

デジタル図書館における基本出版技法：SGML

石塚英弘

図書館情報大学

〒305 茨城県つくば市春日 1-2

Tel: 0298-52-0511, Fax: 0298-52-4326, E-mail: ishizuka@ulis.ac.jp

概要

SGML(Standard Generalized Markup Language) は電子出版のための基本技法であり、電子図書館ないしデジタル図書館を支える重要な技術である。著者はまず SGML 方式の電子出版の概要と現状を紹介し、次いで文書データベースの構造の表現方法を例を示して説明した。また、SGML の機能の特徴と名前の由来も解説した。さらに Mosaic の HTML も解説し、SGML と HTML の違いを示した。また、例として日本化学会の欧文論文誌の SGML 方式電子出版を紹介した。さらに、SGML 関連ツールとそれを使った仕事についても解説した。

SGML: a fundamental technology for publication in a digital library

Hidehiro Ishizuka

University of Library and Information Science

1-2, Kasuga, Tsukuba, Ibaraki, 305, Japan

Phone: +81-298-52-0511, Fax: +81-298-52-4326, E-mail: ishizuka@ulis.ac.jp

Abstract

SGML(Standard Generalized Markup Language) is a fundamental technology for electronic publishing, and is a key technology supporting an electronic library or a digital library. The present author introduces the outline and the state-of-the-art of electronic publishing based on SGML (EP/SGML). He illustrates how SGML represents a structure of a document database. He also explains other features of SGML and why it is called "SGML". He also illustrates Mosaic HTML and the difference between SGML and HTML. EP/SGML of a primary English journal of the Chemical Society of Japan is introduced. He explains some SGML tools and some jobs using them.

Keywords

SGML, electronic publishing, electronic library, digital library, document database, WWW, Mosaic, HTML

1. はじめに

本稿では、SGML(Standard Generalized Markup Language)[1,2] と SGML 方式による電子出版について解説する。その理由は、1) 電子図書館ないしデジタル図書館では SGML 方式による電子出版物が重要な位置を占めるからであり、2)SGML は欧米では以前から各方面で使われていたが、日本では漸くこの 1・2 年は様々な方面で使われるようになってきた程度であり、未だ知らない人あるいは誤解している人が少なくないからである。なお、ここでいう電子出版物とは、CD-ROM だけでなく、コンピュータ・ネットワーク環境で動くシステム（たとえば、WWW Mosaic）上の出版物も含むこととする。著者（石塚）は数年来、SGML 方式の電子出版、特に学術雑誌の電子出版の研究開発に従事し、そのプロジェクトの殆どに参画してきた。そこで本稿では、その経験に基づいて、事例も交えながら SGML と SGML 方式による電子出版について解説する。

なお、未だ SGML が知られていない、あるいは誤解されている理由も理解できる。それは、1)SGML は直訳すると「標準汎用マーク付け言語」となるが、これだけでは内容が分からない、2)SGML が直接関連する事柄だけでも、印刷・出版、文書データベースないし全文データベース、文書記述法、規格など多方面に渡るため、全体像を把握し難い、2) 適用対象は文書一般で、マニュアルなどの技術文書から、議会の議事録、シェークスピア全集まで分野を問わないため、そのような異質なものが扱えるのかという疑問を持たれることがある、3) 具体的・実務的な部分だけでなく、抽象的な概念も含むため分かりにくい、4)Mosaic のデータ入力形式で HTML(HyperText Markup Language)があるが、それが SGML と似ているため、SGML と HTML が同じものと誤解されることがある、などの特徴を SGML は持っているからである。そこで本稿では、この種の誤解を解くことを主眼として、SGML を解説することにした。

2.SGML 方式出版の成立と普及

SGML 方式による出版の事例として、欧米では、SGML の解説書 [3,4]、マサチューセッツ医学会による AIDS の教科書、EU(欧州共同体)の議会の議事録などの先駆的事例に加えて、最近では、大手出版社 Elsevier Science が 1993 年から始めた大学図書館向け雑誌論文配信サービス実験 TULIP などがある。また、SGML 方式出版のためには、DTD(Document Type Definition、文書型定義)を決める必要があるが、DTD を設計した所としては、AAP(米国出版協会)、大手出版社の Wiley や Elsevier などがある。さらに、AT&T の主催する電子図書館実験 Red Sage でも、参加している Springer、Wiley、マサチューセッツ医学会など 21 出版者の中で SGML 方式採用の機運が高まり、94 年 11 月には Red Sage のための DTD 即ち"Red Sage DTD"を正式決定し、95 年中にこの DTD を使って雑誌出版を行う予定となった [5] とのことである。この事例は、SGML 方式が電子ないしデジタル図書館で有効なことが認められ、広く使われるようになってきたことを示している。なお、DTD については後で解説する。

