墳時代, 三國時代, 春秋戰國時代,

奏漢時代, 石器時代, 青銅器時代, …… 등 등 語尾가 「時代」로 끝나는 말들이 나타난다. 커서(cursor)를 찾고자 하는 用語, 예컨대 石器時代로 옮기고 確定단추를 누르면 石器時代의 說明이 나타난다.

前方一致 때도 같은 要領이다.

條件檢索이란 「in front of」란 英語의 意味를 알고자 할 때 「in」「front」「of」와 같이 3개의 關聯된 單語를 入力시키고, 이 모든 것을 包含하는 用語를 찾아내라할 때 쓰이는 檢索方法이다. 위에서는 「in」「front」「of」등 3條件을 썼지만, 最高 5가지까지 條件을 入力시킬 수가 있다.

메뉴檢索은 冊의 目次로 부터 該當되는 項目

을 찾아낼 때처럼 LCD表示板에 表示되어 있는 메뉴 중에서 찾고자 하는 項目을 찾아내는 檢索法이다.

其他 複合檢索·參照檢索이란 것도 있어 찾고자하는 用語自體를 잇고 있는 경우에도 위의 5가지 또는 6가지(前方 및 後方一致를 두가지로 取扱)의 檢索法을 쓰면 웬만한 用語나 項目은 모두 찾아낼 수 있도록 되어 있다.

電子冊의 時代

이리하여 이제 이미 世上은 電子冊(Electronic Book) 또는 電子出版(Electronic Publishing)의 時代로 접어 들어가고 있다.

一般的으로 「電子出版」이란 말은 두가지 意味로 使用되고 있다. 그 하나는 冊이나 雜誌를 만들 때 原稿作成에서 組版에 이르기까지의 모든 出版作業을 電子化하는 것인데, 이 경우 특히, 社內에서 또는 個人집에서 이런 作業을 할 수 있게 한 것을 DTP(Desk Top Publishing)라고 한다. 워크스테이션(WS, Work Station) 또는 워드프로세서(Word Processor)에 다 卓上 輪轉機 등을 結合시킴으로서 이루어진다.

두번째는 CD-ROM이나 LD(Laser Disk) 또는 IC카드 등 비디오 媒體를 통해서 冊의 内容을 出版하는 形態의 冊出版이다. 이미 美國서는 1985년부터 이런 意味의 電子出版이 活潑히 進行되어오고 있다. 그 하나가 CD-ROM으로 된 電子冊이다.

CD-ROM이란 音樂에서 使用되고 있는 CD(Compact Disk)盤에 音樂代身 文字나 그림을 入力시킨 것으로서 12cm 直徑의 것과 8cm 直徑의 것의 두가지가 있다. 두께는 모두가 1.2mm이고 무게는 12cm쪽이 약 15g, 8cm쪽은 약 6.5g에 不過하다.

이 얇고 작은 円盤속에 12cm의 것에는 540MB(Mega Byte), 8cm의 것에는 200MB의 情報를 記錄해 둘 수가 있다. 1B(Byte)는 英語文字

하나를 기록 또는 기억시키는 능력이며,漢字의 경우는 2B로 한글자를 기억시켜 둘 수가 있다.

따라서 12cm CD-ROM에는 國漢英 混用文으로 2億 7,000萬字, 8cm쪽에는 1億字가 기록(또는 기억)될 수 있다.

따라서 12cm CD-ROM 하나에는 웬만한 百科大辭典, 예컨대 Encyclopedia Britanica나, Encyclopedia Americana 또는 Qxford English Dictionary 全12卷이 몽땅 收錄될 수 있다. 그 뿐만 아니라 이 辭典의 어느 項目이라도 CD-ROM Player를 쓰면 單語의 경우라면 1秒 以內, 前方一致나 後方一致 등 여러 單語의 一覽을 檢索해낼 경우에도 數秒 以內에 檢索해낼 수가 있다.

데이터베이스로서의 CD-ROM

이미 美國이나 日本에서는 웬만한 百科大辭典이나 專門分野의 參考書類, 同窓生名簿, 電話番號簿, 出版書籍目錄 등이 CD-ROM의 形態로 出版되고 있다.

