

S G M L 方式による
情報管理誌全文データベース化の可能性と
H T M L による
電子版情報管理誌の試作

森田 歌子*、鈴木 政彦**、宮川 謹至***、浜中 寿***

日本科学技術情報センター 業務部*、技術開発部**、電子計算機システムセンター***

情報管理編集委員会では、情報管理誌の効率的編集と、情報管理誌のより高度な利用を指向するため、情報管理誌の SGML 化について検討を行ってきた。この検討には、JICST 電子計算機システムセンターの次期システム検討グループによる JICST DTD の検討と、編集委員による電子版情報管理誌の試作という、EDP 部門によるアプローチと実際の編集という立場からのアプローチの両者が相互に意見交換を行いつつ進められた。これらの検討は、まだ各々の立場からの検討が、先行しているきらいはあるものの、今後情報管理誌が学協会誌等と連携して行く上での、その在り方を検討する第一歩である。

A TRIAL FOR DEVELOPMENT OF DTD
FOR "JOHO KANRI"
AND
"JOHO KANRI" FULL TEXT DATABASE
BY USING HTML

UTAKO Morita*,MASAHIKO Suzuki**,KINJI Miyagawa***,HISASI Hamanaka***

Japan Information Center of
Science and Technology
Editors Office for JOHO KANRI*
Technology Research and Development Department**
Information System Center***
5-3 Yonban-cho Chiyoda-ku,Tokyo 102 JAPAN

This paper describes a trial for development of DTD for "JOHO KANRI" and "JOHO KANRI" FULL TEXT DATABASE by using HTML. "JOHO KANRI" is a special journal for information processing and it has wider than other journals in its paper style and category. So, it was difficult to make SGML style and DTD for "JOHO KANRI". Here it was made Hyper-Linked "JOHO KANRI" by using HTML and distributed by CD-ROM in order to evaluate, and so on.

1. 情報管理誌の構造とDTD

情報管理誌のSGML化を検討するにあたり、現在行っている編集プロセス等についての先入観を全く捨てて、全く別な視点で情報管理誌を見てみることにした。

そこで、まず最初に、既に発行済みの1年分を取り上げ、収録されている記事の構造の分析を行った(図1)。情報管理誌の記事は、論文、講演、連載、ディスカッション、ニュース、図書紹介、アナウンス、文献の紹介の、8種に分類できた。これは、情報管理誌自体が、情報処理関連の技術等の論文をまとめた雑誌としての性格から、比較的容易に推測される。中でも、本誌の大部分を占める論文は、一般論文、講座形式の論文、講義の論文などがあり、比較的構造がはっきりしている。その他のものについては、やや形式に統一性が少なく、個別に構造をまとめる必要がある。

論文	=	フロントタイトル, 著者の詳細+, アナウンス+, 出版情報, 本文, 後付け?, 参考文献?, 謝辞?
講演	=	フロントタイトル, 著者の詳細, 講演, , 参考文献?
連載	=	フロントタイトル, 著者の詳細, 本文, , 参考文献?
ディスカッション	=	フロントタイトル, 著者の詳細+, ディスカッション
ニュース	=	ニュース?
図書紹介	=	フロントタイトル, 著者の詳細, 出版情報, 図書紹介?, , レポート?
アナウンス	=	アナウンス
文献の紹介	=	フロントタイトル, 著者の詳細?, 内容紹介

図 1

次に、情報管理誌のDTD定義を試みた。情報管理誌のDTDを検討するために、まず、JICSTとして従来から利用してきている書籍・雑誌・論文という観点からのDTDの検討を行った。JICSTでは、入手した資料については、全て資料管理システムで資料の入手からJICST書誌データベース作成、JICST文献データベース作成、JICST英文データベース作成、そして閲覧・複写対応まで一元管理している。それに合わせて、BOOK DTD、雑誌DTD、基本DTDの3種を設定し、各々にエンティティを対応させることを試みた(図2)。このエンティティ