日本での SGML 方式出版 [6] は、90 年 12 月の情報知識学会誌 [7,8]、91 年 3 月に出版された訳本「SGML 入門」[9]、同時期に学術情報センターによって試作された「SGML 実験誌」[10]などを先駆とし、92 年 4 月出版の訳本「実践 SGML」[11]、慶応大学図書館による CD-ROM 版三田商学研究（和文誌と欧文誌の最近 5 年分 150 論文）の試作 [12](92 年 7 月)、学術情報センター紀要の第 5 巻(92 年 12 月)を経て、93 年 1 月からの日本化学会欧文論文誌 (Bulletin of the Chemical Society of Japan, BCSJ) の定期印刷刊行 [13-17] に至る。BCSJ は月刊で毎号 60 論文を収録する大論文誌である。日本化学会という大学が、SGML 方式を採用したことは、日本でも SGML 方式が実用化したことを示している。なお、BCSJ の CD-ROM 版 [14,15,18] も試作された。そして WWW Mosaic 版も試作され、94 年 10 月末から 95 年 1 月末まで期間限定でテスト公開されている。

また DTD の設計では、上記のプロジェクトに加えて、学術情報センターが学術論文用の汎用 DTD を

設計したこと [6] が挙げられる。この DTD は、BCSJ の全文 DB を学術情報センター全文 DB 検索サービスに載せる際に使用されている。

なお、著者は情報知識学会誌、SGML 実験誌、学術情報センター紀要、日本化学会欧文論文誌、学術情報センターによる学術論文用汎用 DTD のプロジェクトに参画した。

3.SGML と SGML 方式電子出版

3.1.SGML 方式電子出版

全体像を把握してもらうため、SGML 方式電子出版の方から解説する。SGML 方式では、印刷出版を 1) 文書データベース (document database, 文書 DB) の作成と、2) 文書 DB からの印刷ないし電子出版物の作成の二段階に分ける。まず、文書 DB を作り、それを基に、印刷、CD-ROM 作成、検索システム用の全文データベースの生成などを行うのである。印刷で言えば、2) は如何にレイアウトするかということになる。そして、SGML が対応するのは 1) の段階である。この考え方の基は ISO の電子印刷出版モデルで、そのモデルでは 2) のレイアウトはさらに、element(要素) 単位の処理と、ページ単位の処理とに別れる。ISO では前者を DSSSL、後者を SPDL に対応させている [19-21]。ただし、実際の処理系、たとえば、LaTeX[22,23] などでは、DSSSL と SPDL の機能を併せ持っているので、本稿では 1) と 2) の二つに分けて説明した。

ここで、文書 DB や element は SGML 特有の用語なので、簡単に説明しておく。「文書」とは、たんなるテキストだけでなく、図表、写真、数式等々、文書に現れる全てを含むことを表している。「文書 DB」とは、文書を構造を持った DB と考えることを示す。文書には、章、節、段落などの階層構造があり、また、本文から図、表、文献、注などへの参照といった参照関係がある。これらの構造を表現したものを文書 DB という。そして、element とは DB の構成要素である。ただし、章も節も段落も element と言い、章は節で、節は段落で構成されていると言うから、element といっても最小単位ではない。

なぜ、文書 DB を導入したか。その理由は、印刷物、CD-ROM、検索用 DB など多様な電子出版を行うには文書 DB を基にした方が効率的だからである。以前は、印刷といえば電算写植 (CTS) で、全文 DB を作る場合は電算写植データからプログラムで変換して作っていた。しかし電算写植では、1) 図表、写真などは別扱いのため、プログラムで生成できるのはテキスト部分のみとなる、2) 電算写植データは印刷・レイアウトのためのデータ (ゴチック、14 ポイント、センタリング等の印刷制御コマンド込みのテキスト) であって、それがどの項目 (たとえば、章のタイトル) かは書いてないため、どの項目かを判定しきれない場合もあり得るし、レイアウトが異なってタイトルでも左寄せになればプログラムの書き直しになる、3) 印刷制御コマンドはハード・ソフト個別のため、それぞれに変換ソフトを作る必要がある、などの欠点があった。SGML 方式はこれらの欠点を解消する。

以上説明したように、SGML の技術は文書 DB 構築のための技術であって、印刷の機能やデータベース検索機能とは切り離されている。このことは、SGML に対する誤解を呼ぶ種となった。たとえば、印刷出版だけに興味を持つ人は、印刷機能を持たないことに失望を感じたかもしれない。また、データベースのみに興味を持つ人は、検索機能がないことに意外な感じを持っただろう。しかし、今ではこの機能分担は寧ろ正解であったといえる。その後に進歩した印刷ソフトや検索ソフトの多くが SGML 文書 DB に対応するようになったからである。