예컨대 筆者의 책꽂이에는 1988年度版의 世界 CD-ROM 總覽 Vol. 1, 2가 놓여 있는데, 그 冊字안에는 1988年 4月 現在까지 出版된 CD-ROM 電子冊의 一覽表와 그 內容紹介가 들어 있다.

또 美國이나 日本에서는 CD-ROM으로 된 出版書籍目錄이 있어 顧客이 願하는 冊을 即席에서 찾아준다.

예컨대 손님이 量子力學(Quantum Mechanics)에 關한 冊 중 그 著者의 이름이 D로 시작되는 것을 찾는다 하면, 書店員은 CD-ROM Player를 써서 條件檢索으로 손쉽게 그런 冊을 찾아줄 수 있다. 즉 먼저 條件檢索 단추를 누른 후 Quantum Mechanics란 말을 入力시킨다. 다음 條件으로서 著書名 D라고만 入力시키면 된다. 그러면 CD-ROM Player의 LCD 表示板에는

Robert H. Dicke and James P. Wittke :

Introduction to Quantum Mechanics

P. A. M. Dirac : The Principles of Quantum Mechanics

.....

.....

등 著者名이 D字로 시작되는 「Quantum Mechanics」라는 術語가 들어가 있는 冊의 一覽表를 表示해준다. 그래서 쉽게 찾고자하는 冊을 찾아낼 수가 있다. 만약 이 중 Dirac이 쓴 冊을 좀더 자세히 알아보려면 커소르(cursor)를 移動시켜 Dirac 冊쪽에 옮겨놓은 후 단추를 누르면 된다. 그러면 Dirac冊이 언제 몇版이 出版되었으며, 몇페이지이며, 그 內容은 어떠한 것인가 하는 概要, 그 內容의 水準(一般用, 中高校生用, 大學生用, 大學院生用, 專門用 등등)까지도 알려주며, 出版社의 이름과 住所와 電話番號도 알려준다.

따라서 書店員이라면 누구나 손쉽게 손님이 찾고자 하는 冊을 찾아줄 수가 있다. 설사 손님이 알고 있는 冊名이나 冊의 內容의 애매하고 斷片的인 경우라 할 지라도 條件檢索의 條件을 한가지 두가지 물어가면서 追加시켜 간다면 결국은 目的하는 바의 冊을 찾아 줄 수가 있게 된다.

이것은 뒤집어 얘기하면, 만약 이런 出版書籍一覽의 內容이 들어 있는 CD-ROM을 한장 갖고 있다면, 누구나 지금까지 出刊된 書籍에 대해 해박한 知識을 얻을 수도 있다는 말이 된다. 쉽게 말해 이런 CD-ROM으로 된 電子冊은 그 冊의 內容을 단순히 읽는다는 것외에, 갖가지 知識을 檢索하거나 統計를 내는데도 쓰일 수가 있다는 것이된다. 쉽게 말해 電子冊은 그 自體가 하나의 데이터 베이스(Data Base) 였던 것이다.

出版界의 흐름을 바꾸게 될 電子冊

1985年 美國의 첫 出版을 한 12cm 直徑의 CD-ROM은 容量은 2億 7,000萬字(英文 5億4,000萬字)나 되어 매우 컷지만, 그것을 읽어내주는

CD-ROM Player가 컷고(Desk Top PC水準), 또 그 값도 비싸서 一般大衆이 구입하기에는 부담이 컸었다.

그러나 이제 Data Diskman과 같이 작고(약 10.8cm×15.9cm×4.2cm), 가볍고(550g), 조작하기 쉽고, 싼(약 30萬원水準) CD-ROM Player가 나와 電子冊은 단숨에 그普及에 박차가 가해질 것 같다.

물론 現在의 Diskman에는 改良할 點도 많지만(예컨대 表示文字數 15字×10行의 擴張같은 것), 그것은 時間이 해결해 줄 것이다. 예컨대 몇年内로 손쉽게 갖고 다닐 수 있는 보통冊 크기의 表示板을 갖는 8cm型 CD-ROM Player가 나와 圖書館의 役割을 一部 代替해 줄 것 같다.

