

CD-ROMによる電子縮刷版システム

—読売新聞社の事例と今後の展開—

Electronic Pocket Edition Based on CD-ROM

設 楽 淳 二*

Junji Shitara

北 澤 行 司***

Yukiji Kitazawa

武留井隆夫**

Takao Murui

芹 田 寿 樹****

Toshiki Serita



電子縮刷版システムの構成 電子縮刷版システムは、新聞紙面をそのままCD-ROMに収納していることにより、キーワードで検索した記事のパソコン画面への表示と印刷ができる。

電子縮刷版システムは、世界で初めて新聞紙面をイメージデータとしてCD-ROMに書き込み、さらに、記事検索のための検索ソフトとキーワードを付加したマルチメディアシステムである。発行元である読売新聞社から提供される紙面に検索キーを付与した電子縮刷版化することにより、パソコンによる写真、図表などを含めたままの新聞記事検索が可能となる。このシステムは検索性に優れた記事テキスト版のCD-ROMと、一覧性に優れた紙の縮刷版の長所を融合させた画期的なものと言える。

電子縮刷版では、1か月分の紙面イメージデータ

を月の前半、後半に分け2枚のCD-ROMに書き込み、キーワードおよび記事見出しはそれぞれのCD-ROMに1か月分を収録、どちらからでも1か月の記事検索を可能にしている。平成6年1月号を2月15日に発行、その後は毎月2枚ずつを月次発行している。平成7年初めには紙面イメージデータを書き込まずに、1年分の記事見出しとキーワードだけを集成した「平成6年度索引」のCD-ROM1枚を発行し、合計25枚のセットで1年分の紙面イメージデータ、記事見出しおよびキーワードを提供する。

*読売新聞社 編集局情報調査部 **日立製作所 システム事業部

****日立製作所 情報メディア事業部

***日立製作所 流通システム営業本部

1 はじめに

電子縮刷版システムは、オンライン記事データベースや最近各新聞社で出版されている紙面のテキスト部分だけをCD-ROM化した出版物などでは扱えなかった新聞紙面そのものを、CD-ROMに収納した新しいマルチメディアシステムである。

電子縮刷版のCD-ROMにはキーワードと検索ソフトが組み込まれているため、従来の紙の縮刷版では不可能であったパソコンによる容易な検索が実現できた。

電子縮刷版システムは読売新聞社と日立製作所が共同開発し、公共図書館を主な導入先として平成6年1月号から発行を開始している。

ここでは、テキスト版CD-ROMと冊子体の縮刷版双方の長所を融合させた電子縮刷版システムの特長とその機能、概要について述べる。

2 電子縮刷版システムの特長

電子縮刷版システムでは、31ページの写真に示すようにパソコンの画面上で、あるいは紙面のプリントの際に、写真や図表を含んだ紙面イメージそのものが得られるので、おのおののニュースの扱いが一目でわかる。従来出版されているテキストデータだけのオンラインデータベース、テキストCD-ROMでは、イメージ情報が欠落しているのが新聞の最大の特長の一つである一覧性、つまりぱっと見てニュースの重要度がすぐにわかるという利点が失われている。この電子縮刷版では、一覧性を失うことなしにオンラインデータベースと同等の、容易で確実な検索機能を実現している。

記事検索のためのキーワードは、オンラインデータベースとして実績のある読売新聞社の“YOMIDAS”から抽出し、それぞれの記事に付与している。漢字とふりがなをセットにしたキーワード総数は月当たり約100万語あるので、検索者自身が思い付いたことばでの自由な検索で必要な記事を確実に参照できる。掲載日があいまいな記事を探す場合にも、その記事に関連しそうなキーワードを入力することにより、求める記事を短時間で確実に検索できる。さらに、キーワードを集約した年間索引CD-ROMを利用すれば、1年間分、約8万本もの記事をまとめて効率よく検索できる。