には、基本DTD構造構成、構成要素、要素、数式、化学式等が考えられる。これらを、情報管理誌の分析により得られたDTDに対応つけてみた(図3)。雑誌DTDとしては、図4の様なものが考えられる。また、これらが参照するエンティティとしては、図5の様なものが、また構成要素としては、図6の様なものとなる。例として、2次情報DTDならびに基本ドキュメントDTDを示す(図7、8)。

以上のような検討を基に、現在情報管理誌DTDの定義を進めている。情報管理誌の場合、他の学術雑誌に比べ、ややカバーする範囲が広い。即ち、論文形式以外のものまで含まれるため、更に多くの発行例を分析し、DTDの定義を進めて行く必要がある。

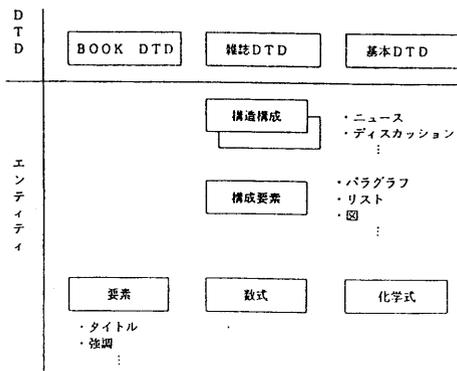


図 2

JICST DTD TREE構造

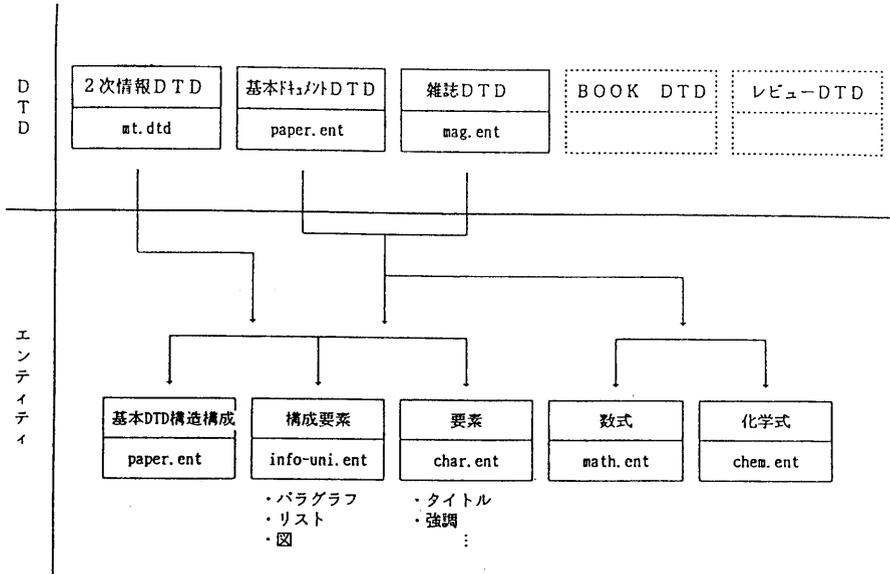


図 3

mag.dtd (雑誌 DTD)

```

mag [ -- ]
  NChar.ent :
  NInfo-uni.ent :
  Npaper.ent :
  cover+ [-O]
    | seritle --- PCDATA
    |
    | date+ [-O]
    |   | pub-yr [-O] --- PCDATA
    |   | pub-mo? [-O] --- PCDATA
    |   | pub-day? [-O] --- PCDATA
    |   |
    |   | vol-no+ [-O] --- PCDATA
    |   | no+ [-O] --- PCDATA
    |   | orgname+ [-O] --- PCDATA
    |   |
    | front? [-O]
    |   | t1+
    |   | p+
    |   | h1+
    |   | contents
    |   | h1+
    |
    | paper+ [-O]
    | back? [-O]
    |   | t [-O] --- (Nfulltxt:)
    |   | (Minfo-uni:)
    |   | topic* [-O] --- (Nfulltxt:)
    |   | pub-corp+ [-O] --- (Nfulltxt:)
    |   |
    |   | pub-addr+ [-O] --- (Nfulltxt:)
    |   |
    |   | price* [-O] --- PCDATA
  