3.2. 文書 DB の構造表現の概略

文書 DB の構造は、DTD、テキスト・データ、図や写真を収めた外部エンティティの 3 つで表現される。DTD は文書 DB の構造を定義する。一般的なデータベースの用語でいえば、DTD はスキーマに相当する。

そして、テキスト・データはDTDに従って書かれる。テキスト・データには、文章だけでなく、表や式も含まれる。なぜなら、表や式も文字で書けるからである。図や写真は別ファイルにビット・データで表現され、その形式はTIFF, PICT, EPS(encapsulated PostScript)などSGMLとは別の規格となるので、外部エンティティといわれる。なお、その図や写真の表現形式が何かは、DTDの中で該当する図や写真ごとに書く。本文と図や写真の参照関係については、DTDに参照関係の表示法を定義し、テキスト・データ中に参照関係を記述する。

前節(3.1)に述べたように、文書DBの構造には階層構造と参照関係があるが、階層構造にはelement間の前後関係も存在する。たとえば、文書では、「本のタイトル」、「本体」、「索引」の順になり、これが逆転することはない。また、「章」は「章のタイトル」と「章の本体」から構成されるが、「章のタイトル」の方が先に来る。

ここで「デジタル図書館における基本出版技法:SGML」というタイトルの解説(仮にreviewと名付ける)があるとして、これを例にしてDTDとテキスト・データを説明する。この解説が収録されている逐次刊行物「デジタル図書館」の印刷様式から推定して、文書は図1に示す階層構造を持つとする。すなわち、reviewはfront matter(前付, fmと略), body(本文, bdy), back matter(後付, bm)という3つの要素(element)が、この順で並んでいるとする。また、fmは要素:title, author, affiliation, abstractの順で構成される。bdyは章で、章は節で、節は段落で構成される。あるいは直接段落となってもよい。等々とする。

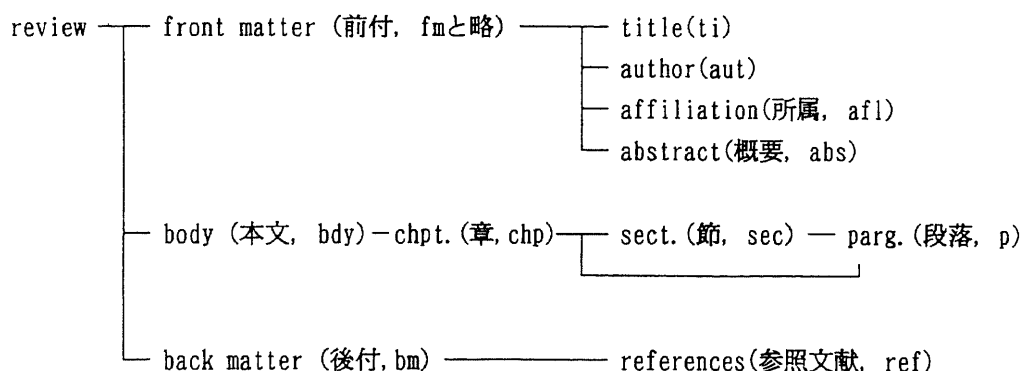


図1. 解説“review”の階層構造

3.3.DTD と SGML

図1の構造をDTDとして書けば、図2となる。DTDはDocument Type Definitionの略で、JISでは文書型定義と訳している。なお、DTDは文書構造をSGMLの構文を使って定義・記述したものである。言い換えれば、SGMLはDTDを書くための言語である。SGMLの説明が中々出てこないと思われたであろうが、その理由はDTDの後で説明した方が分かりやすいと思ったからである。

図2の各行はELEMENT文で、個々の要素がどの要素から構成されるかを定義している。なお、図1

では章のタイトル (chpti) は省略したが、DTD では要素に加えてある。また、ref の ELEMENT 文は省略してある。

```
<!ELEMENT review - - (fm, bdy, bm?)>
<!ELEMENT fm - 0 (ti, aut, afl, abs)>
<!ELEMENT bdy - 0 (chp+ | sec+ | p+)>
<!ELEMENT bm - 0 (ref?)>
<!ELEMENT chp - 0 (chpti, (sec+ | p+))>
<!ELEMENT p - 0 (#PCDATA)+>
```