紙面イメージのプリントに関しては、A4の用紙に紙面イメージそのものを専用レーザビームプリンタで1枚当たり8秒の高速で印刷できる。高い印字品質を保った



図1 紙面プリントアウト

紙面的一部分を切り出し、A4サイズに横長で印刷することもできる。この場合、クレジットとして発行社名と日付が印刷される。

まま紙面1ページ分を縮刷版と同じサイズで縮小印刷したり、または特定の記事領域の部分を縦長や横長に元の新聞の文字と同一の大きさで、あるいは拡大した大きさに自由に印刷できる(図1参照)。紙面の部分印刷をした場合には、クレジットを兼ねて新聞社名と掲載日、朝・夕刊別、ページ数が用紙の右下に印刷される。

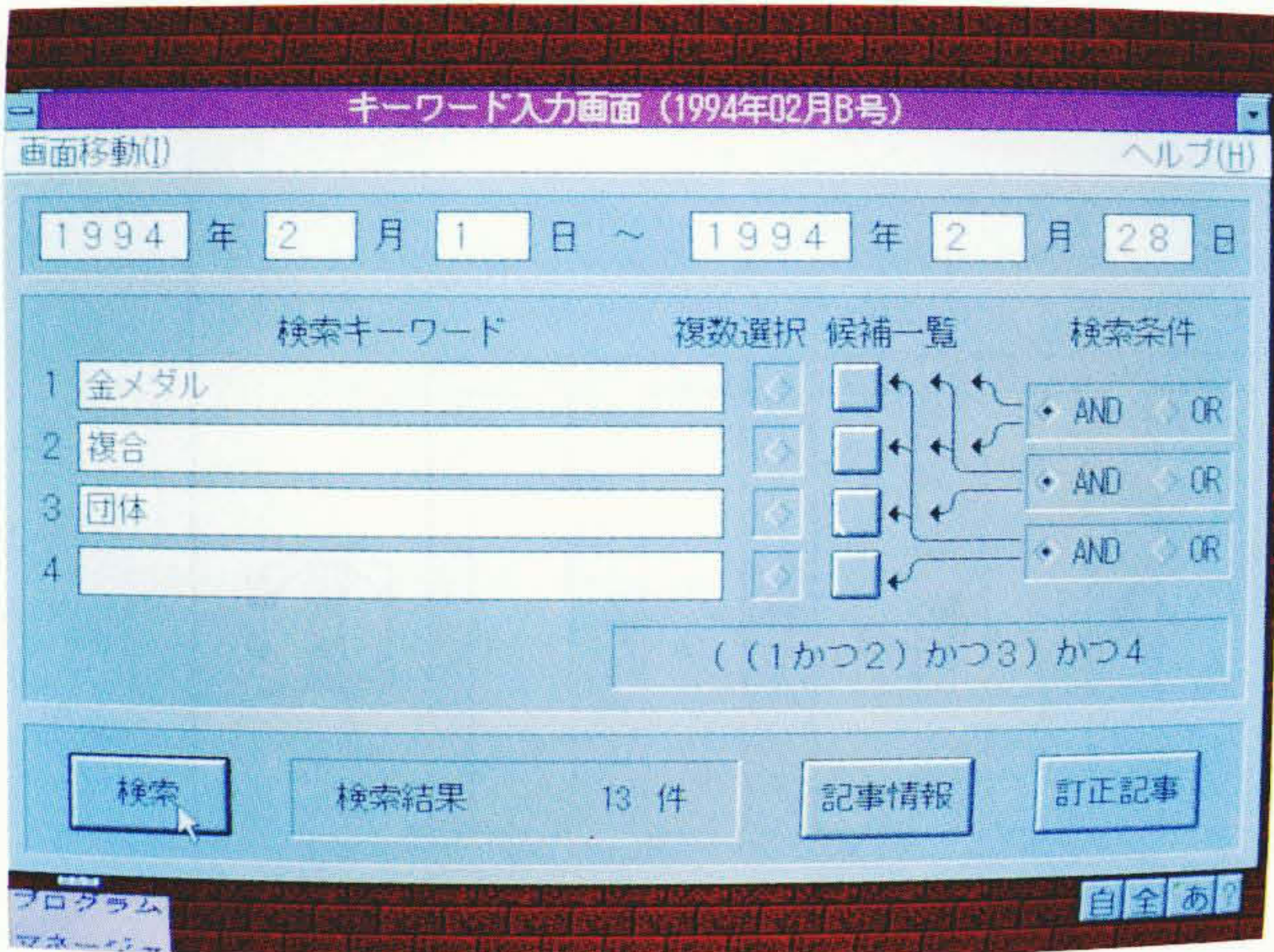
情報提供のスピードという面からでは、テキストデータだけのニュース記事CD-ROMが1年経過した後に1年間分のテキストデータをまとめて提供するのに対し、電子縮刷版は記事が紙面に掲載された翌月の中ごろには発行されるので、よりスピーディーな情報提供が実現できる。