```

図 4

参照エンティティの説明

エンティティ名称	エンティティ (3-1/1/1)	内 容
基本要素	%Char.ent	構成、構成をしない要素をまとめたもの。段落の要素と文字列を表現する場合に使用する。
構成要素	%Info-uni.ent	構成情報の要素をまとめたもの。構成要素とは、段落、強調などの単純な構成を表現するものである。
論文定義	%Paper.ent	論文の構成を表現する要素をまとめたもの。表紙、雑誌のまえがき、論文、出版情報などの構成を表現するものである。
言語区分	%Lang	アブストラクト、キーワードなど複数の言語を表現し、表示する場合の言語属性である。複数の言語を使用する要素が多くなるためにエンティティで宣言した。
読み (ふりがな)	%Yomi	データベース構築用に使用する情報である。アブストラクト、キーワードなどのフリガナを使用し、リスト、マークを行う。フリガナを使用する要素が多くなるためにエンティティで宣言した。
図表付加情報	%Illustr-model	図表のキャプション、説明など同一の要素をまとめたもの。メンテナンスの向上のためエンティティ宣言した。
図表サイズ情報	%Size	図表のサイズ、ファイル名、id など属性として必要な情報をまとめたもの。
図表ファイル名情報	%Loc	メンテナンスの向上のためエンティティ宣言した。
図表 id 情報	%Id	
基本文字要素情報	%Text	文字、文字列を表現する最長の要素である。文章で最低限の文字列を表現する場合に使用する。
基本段落要素情報	%Fulltxt	1つの段落内で表現する要素 (レファレンス、引用など) をまとめたもの。
%Chem-grp		化学情報 DTD の化学式を取り込んだ部分である。化学記号をまとめたものである。
%P.trans		AAP DTD の数式を取り込んだ部分である。数式で使用する単位をまとめたものである。
%M.math		AAP DTD の数式を取り込んだ部分である。数式をまとめたものである。

図 5

2. HTMLによる電子版情報管理誌の試作

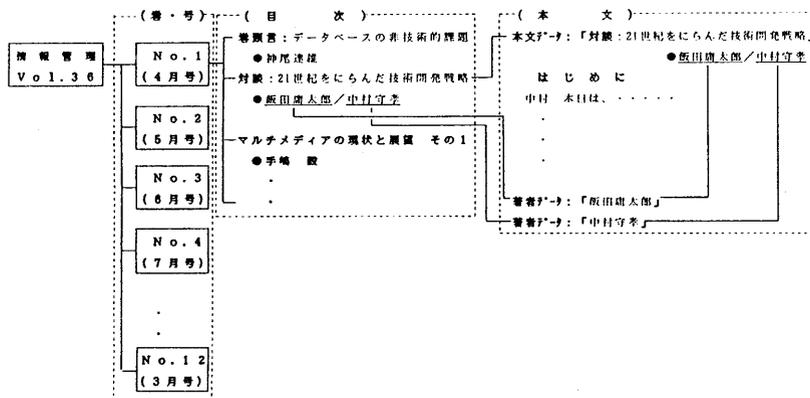
情報管理誌の分析と並行して、情報管理誌のSGML化により新たな展開も並行して検討する必要がある。即ち、情報管理誌のSGML化により、情報管理誌の全文データベース化、あるいは副次編集物の作成が容易になることは、容易に類推できる。しかし、片や依然として冊子体としての情報管理誌は存続していくことになるため、従来からの編集過程をそのまま踏襲しつつSGML化する事は作業が増えるだけで、SGML化によるメリットを見出し難く、そのシステムの恩恵が受けにくい。そこで、発想を大きく変えて、情報管理誌を電子媒体として出版するものと考えてみた。即ち、SGML化され、きちんとタグ付けされたデータであれば、当然電子出版ということはある。一方、情報管理誌は、冊子体として広く普及しており、雑誌のイメージを維持しつつ、電子的な媒体を指向することも必要となる。

