

## 文書管理システムの現状と将来†

### — 新しい技術の業界への応用—

阿部 詔一 ††

従来清書を補助してくれる機器として普及してきたワードプロセッサや、印刷用版下を簡易的に構築するためのシステムである DTP が、急速にデータベース構築のためのエディターとして扱われるようになり、米国で約 10 年前に起こったデータベースパブリッシングの概念が日本に於てもやっと一般化されつつあります。一方では、光ファイリングのような大容量媒体を利用したシステムも、急速にソフトの強化とオープン化を進めており、最近では、もはや専用機としての販売は非常に難しくなっているであろうことが察せられます。

このようなデータベース重視の方向は、従来日本の各ワープロメーカー、コンピュータメーカーが共通に取ってきたファイル形式やソフトウェアの API の非互換化による自社市場の維持という考え方を根底から覆すものであり、唯一と言っていいほど日本メーカーが自国の市場の中で優位性を保って来たといえる日本語の処理機能が、オープン化によって崩壊して行く可能性が非常に高まったことを意味します。

この動きを契機に、ソフトウェアやファイルの互換化、標準化は益々進んで行く事になり、更にその標準化の流れのほとんどが日本ではなく、米国から来ることになるでしょう。従ってこの流れは、多国籍間の文書共有の標準化を促し、機能性よりも互換性を求めるプロダクトが重要視されて行く傾向が、益々高まって行くこととなりそうです。

さて、ファイルの標準化の流れに関しては、米国防総省が中心となって 1984 年から開始した CALS 計画に基づくファイル形式が急速に世界的な標準化を生みつつありますが、一昨年あたりより、やはり米国の軍事通信インフラを用いたインターネットの普及が日本に於ても活発化しており、デファクトスタンダードとしてのファイル形式がより負荷価値を付け標準化の方向を歩んでおります。当社でもその双方をにらみ、SGML、HTML、WWW、JPEG、MPEG 等に対応した文書管理システムを構築致しておりますが、最近の情報系のオープンシステムの願

望ニーズとして、従来のクライアントサーバーシステムの中心であった SQL 型 RDB と Windows4GL の組合せを凌ぐ勢いでこれら標準化を求める動きが、特に大学や研究機関に於いて活性化してきたことが感じ取れます。

どのようなファイル形式を取ったとしても、現在の紙という便利な記憶媒体を離れて行くためには、情報を管理するためのいろいろな要素技術が必要となってきます。図形の表現及び蓄積のためには CAD 化することが必要ですし、効率のよいデータ交換のためにはベクター、ラスター両方の CAD 化及びその双方の変換技術が必要となって参ります。更に、どのような部材情報でも相互に共有、管理できる SQL 型の RDB が必要です。テキストに関しては、統制化されたキー情報だけではなく、記載された全情報を検索可能とさせるような高速の全文検索エンジンと、表記の揺れを吸収するような辞書の存在が必要となって参ります。

また情報のマルチメディア化は急速な情報量の拡大を引き起こし、媒体や回線の容量不足を補い、更に高速化を図るニーズも加わって、データ圧縮技術が駆使されることとなります。

全文検索エンジンは、現在 OSI の標準化の流れの中で、SQL の標準規定の中で採用が検討されておりますが、言語によって高速検索のインデックス化の方法が異なるため、標準化は非常に難しいと思われれます。当社でも独自のエンジンを開発致しておりますが、他社に例を見ない特長として、OS にダイバインドしないという大きな利点があります。といっても、あらゆる OS という訳ではなく、現在オープンシステム市場の中で一般化している OS (UNIX、MS-Windows、Netware、OS/2、DOS、NT) に対し、共通 API にて使用可能なライブラリとなっております。

また他社の商品が海外システム (米国、カナダ、オーストラリア、ドイツ等) の移植商品が多い中で、言語学技術を用いて独自に開発した辞書を加え、漏れの少ない高速検索技術を確認しているところが大きなポイントと言えます。

全文検索の機能を駆使する上で、絶対的に起こりうる問題点が表記の揺れです。全文検索に限らず、文書管理の

† Shoichi Abe. 本稿は、日本原子力学会 1994 秋の大会「放射性廃棄物部会企画セッション」での講演要旨である。

†† エム・シー・ワードセンター、システム事業部 System Business Division, MC WORD CENTER, Inc.

普及に伴い情報系のデータベースが広がれば広がるほどそこに発生する膨大なキーワードの統制も不可能になって参ります。これらの問題を解決するのが、当社が国内では唯一商品化した検索を目的とした約 20 万語の汎用シソーラス辞書と、約 63 万語に及ぶ専門用語シソーラス辞書です。

元々日本語は表記言語と表音言語が混在し、外来語の数も圧倒的に多い 2 バイトコード言語です。全文検索システムメーカーの多くは、閾値と称する曖昧度設定によってアルゴリズムで曖昧性をカバーする方法を取っておりますが、この方法は英語のような表音言語には例えば英文ワープロのスペルチェッカーの様な有効性を発揮致しますが、表記言語にこれを使うと、全く意味の違うものまで捕らえてしまう場合が多々あり、検索ノイズが多く発生するため、使用に耐えなくなる恐れがあります。従って、今後もこれら概念体系を示す辞書は益々必要になって来ると思われ、当社では既に本学会関係者が最も必要とされるであろう電力、原子力、放射性廃棄物用語に関する同義概念辞書化も商品化を完了致し、メンテナンス体制も確立致しております。

今後は米国に於いても、更に深く広いデータベースへの発展が指向されるため、現在はまだ独自性を持つドキュメントフローの標準化や、各データや書式、ソフトモジュール等のオブジェクト化や更にその標準化が益々進むこととなるでしょう。

何れにせよ、当放射性廃棄物業界関連文書に於いても、その保存を必要とする期間の長さや、関与する個人、組織の幅の広さから、ユーザー、監督官庁、業界設備機器メーカー、コンピュータ業界、関連印刷会社、業界出版社などを巻き込んで、電子化やペーパーレス化が益々広がって行くことは確実です。

大きな設備投資や、ソフトウェア投資は今の時点ではすぐに必要ではありません。まずは各位が標準化の方向性をよく見極め、現在の文書管理の方法や、データの持ち方を統一化し、来るべき将来への布石を打って置くことが重要だと考えます。