電子縮刷版の媒体は、音楽用のCDとまったく同じものである。キャディ(収納ケース)を含めた1か月分2枚のボリュームとしても、従来の大きな縮刷版と比べて重さで $\frac{1}{14}$ 、体積で $\frac{1}{10}$ となるので、取り扱いも容易となるばかりでなく、保管スペースの面でも大きなメリットが得られる。

3 検索機能とイメージROM

電子縮刷版のキーワードは漢字、ふりがなおよび同義語だけでなく、「トラック」や「バス」に対する「自動車」などの上位語がそれぞれの記事に付与されており、以下に述べるような操作方法で検索者自身が思い付いたことばで、かつ漢字、仮名のどちらでも必要な記事の検索が容易に確実にできる。

検索画面は初期画面、キーワード入力画面、見出し一覧画面、および紙面イメージ表示画面の4画面で構成する。初期画面では、キーワード検索か掲載日・ページ数



使いやすい検索画面

図2 キーワードによる記事検索
キーワードは最大4件まで入力が可能であり、論理演算もできる。

検索のいずれかの方法が選択できる。キーワード検索を選ぶと、図2に示すようなキーワード入力画面が表示され、キーワードを入力し検索を実行すると、そのキーワードでヒットした記事の件数が画面下部に表示される。ヒットした件数が多過ぎて検索の絞り込みを行いたい場合には、最大4件までのキーワード間での論理演算としてのANDなどが行えるので、適当な記事の件数に絞り込みができる。

検索の際にキーワード自体を正確に記憶していない場合でも、キーワードとして入力したことばと前方一致する隣接語表示機能がある。それによって適切なキーワードを選択できるので、電子縮刷版システムを利用する場合には専門家としての特別な検索技術がいっさい不要であり、初心者でも簡単な説明を受けただけで、マニュアルなしで必要な記事の検索を行うことができる。

画面の検索ボタンを押して見出し一覧画面を開くと、ヒットしたすべての記事の掲載日、朝・夕刊の別、掲載面数、記事段数、写真の有無などを見出しのテキスト一覧が表示される。これらの記事の見出し一覧の中から見たい記事を選択すると反転表示され、さらにマウスボタンをクリックすると、記事を含む新聞イメージが画面に表示される。画面に表示されたイメージは拡大、縮小、スクロールが自由に行え、必要な記事の部分を画面でそのまま読むことができる。

ある記事の関連記事あるいは続報記事を検索する場合には、紙面に付与されている面名によるページ検索機能があるので、改めてキーワードを入力しなくても翌日や前日の同じ面を開くことができる。また、見出し一覧の

画面で参照したい紙面を記録し、のちほどその面を順に参照したり印刷できる「しおり機能」、あるいはあたかも紙の縮刷版をパラパラとめくってゆくような操作ができる「自動改ページ機能」もある。なお掲載日、ページ数がはっきりと特定できる記事の場合には、キーワードをいっさい使用せず初期画面からそのまま掲載日、朝・夕刊の別、ページ数を入力して紙面を表示する機能もある。

電子縮刷版システムは、汎(はん)用的なソフトウェアであるMS-Windows3.1[※]環境のパソコンにSCSI (Small Computer System Interface)でイメージROMドライブを接続するだけで利用できる。CD-ROMに組み込んだ検索ソフトのドライバソフトはFD(Floppy Disk)で提供する。レーザビームプリンタは、A4用紙1枚当たり8秒の高速で印刷するため、専用インタフェースでイメージROMに直結している。イメージROMの仕様を表1に示す。