図 2. "review" の階層構造を表現した DTD

SGML の文法では、大文字小文字は同じものとして扱われる。ここでは要素の名前が目立つように小文字にしてみただけである。ここで使われている要素名は省略した名前で見分けにくいと思う読者も少なくないだろう。要素名は DTD の ELEMENT 文で定義するのであるから、分かりやすい名前を定義してもよい。通常は名前の長さは 8 文字以下と定められているが、DTD で長さを変えることも可能である。また名前の頭の文字は A から Z (小文字でもよい) であるが、その後ろには A から Z のほかに、数字、ピリオド、ハイフンなども使うことができる。この範囲で適当な名前を考えればよい。

ELEMENT 文の形式は、<!ELEMENT、要素名、開始・終了タグ省略の有無、開始タグから終了タグの間に出現できる構成要素、> である。図 2 の 1 行目の記述は、review という要素は fm, bdy, bm という 3 つの要素で構成され、また、この順で出現することを示している。fm, bdy についても同様である。なお、#PCDATA とは解析対象文字データという意味で、予約語である。データ中に実体参照や開始タグなどのマークを含んでいるかどうかを SGML パーサが解析する必要があることを示す。# は予約語を示す記号である。SGML パーサについては後で説明する。

これら ELEMENT 文の記述は図と同じことを表現しているが、DTD の方が構造をより厳密に記述できる。例えば、存在しなくともよいもの (optional, 出現頻度 0 または 1) は "?" で示す。bm, ref などは optional である。また、章や段落などは一回以上繰り返し出現する (1 から N 回) が、これは "+" で示す。なお、無いこともあり、繰り返すこともある場合 (0-N 回) は "*" で示す。また、構成要素によっては順に現れるのではなく、どちらかが現れる場合もある。これは | (OR 記号, 縦棒) で表現する。このほか & (AND 記号) もあるが、これは & の両側の要素が必ず現れること、ただし順序はどちらでもよいことを示す。

開始・終了タグ省略の有無についても説明しておく。2 行目は "- -"、3 行目は "- 0" となっている。左は開始タグ、右は終了タグを示し、"- -" は省略できないことを、"- 0" は omitted を示す。たとえば、要素 review は開始・終了タグともに省略できないことを、fm は終了タグは省略できることを示す。なお、開始タグ、終了タグは文書テキストの所で説明する。

なお、文書は、単行本、論文、レポートなど文書の種類によって、その構造が異なるから、種類ごとに DTD は異なる。DTD を自分で書くには SGML の構文を知る必要があるが、種類によっては DTD が既に用意されていることがある。その場合は、それを使えばよく、自分で書く必要はない。

3.4. テキスト・データ

review のテキスト・データを図 3 に示す。ここで、<review>, <ti>, <aut> など、<要素名> の形式になっているものは、それぞれ、要素名が、review, ti, aut の要素の始点を示すマークであり、これを開始

タグという。また、要素の終点は</review>のように、</要素名>で示す。これを終了タグという。この例で<ti>や<aut>の終了タグ</ti>や</aut>が無いのは、この文書テキスト用の DTD で、これらの省略が許されているからである。なぜ、許されるかと言えば、ti の次は aut, aut の次は afl と定義されており、そのため<aut>が来れば ti の終了が、<afl>が来れば aut の終了が判定可能だから省略できるのである。一方、</review>の方は全体の終わりを示しており、これがないと review の終わりが判定できない。そのため、省略できないのである。

なお、文章中に記号として”<”や”>”を使いたくなることもある。その時は、DTD で”<”を<、”>”を>と定義しておけば、<や>と書けばよい。ここで、lt は lower than、gt は greater than を意味している。この手法は、外字を定義するのにも使える。

ところで、図 3 のようなマーク付き原稿を作るのは面倒と思われるであろう。実際には、<要素名>、</要素名>無しにテキストを入力する方法があるから安心してほしい。たとえば、「SGML 対応エディタ」の使用、自分で違和感のない記号を DTD で定義し、その記号を正規の開始・終了タグにソフトで変換する方法などがあるし、最近ではワープロ原稿からの自動変換 [24] もある。

```
<review>
<fm>
<ti>デジタル図書館における基本出版技法：SGML
<aut>石塚英弘
<afl>図書館情報大学
    (中略)
<abs>SGML は電子出版のための基本技法で.....
    (中略)
<bdy><chp><chpti>はじめに
<p>本稿では、SGML.....(中略)..... について解説する。
<p>なお、未だ SGML が.....(中略)..... することにした。
<chp><chpti>SGML 方式出版の成立と普及
<p>SGML 方式による出版の事例.....(中略)..... 後で解説する。
    (中略)
<chp><chpti>SGML と SGML 方式電子出版
<sec><secti>SGML 方式電子出版
<p>全体像を把握してもらうため、.....(中略)..... 説明した。
    (中略)
</review>
```