そこで、電子媒体として、最近一般化しつつあるWWW (World Wide Web) で規定されているHTMLを用いて、情報管理誌の電子出版イメージを探るため、バックナンバーVol. 36からデータを抽出し、電子版情報管理誌を試作した。HTMLを取り上げた理由は、HTMLがSGMLのサブセットであるためSGML化したデータから移行が容易であること、WWW

の普及によりフリーウェアのビューアが多くあり、ビューアのためのシステム開発が少なくなる、などがあげられる。今回の試作は、各種評価の目的もあり、CD-ROMとして情報管理誌Vol. 37 No. 11 (1995年2月号)に付録として添付した。

以下、今回作成した電子版情報管理誌について説明する。この電子版情報管理誌は、もともと編集部にMS-DOSのテキストファイル形式で保管されていたものを、基に作成した。電子版ということではあるが、冊子体イメージに近いものを指向するため、Vol. 36に対する各号(1~12号)を目次としてつけた。文字だけでも、目次として十分な機能を持つが、ここでは各号の表紙もインライン・イメージデータとして利用するとともに、このイメージをリンク元にも利用した。各号の中に対しては、号毎の目次を設定し、各記事を対応するものとして本文にハイパーリンクを張った。従って、目次の中から標題を参照すると、記事にジャンプすることになる。情報管理誌の場合、著者名も目次の1項目となっているため、目次の中の著者名については、個別に著者情報を参照できるように、別個にジャンプできるようなハイパーリンクも設定した。この著者名参照は、目次からだけでなく、当然本文中の著者表示からも参照できる(図9, 10, 11)。

電子版情報管理のリンクの例



3. SGML方式による情報管理誌の問題点

今回、HTMLによる電子版情報管理誌の試作を行い、また並行して情報管理誌のDTDの検討を行うなかで、情報管理誌の性格から派生する多くの問題点が存在していることが分かった。その問題点は大きく分けて、SGML化のプロセスそのものにかかわるものと、従来からの冊子体の編集・発行のプロセスにかかわるものの二つに分けられる。

今回のDTDの検討とHTMLによる試作は、初めに述べたとおり従来の編集プロセスから離れ、既に冊子体として発行されたものを基に行ったが、現実的には紙媒体による冊子体の発行は今後とも継続して行う必要があるため、印刷媒体と電子媒体による同時あるいはタイムラグの小さい出版、更には全文データベース化を図る上での問題点を述べる。

(1) 情報管理誌は学会誌と商業誌の中間的な性格を持つ雑誌という位置付けにあり、その扱う分野は科学技術だけでなく、社会・経済や文学、歴史さらには美術・芸術など、非常に多岐にわたっている。また内容も論文や解説、あるいは半ば広告・宣伝の意味を含むものまであり、単一の執筆規定での扱いが難しい。そのため、ドキュメントタイトル、章・節などの定義が難しいものがかなり含まれ、その扱いが繁雑である。また、商業誌的部分としてイラスト的性格を持ったレイアウトや飾り等も相当含まれる。その扱いについては基本的には編集の意図、すなわち見栄えを良くし読み易さを多く求めるか、逆に内容を中心として定型的な紙面作りとするかといった観点に切り分ける必要がある。

(2) 現在、執筆規定に原稿用紙を使用した手書きまたはワープロ使用の場合はMS-DOSのテキストファイルで、と規定している。しかし、現状は約3割の手書き原稿とワープロ原稿(FD)7割となっている。この中でFDにより提出される原稿についても、執筆者の都合により様々なワープロ専用機のデータであったり、MS-DOSであってもワープロソフトの各種機能を駆使した