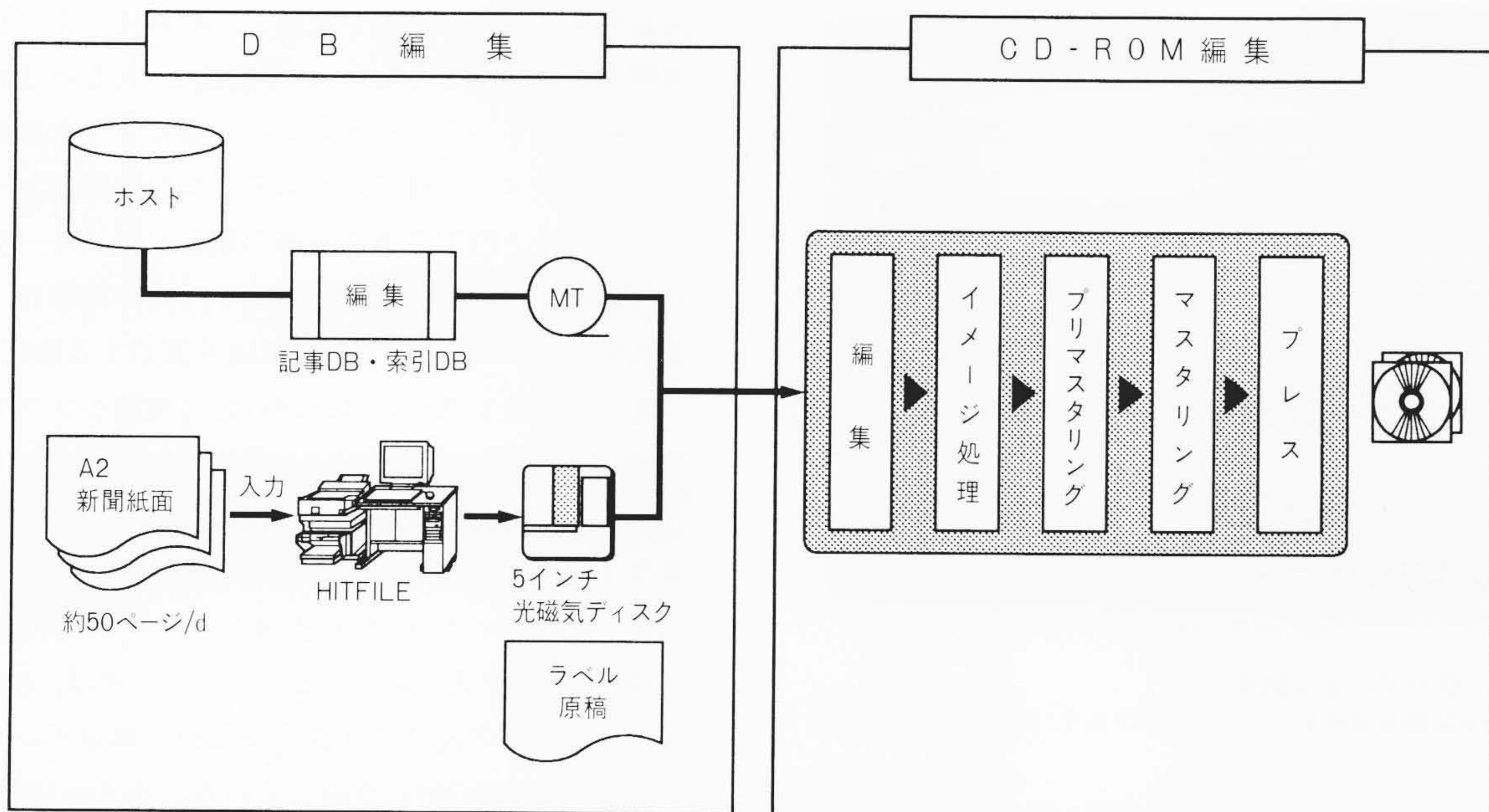
イメージROMは、日立製作所が特許公報システムなどのイメージデータ処理用に開発したものである。紙面イメージデータを高速にディスプレイに表示したり、プリンタで印字するためのハードウェアによるデータ伸張機能を搭載しているので、圧縮して書き込まれている紙面イメージデータの高速の伸張や画面上でのすばやいスク