図 3. "review"のテキスト・データ

3.5. 非階層構造の表現

文書の構造には、章・節・段落といった階層構造と、注・参考文献・図表などの非階層構造とがある。節は章の中に、そして段落は節または章の中に必ず存在する。章が節の中に存在することは原理的にない。しかし、注・参考文献・図表などの存在場所は自由度が高い。たとえば、注は脚注として同じ頁の下に置く方式もあるし、章の終わりにまとめて置く方式もある。参考文献も脚注のように下に置く方式と、章の

終わりに置く方式とがある。図や表はどこにも存在しうる。SGML ではこの種の要素を floating element (浮動要素) という。また、注・参照文献図表などは存在場所の自由度が高いたくだけでなく、本文との間の参照関係もある。重要な文献や図表は何度も参照されることがある。注・参照文献・図表などは存在場所の点でも参照関係の点でも非階層的なのである。

そこで SGML 方式では次のようにして非階層構造も表現できるようになっている。

1) 浮動要素の定義

まず、図や注などが文書データベース中の任意の場所に存在できることを DTD で定義しておく必要がある。これは例えば次のようにして行う。

```
<!ENTITY % floats "fig | fn">
<!ELEMENT review - - (fm, bdy, bm?) +(%floats;)>
```

ここで、%はパラメータ・エンティティを示す記号である。第1行は %floats という名前のパラメータの値は fig または fn であることを定義しており、%floats の内容は図や脚注とするという意味になる。2行目の "+" は包含 (inclusions) を示し、任意の場所に () 内のものが存在してよいことを意味する。%floats の後ろに ";" が付いているのは、% など記号で始まるエンティティを引用する時は最後に ; を付けることになっているからである。

2) ID を使った参照

これは、被参照データと参照指示データとを ID を使ってリンクする方法である。ID の操作は SGML の属性機能を使って行う。note (注) を例として説明する。

DTD のうち、被参照データについては、

```
<!ELEMENT note - - (p)>
<!ATTLIST note id ID #IMPLIED>
```

と定義する。ここで、ATTLIST 文は属性リストを宣言する文で、その形式は、<!, ATTLIST、要素名、属性名、属性の宣言値、属性の省略時値、> である。また、属性名から属性の省略時値までは組で繰り返し可能である。note の場合は、要素 note が id という属性名を持ち、それは ID (identifier value、識別子値) であり、属性値が与えられていない場合は #IMPLIED (アプリケーションが省略時値を決める) であることを示す。

参照指示データは、note の refer という意味で要素名を ntref にすれば、

```
<!ELEMENT ntref - 0 EMPTY>
<!ATTLIST ntref refid IDREF #IMPLIED>
```

と定義される。ここで、EMPTY とは構成要素が無いことを示し、IDREF (identifier reference value、識別子参照値) は被参照データの識別子を参照することを示す。

そして、被参照データの記述は、たとえば、

```
<note id=note1>開発者 3 人の頭文字を並べたものでもある。</note>
```

となり、本文の方の記述は

.... 汎用マーク付け言語 GML<ntref refid=note1> はマニュアルや仕様書....

となる。この方式ならば、

```
....GML<ntref refid=note1> 以外でもシステムの命名の仕方には....
```

のように、本文の別の所で同じ注を参照することができる。

3) 図や画像の取り扱い

図や画像は、その番号・キャプション・図や画像の本体で構成される。番号やキャプションは文字データである。しかし、図や画像の本体はドット・データすなわちビット・データであることが多く、SGML による解析の対象外である。そのため、次に示すように取り扱いが少し異なる。

- a) 図や画像のデータは別ファイルに置く。
- b) 図や画像の表現形式には、PICT, EPS, TIFF などがあるので、それを区別する必要がある。
- c) 図や画像の大きさを指定する必要がある。

```
<!ELEMENT figure -- (nf?, figcp?, figbdy?)>
<!ATTLIST figure id ID #IMPLIED
                 file CDATA #IMPLIED
                 type CDATA #IMPLIED
                 sizex CDATA #IMPLIED
                 sizey CDATA #IMPLIED>
<!ELEMENT figref - 0 EMPTY >
<!ATTLIST figref refid IDREF #IMPLIED>
```

図 4. 図の DTD

図 4 に図の場合の DTD を示したが、画像でも基本的に同じである。nf は図の番号、figcp はキャプション、figbdy は図本体である。キャプションには図のタイトルと説明が含まれている。file はファイル名で、その値は例えば、"b:\figure\figure1.dat" のようになる。ここで、"b:\..." はディスクドライブ B のディレクトリ figure にある figure1.dat というファイルとして保管されていることを示す。なお、CDATA とは予約語で SGML の解析対象外の文字データのことであり、type は PICT ほか表現形式を示す。図の大きさは sizex と sizey で表す。要素 figref とその属性については既に述べた手法と同じである。