もの等もあり、テキストファイルへの変換が難しい場合もかなりある。

(3) 情報管理誌は編集委員会による企画を基にその内容を決めており、基本的に投稿原稿ではなく執筆依頼により編集している。その執筆依頼先は多岐にわたり、論文執筆のレベル等もさまざまであるため、原稿を受領してから編集の意図に沿って手を入れる場合が多い。また執筆者による著者校正を原則としており、ゲラ段階での加筆・訂正が多い。そのためSGML文書作成をどの段階で行うかによって基本的に処理の流れやシステムの構成が違って来る。これは、本誌のみの問題ではないが、電子版や全文データベースに、従来の編集・印刷工程では避けられない”手による”切り貼り作業による修正をどこまで反映させるかによっても違って来る。この問題もまた、基本的な編集に対する意識の問題に帰するのではないかと考える。

4. SGML化による編集の効率化

雑誌の編集という作業に共通の問題や、それぞれの雑誌等の固有の事情はあっても、最近の学会や出版業界の情勢を考慮すれば、電子媒体による出版や全文データベース化は避けて通れないのが現状であろう。しかし、それが従来の編集作業の後戻りとなったのでは、現場での実用化は困難となる。やはりSGML化によって編集の効率化が図れるか否かが今後の大きなポイントとなるのではなかろうか。編集という立場から前向きに対処していくには、編集の効率化を図るに当たって、考慮すべき点がいくつか考えられるが、次の様な点に関し基本的な検討を行っていけば、印刷媒体と電子媒体の同時発行への足掛かりとなるのではないかと考える。

(1) 誌面のできあがり(文字、レイアウト、イラスト、飾り等の出来栄)が低下しないか。1号分のページ数を確定するため、台割り等の段階で行うべき調整、即ち、図、表、写真等の大きさや飾り等の調整、が可能かどうか。

- (2) 原稿受理から発行までの制作期間が短縮されるかどうか。少なくとも、現状維持で在ることが必要となる。(現在は原稿受領から発行まで平均40日)。
- (3) 制作コストの削減が可能であること。
- (4) 執筆者の負担が増大しないこと。
- (5) 事務局(編集部)の作業量が増加しないこと。人手による作業の種類、量ともに増加を来さないこと。

情報管理誌における今回の電子版の試作とD T Dの検討は、まだ初手についたばかりであり、上記の5点に関して、その方法の具体的な手掛かりは見いだしていないが、今後更に検討と試作を重ね、実現の可能性を模索して行きたい。

5. 参考文献

- (1) JIS X 4151, 文書記述言語 S G M L, (1992)
- (2) 石塚英弘, S G M L と全文データベース, 情報管理, Vol.37 No.2, p148-159
- (3) 石塚英弘, S G M L による情報知識学会誌の印刷編集について, 情報知識学会誌, Vol.1 No.1, p24
- (4) S G M L 懇談会 News 編集部, S G M L 懇談会 News 作成用 D T D の紹介, S G M L 懇談会 News Sept.1992
- (5) 石塚英弘, S G M L 形式による学会誌全文データベースの構築と印刷, 情報知識学会誌, Vol.2 No.1, p23-48
- (6) 伊藤卓, 日本化学会欧文誌 (Bull. Chem. Soc. Jpn.) の全文データベース化と電子出版化への移行について, 化学と工業, 第46巻, 第1号
- (7) 作井正人, TeX とデータベース, 情報管理, Vol.37 No.3, p246-256
- (8) 高雄信一, ここまで来た電子文書処理, 行政と A D P, 1994年11月号
- (9) S G M L のススメ, オーム社, 1993年
- (10) WWWパブリッシング入門, インターネットマガジン編集部, インターネットマガジン, No.3