※) MS-Windows3.1は、米国マイクロソフト社の登録商標である。

表1 イメージROMの仕様
イメージ処理機能を搭載しており、データの圧縮・伸張、拡大・縮小、回転などが高速に行える。

機 種	CDR-5150
タイプ	スタンドアロンタイプ(水平置)
インタフェース	SCSI(ミニコネクタ)×2 日立バス(増設CD-ROMドライブ接続用)×1 ビデオI/F(レーザビームプリンタ接続用)×1
レーザビームプリンタ出力能力	●A4サイズ200ドット/インチ・400ドット/インチの解像度対応 ●印刷速度:最大8枚/min(連続時)
イメージ処理機能	●イメージデータ圧縮・伸長 ●任意縮小・拡大・回転(左右90°) ●キャラクタコードの文字フォント展開 ●イメージデータのイメージバッファ内複写機能
対応ディスク	CD-ROM規格
環境条件	温度 5~35℃
	湿度 15~80%RH
外形寸法	幅360×奥行360×高さ65(mm)
電源	AC100V±10% 50/60Hz
消費電力	10W(最大25W)
質量	約5kg

注:略語説明 I/F(インタフェース)



注：略語説明 DB (Database), MT (磁気テープ)

図3 電子縮刷版の作成

記事DB・索引DBをホストで処理し、検索に必要なキーワードを磁気テープに変換する。イメージデータは光ディスクファイルシステム“HITFILE”から入力する。

ロールができる。

4 CD-ROMの製作

電子縮刷版システムでは、紙面イメージデータおよび優れた検索性を実現するための記事見出しとキーワードをCD-ROMに書き込まなければならない。まず紙面イメージデータに関しては、データ品質の面からは紙面編集システムであるCTS(Computerized Typesetting System)の出力部分からデータを取り出してCD-ROMに書き込む方法が考えられる。しかし、この方式ではCTSの出力データを電子縮刷版システムのデータフォーマットに合わせるための専用プロセッサと、データ変換のためのソフトウェアが必要となり、開発コストおよび期間の面で実用的ではない。このため、紙に印刷された紙面そのものを白黒の電子縮刷版システムのデータとして電子化する入力方式を採用した。

紙面イメージデータの圧縮についても、1か月分の紙面イメージデータを一定の品質を保って圧縮し、実用的な枚数の電子縮刷版のCD-ROMに収録して発行しなければならない。検討の結果、A2サイズのスキャナで紙面をそのまま200ドット/インチの解像度で入力し、データ圧縮を行えばCD-ROM2枚に1か月分の紙面イメージデータ、および検索用のキーワードを収録できることがわかった。このため、入力装置としては汎用の光ディ

スクファイルシステム“HITFILE”を使用し、普通の新聞紙面を200ドット/インチの解像度で入力することとした。電子縮刷版の製作工程を図3に示す。

テキストデータである記事見出しとキーワードの作成については、“YOMIDAS”を有効活用して使い勝手の良い検索システムを短期間で開発した。すなわち、大型ホストコンピュータで電子縮刷版システムに合わせた形で記事見出しとキーワードデータを編集・加工するソフトウェアを開発した。そして、読売新聞社で毎月初めに記事見出しとキーワードデータをそのソフトウェアによって編集・加工し、磁気テープに出力したものを作成し、それをCD-ROMに書き込むこととした。この方式により、記事の掲載の翌月半ばには紙面イメージデータをそのまま参照でき、かつ検索性に優れた電子縮刷版を発行することができる。

5 おわりに

ここでは、新聞紙面イメージと検索キーワードをCD-ROMに一体化した電子縮刷版システムについて述べた。このシステムは、検索性と一覧性に優れた画期的なものである。将来的にはこの電子縮刷版をネットワーク上のサーバにセットしておくことにより、自分の席に居ながらにしてパソコンで自由自在に記事の検索ができるようになり、オフィスの生産性の向上に役立つと確信している。