4) その他の非階層構造

表は DTD で定義し、それに従って表そのもののデータを書き、属性を用いて参照すればよい。紙面の都合上、説明は割愛する。また、それ以外の非階層的な構造の定義も属性を用いて行う。Bryan の本 [9] に詳しく書かれているので、詳細についてはそれを参照されたい。

3.6. SGML の機能と名前の由来

SGML は国際標準 ISO 8879 であるが、これまで解説したように、SGML は DTD を書くための言語であるから、文書の構造・構成を標準化した、すなわち DTD を標準化した訳ではない。SGML の考え方としては、全ての文書を記述できる DTD を決めるのではなく、対象とする文書とそれを扱う人々に対して DTD を定めようという訳である。SGML 方式の文書 DB は、DTD とそれに従ったテキスト・データや外部エンティティから構成されているから、他の DTD に基づく文書 DB にプログラムで変換できる。そ

れが SGML 方式の謙虚な点であり、实际的・現実的なアプローチと言える。最近、SGML が注目されている理由もその点にあると思っている。

ところで名前の由来であるが、SGML では文書を構造化するために element を識別するが、それに開始タグと終了タグを用いた。これが generalized markup である。本来 mark up とは、センタリング、フォントや字の大きさの指定などを編集者が赤字で書き込むことであった。これを電子的に指定するので markup と一語にし、また構造を指定することにより、印刷用のハード・ソフトの個別のコマンドから自由になり、一般化 (generalized) された。そこで、generalized markup となった。language は言語を、standard は ISO のような標準となったことを示す。確かに分かりにくい名前ではある。

4. WWW Mosaic と HTML

阪口の解説 [25] にもあるように、Mosaic は SGML に準拠しており、そのためのデータ入力言語が HTML である。Mosaic の DTD は文書の種類に依らず一つで、DTD そのものは公開されておらず、テキスト・データの書き方が HTML として公開されている。HTML は文書の階層構造は表現しておらず、ハイパーテキスト [26] のリンク機能で代わりをするようになっている。また、element の種類は少なく、DTD を直せないため追加もできない。一方、Mosaic の強力なハイパーテキスト機能が使える点は便利である。

SGML 文書 DB と HTML の違いを示すために、先程示した解説の SGML のテキスト・データを HTML 形式に書き換えて、図 5 に示す。文書のタイプは一つしかないから、最初は <HTML> で最後は </HTML> となる。<H1>、<H2> はヘディング (見出し) で、1, 2, 3 の 3 段階ある。1 の方が字が大きい。 を使えば、ボールド (太字) になる。どう使い分けるかはデザインの問題である。段落は終わりに <P> を付けることになっている。

参照関係は、参照箇所の文字列を と で挟む。ここで、A はアンカー (錨の意味、ハイパーテキストのリンク元を示す) の A である。また、" " 内は被参照データの場所を示す。

ここに示した例は説明用のごく簡単なもので、もっと凝ればより良いものができることを指摘しておく。

SGML 文書 DB の方が DB としてより詳しい情報を持っている。そのため、SGML 文書 DB から HTML はプログラムで生成することができる。しかし、その逆は難しい。そのため、HTML で文書を保持するよりも、SGML 文書 DB で持っていた方がよい。HTML の方は検索表示用と考えるのが妥当であろう。

5. SGML 文書 DB の一例としての日本化学会欧文論文誌

日本化学会欧文論文誌は 93 年 1 月号から SGML 文書 DB となっている。印刷はプログラムで LaTeX に変換して行っている。また、この DB は学術情報センターの全文検索システムへも提供 [18] されている。学術センターのシステムは学術論文用の汎用 DTD を持つことにより、SGML 文書 DB へのインターフェースを持っているので、日本化学会と学術センターの二つの SGML 文書 DB 間の自動変換が行われている。これ以外に、MS Window 上のハイパーテキスト・システム Multimedia Viewer で動く CD-ROM も試作 [18] されている。このビューアは RTF (Rich Text Format) のインターフェースを持っているので、SGML 文書 DB から RTF へのプログラムによる変換を行って実現した。また、94 年 10 月末から 95 年 1 月末までは Mosaic 版もテスト公開している。これもプログラムによる変換で実現した。なお、Mosaic 版には 93 年 1 月号から 94 年 6 月号までしか載せていない。DB が有りながら、新しい号を載せなかった理由は、印刷物の売れ行きの邪魔を避けるためである。また、これはあくまでテストであり、継続は今のところ考えていない。

```

<HTML>
<TITLE>デジタル図書館における基本出版技法：SGML</TITLE>
<H1>石塚英弘</H1>
<H2>図書館情報大学</H2>
    (中略)
<H2>概要：</H2>
SGML は電子出版のための基本技法で.....<P>
    (中略)
<H1>はじめに</H1>
本稿では、SGML.....(中略)..... について解説する。<P>
なお、未だ SGML が.....(中略)..... することにした。<P>
<H1>SGML 方式出版の成立と普及</H1>
SGML 方式による出版の事例.....(中略)..... 後で解説する。<P>
    (中略)
<H1>SGML と SGML 方式電子出版</H1>
<H2>SGML 方式電子出版</H2>
全体像を把握してもらうため、.....(中略)..... 説明した。<P>
    (中略)
</HTML>