예컨대 그 때가 되면 「어거서 크리스티(Agatha Christie)의 모든 것」 또는 「어거서 크리스티 推理小說全集」이라는 이름의 8cm版 CD-ROM冊이 나와 讀書界를 휩쓸지도 모른다. 이 CD-ROM 한장만 있으면 어거서 크리스티의 推理小說은 長篇이건 短篇이건 모두 읽을 수 있으니 말이다. 물론 表示板이 너무 작다고 생각하는 사람은 안房에 있는 30인치 또는 33인치 TV에 비추어 볼 수도 있을 것이다.

그러니 이제 讀書界 또는 出版界는 電子冊(電子북 또는 E북이라 略稱)의 出現으로 그 흐름이나 性格마저도 달라질 것 같다.

電子북이나 電子出版 開發에 힘쓸 때가 오고 있다.

물론 電子冊이 나왔다고 해서 出版界가 完全히 電子북으로 바뀐다는 것은 아니다. 왜냐하면 지금까지의 冊에는 그 나름대로의 좋은 點이 많기 때문이다. 우선 푹푹말아서 갖고 다닐 수도 있고, 感銘깊게 읽은 부문에 줄을 치거나 欄外에 感想이나 주석을 달 수도 있고, 아주 重要하다면 그 부분만 복사하거나 또는 아예 찢어서 포켓에 넣고 다닐 수도 있다. 또 한권 두권의

內容이라면 在來式의 冊쪽이 훨씬 값도 싸다.

그러나 電子冊은 電子冊대로 長點을 갖는다. 6.5g의 直徑 8cm CD-ROM은 간단히 포켓에 넣고 다닐 수가 있다. 그런데 그 안에는 보통의 冊으로 따져 200~300卷정도의 內容이 들어 있으니 冊 한권의 單價로 따지면 무척 싸게 된다. 또 단순히 읽는 것 뿐만 아니라 檢索法에 依해 각종 關聯知識도 얻을 수 있다. 特히 辭典類나 名簿類로서는 더이상 이보다 便利한 것은 없을 것이다.

어쨌든 이런 理由로 앞날(아마도 '90年代後半부터)의 出版界는 많이 그 樣相이 달라질 것이 예상된다.

따라서 우리가 「電子出版」이란 것을 단순히 自動組版정도의 뜻으로 해석할 것이 아니라 CD-ROM으로 된 電子북을 出版하는 것이 오히려 더 主된 것이라 해석하고, 電子북 出版에 힘을 기울여야 할 줄로 안다.

筆者가 '89年 12월에 열린 臺灣의 「情報의 달」 行事를 參觀했을 때 놀란 것은 臺灣에서는 이미 國語·英語 教材用의 CD-ROM 셋을 몇10個씩이나 商品化해 놓고 있다는 點이다. 世界 6位의 電子產業國이라 우리들 스스로는 자랑하고 있지만, 컴퓨터나 그 소프트웨어 또는 그 周邊分野에서는 오히려 臺灣이 우리를 앞서고 있는 것이 實情이다. 물론 臺灣을 뒤따라가는 것이 重要한 것은 아니지만, 그렇다고 臺灣도 이미 시작하고 있는 電子북을 우리는 이제야 겨우 시작('90年 가을경 國內서 첫 試圖)하고 있다.

電子북이나 電子북을 읽게 해주는 CD-ROM Player는 어쩌면 PC以上으로 좋은 輸出商品이 될 可能性도 많다. 그 뿐만 아니라 國內用으로도 CD-ROM은 辭典類나 各種 統計資料·名簿(顧客管理包含)·出版書籍目錄 등과 같은 基礎 데이터 베이스(Data Base)用으로도 그 需要가 많을 것이 예상된다.

어쨌든 다가올 情報化社會를 위해 우리는 電子出版·電子북의 開發에도 힘을 기울여야 할 줄로 안다.