```

図 5. "review" の HTML 形式データ

6.SGML 関連ツール

ここでは、SGML 方式の全文データベースを作って利用する時に、どのような仕事があるのか、また、それはどの SGML 関連ツールを使って行うのかを説明する。

1) DTD の作成

自分が扱う全文データベース用の DTD が既に存在する場合は、それを利用すればよいので、作る必要はない。そうでない場合は、既存の DTD を改良するか、あるいは自作すればよい。作った DTD が SGML の文法に合っているかは SGML パーサ (parser) でチェックする。

パーサとしては、たとえば、SEMA Group の Mark-It, Exoterica の OmniMark, 富士通の SGML/Base などがある。また、public domain のソフトとしては、sgmls がある。

2) 文書テキストの作成

DTD に従った文書テキストを作る方法は、A) SGML エディタを使う、B) 既存のワープロ・ソフトを使うの二つに大別される。SGML エディタを使えば、プロンプトに従って該当データを入力することによって容易に文書テキストを作ることができる。B には B1) 特定の記号を DTD で設定して入力し、SGML パーサを通して変換する、B2) ワープロ・ソフトのスタイル・シートの機能を使う、B3) 同じくマクロ機能を使う、B4) 変換ソフトを使うなどの方法がある。このようにすれば、B でも開始タグや終了タグの入力無しで文書テキストを作ることができる。日本化学会では 94 年 4 月から MS Word, WordPerfect, PageMaker で使用できるスタイルシートを使った電子投稿の実験も行っている。なお、B2, B3 などの場合は SGML パーサを通してチェックした方がよい。

SGML エディタとしては、たとえば、SoftQuad の Author/Editor, 松下電送の SGML-Plus などがある。また、ワープロ形式から SGML 形式への変換ソフトも最近いろいろ出てきている。

3) 文書テキストの印刷

SGML 文書用のフォーマッタを使うか、SGML 形式から変換して LaTeX で印刷するか、または印刷ソフトで SGML 対応機能を持ったものを使えばよい。たとえば、LaTeX 出力が可能なものとしては、Mark-It, SGML/Base, SGML-Plus などがある。また、SGML 対応の印刷ソフトとしては、日本ユニテックの GRACE, Interleaf の Interleaf5<SGML> などがある。

4) 電子図書の作成

SGML に対応しているものとしては、DynaText, WWW, Guide などがある。それ以外でもそのソフトへの変換プログラムを書けばよい。

7. おわりに

SGML による全文データベースは欧米では広く使われるようになってきているし、日本でも漸くその傾向が出てきた。これは、SGML 関連のツールがいろいろと出てきたこと、紹介の文献や本が出版されるようになったことなどによると思われる。また、SGML に動画や音声を取り入れた HyTime も ISO 規格になり、94 年夏には JIS 規格にもなった。

さらに注目されることとして、アメリカの CALS の民生化がある。CALS はこれまで SGML の普及を進める原動力となっていたが、CALS が国防省のプロジェクトであったため、SGML は自分には関係ないと思う人も少なくなかった。しかし、93 年 CALS の本称が Continuous Acquisition and Life-cycle Support に変更され、一般官庁、民間企業間の取引にも適用しようとする動きが活発になってきた。そして、CALS はネットワーク環境 NII(情報ハイウェイ) と結びつき、新たな展開を見せている。

今後、SGML の浸透は必然となろう。特に電子図書館ないしデジタル図書館では SGML は必要不可欠となっていくだろう。今後の進展に期待したい。

なお、本稿は頁数の関係もあり、触れずに終わった点も少なくない。そこで最後に解説書を紹介しておこう。SGML とは何かを手短かに知るには文献 [27] が良い。また、各分野の専門家が「SGML は如何に有効か」を書いた本 [28] も参考になろう。Herwijnen[11] は SGML 方式を実行するにはどうしたらよいかを書いている。Bryan の本 [9] は SGML の機能に詳しい。SGML の元祖 Goldfarb の本 [29] は SGML の全貌を明らかにしようとした本である。そして、本稿が SGML の理解の一助になれば幸いである。

文献

- [1]ISO 8879-1986, Information processing - Text and office systems - Standard Generalized Markup Language (SGML), Oct. 15, 1986.
- [2]JIS X 4151, 文書記述言語 SGML, 日本規格協会, 1992 年 7 月, 231p.
- [3]Martin Bryan. SGML: An Author's Guide to the Standard Generalized Markup Language. Addison-Wesley, 1988.
- [4]Eric van Herwijnen. Practical SGML. Kluwer Academic Publishers, 1990.
- [5] 根岸正光, 私信.
- [6]H.Ishizuka. The reception of SGML based electronic publishing by Japanese scientific community. Proc. 47th FID(Int. Fed. Inf. Doc.) Conf. Cong., pp.505-508 (Oct. 1994, Omiya).

- [7] 情報知識学会誌, Vol.1, No.1, 1990 年 12 月, 98pp.
- [8] 石塚英弘. SGML 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, pp.23-48(1991).
- [9] Martin Bryan 著, 山崎俊一監訳, 福島誠訳. SGML 入門. アスキー出版局, 1991 年 3 月, 378pp.
- [10] 根岸正光. 『SGML 実験誌』の出版について. SGML 実験誌 1991, p.i-iii(1991), 学術情報センター.
- [11] Eric van Herwijnen 著, SGML 懇談会実用化 WG 監訳. 実践 SGML. 日本規格協会, 1992 年 4 月, 305pp.
- [12] 齊藤雅. 学術論文 CD-ROM システム. 情報処理学会人文科学とコンピュータ研究会資料, 15-1, pp.1-7(1992).
- [13] 伊藤卓. 日本化学会欧文誌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, 46 巻 1 号, pp.92-95 (1993).
- [14] 石塚英弘, 伊藤卓, 榎敏明, 千原秀昭, 中西敦男, 田中洋一. 日本化学会欧文誌の SGML 形式全文データベースの構築・印刷そして検索. 情報処理学会情報学基礎研究会資料, 29-1, pp.1-8 (1993.5).
- [15] 齊藤伸雄. オブジェクト指向による論文誌ビューアの開発. 日本印刷学会年会 (1993.5).
- [16] 伊藤卓, SGML を用いた学術論文誌の電子出版化—日本化学会欧文誌の場合—. 情報の化学と技術, Vol.43, No.12, pp.1089-1096(1993).
- [17] 上嶋邦義, 学術論文誌の全文データベース化と電子出版の実際. 医学図書館, Vol.41, No.1, pp.38-45(1994).
- [18] 石塚英弘, 伊藤卓, 千原秀昭, 根岸正光, 中西敦男, 田中洋一. 全文検索システムのリーソースとしての SGML 方式データベース. 情報処理学会情報学基礎研究会資料, 33-6, pp.29-34 (1994.5).
- [19] 安達淳「国際規格; 文書スタイル意味指定言語 (DSSSL) の概要」情報処理, Vol.33, No.2, pp.176-184(1992).
- [20] 高橋亨「標準ページ記述言語 (SPDL) とその動向」情報処理, Vol.33, No.6, pp.689-697(1992).
- [21] 小町祐史編著. 電子出版技術入門. オーム社, 1993 年, 175pp.
- [22] Donald E. Knuth. The TeXbook. Addison-Wesley, 1984. (齊藤信男監修, 鷺谷好輝訳. TeX ブック. アスキー出版局, 1989).
- [23] Leslie Lamport. LaTeX: A Document Preparation System. Addison-Wesley, 1986, 242pp. (E. Cooke・倉沢良一監訳, 大野俊治・小暮博道・藤浦はる美訳. 文書処理システム LaTeX. アスキー出版局, 1990).
- [24] 石塚英弘, 伊藤卓, 竹内敬人, 千原秀昭, 中野英彦, 眞野倅一, 吉村忠与志, 中西敦男, 田中洋一. 電子投稿による SGML 方式全文データベースの作成—日本化学会の実験—. 情報処理学会情報学基礎研究会資料, 35-1, pp.1-8 (1994.11).
- [25] 阪口哲男. Internet における情報資源の利用と提供の実際. デジタル図書館, No.1, pp.43-50 (1994.8).
- [26] Jakob Nielsen. HYPertext & HYPERMedia. Academic Press, 1990. 齊藤孝訳. HYPertext & HYPERMedia. HBJ 出版局, 1991, 257pp.
- [27] 吉岡誠編著. SGML のススメ. オーム社, 1993 年 11 月, 167pp.
- [28] 根岸正光・石塚英弘編. マルチメディア選書, SGML の活用. オーム社, 1994 年 12 月刊行予定.
- [29] C.F.Goldfarb. The SGML Handbook. Oxford University Press, 1990, 633pp.