

SGMLをインフラとした情報通有の実現

吉岡 誠

富士通(株) 光ディスク応用開発部

情報通有の視点ではSGMLで記述した文書は、個別ワードプロセッサで記述した文書より、有利である。ただ、SGML文書の入力には普通の人にはなじまない、そこでSGMLを普及させるためには、SGMLをあまり意識しない入力方法の提供が必要で、当面はSGMLへの自動変換が可能な通常の平文の書き方を尊重した簡易タグを利用した文書記述方法を推奨したい。ISO規格化が未完の体裁定義の標準化については早期策定と関連ツールの提供が必要である。媒体格納については、データベース的価値の高いもの、版数管理の必要なマニュアル等はSGMLのデータベース形式で、個人用ファイルには平文の延長で簡易タグ形式で格納するので良いと思われる。

Toward Information Sharing with SGML as an infrastructure

Makoto Yoshioka,

Optical Disk Application Development,

File Systems Group, Fujitsu Limited

SGML written document has an advantage over a document written with proprietary word processors for information sharing.

But applying SGML we encounter problems:

1) Marking up SGML document is not an easy task for ordinal word processor users. There are several methods and tools assisting user to input SGML markup, still ordinal user seems unhappy to use these methods and tools. So, Author hope to recommend a standard practically useable simple tag based structuring method for documents.

The point of this format is it is easy to understand from people and easy to analyse from computer. This format may be used for interchange information with network.

2) For formatting SGML document the delay of standardizing DSSSL (Document Style and Semantic Specification Language) is a problem. Sometimes TeX is used, but there are problems in formatting Japanese SGML document through TeX. Interim step we can use Word Processor's Style sheet for formatting Japanese SGML document.

If we have (a) SGML and/or (b) standard Simple-tag based documents, what format should be used in the storage media. The author's recommendations are:

format (a) for Manuals which needs revision control, and for Database oriented documents which needs standard query language,

format (b) for personal information.

## 1. SGMLの利点

SGMLはStandard Generalized Markup Languageの略称である。SGMLの利点は：

-国際規格ISO8879:1986(国内規格は「文書記述言語SGML」JIS X 4151-1992)であり、システム間での文書交換を容易にする規約である。

-SGML文書はテキスト列であり、マークにより構造化されている。テキスト列であるから、人間にも読めて、通信との相性が良い。計算機による構造解析が簡単で、一つの文書を多角的に利用・再利用できる。データベース化も容易である。

-印刷/表示用の出力体裁はやはりテキスト列として別に定義する。出力体裁定義を各種出力装置に適した形式で定義することにより、同一文書をそれぞれの出力装置に適した体裁で出力が出来る。同一装置に対しても利用目的に従って、同一文書を異なった体裁で出力することも可能である。

-これに比較して通常のワープロ文書ではテキスト情報とバイナリの体裁定義が混合したバイナリ形式を、各商品固有の形式で持っているために、他システムとのスムーズな交換、多角的利用、計算機による解析も困難で、自在な体裁定義への対応も不可能となっている。

HyTime(Hypermedia Time-based Structuring Language:ISO規格)により、SGMLがハイパーメディアにも適用できるようになった。

## 2. SGMLを取り巻く標準化の課題

-出力体裁に関する国際規格DSSSL(Document Style and Semantic Specification Language)の規格化が遅延している。

-ハイパーメディアの入出力制御にかかわる国際規格SMML(Standard Multimedia Scripting Language)の規格化が未完である。

### 3. SGML推進上の課題

#### 3.1 SGML文書入力の問題

SGML文書はDTD(Document Type Definition)で定義された構造に基づいて、マークを付けて記述する必要がある。ところが、マーク付けはワープロ記述に慣れた普通の人には抵抗がある。又、マーク付けをした文書は人間には少し読みづらいと言う問題もある。

この問題を速やかに解決しないと、SGML文書の流通を爆発的に広めることは難しい。現在実施されている各種方法を挙げながら、その功罪を論評したい。

##### 3.1.1 入力方式各論

###### 1) マーク後付け型

適当なEditor, WP(Word Processor)を利用して、先ず、平文を完成させ。これに、後から必要なマークを付け、SGMLパーサを利用してエラー検出/訂正を繰り返してSGML文書を完成させる。ルーチンワークでマニュアルを書くような人にとっては、効率的な方法であるが、一般の人にとっては、マーク付けそのものが、苦痛。マーク付け後の文書は当然読みづらい。

###### 2) SGMLマーク入力支援型

既存のEditor, WPの付加機能として、入力可能なマークの一覧を出して、これを選択させることにより、マーク付け入力を支援するもの。SGML文書として完成させるにはマーク後付け型と同様の手順を踏む必要がある。マーク付けは少し簡単になるが、マーク付け作業から開放されるわけではないし、マーク付け後の文書は読みづらい。

###### 3) スタイルシート利用型

既存のワープロ文書入力時に個々の要素にスタイルシートを適用させ、個々のスタイルをマークに変換することにより、SGML文書とする。SGML文書としての完成はマーク後付け型と同様の手順を踏む必要がある。入力時にSGMLそのものを意識する必要は無いので、抵抗は少ないと思われる。ワープロ上ではSGMLマークによる文書の読みづらさは表面には見えない。

#### 4) SGML構造化Editor型

文書入力時に、選択したDTDに基づいた文書構造の入力しか許さない。マーク付けはSGMLを意識した構造化Editorが支援する。常に正しいSGML文書が入力できるのだが、普通のユーザには、こうしたスタイルでの入力の強要はなかなか馴染まないようである。

#### 5) 簡易タグ構造化文+オートタガー型

簡易タグをベースに構造化した平文入力を解析して、自動的にSGMLのマークを付ける方式。通常の平文入力の形式と余り変わらないので、入力にも抵抗が無く、人間にも読み易い、また、構造化してあるので自動的にマーク(タグ)を付けるのも容易。SGMLとしての文書の完成手順はマーク後付け型と同様の手順を踏む必要がある。

### 3.1.2 当面の推奨案

筆者は当面は5)を推奨したいと思っている。なぜなら、この形式ならどのような入力システムでも利用可能だし、オートタガーはバイナリのフリーウェアをばらまけば良いからである。

この場合、簡易タグをどのようなものにするかが、普及度合いに大きく影響すると思われる。現在検討中の案の考え方を少し紹介する:

- 簡易タグは英語にも適用可能とするためにASCIIベースとする
- 通常の手書き方はなるべく認める
- 人に読み易く、機械に解析しやすい構造とする
- 具体的には
  - TITLE:, TO:, FROM:, DATE:等の通常良く利用するタグはそのまま認める
  - 章, 節, 項等の通常の記述形式である3.1.2等はそのまま認める
  - 箇条の開始として通常利用する「-」, 「・」等はそのまま認める。その開始の深さで箇条のレベルを判断する。
  - 空白後に開始する文はパラグラフとする。
  - 文の中に出現しても良いマークの種類を決める

以上のようなものであるが、案が出来次第ネットワーク上でばらまいて、多くの意見を集めこれを反映して、簡易タグの完成度を高めたい。できれば、業界で受け入れられるような物にまで仕立て上げたい。

更には、通信での交換形式としても、この形式を採用すると良いのではないかと考えている。

### 3.2 SGML関連規格策定の遅れ

DSSSLの規格化が遅れているために、現在はSGMLで文書を作成しても、体裁情報を含めて、標準形式で文書を交換することはできない。現状ではTeX等を整形システムとして利用しているが、日本語組版システムとしてSGMLからTeXを利用する場合には結構、苦勞も多い。

当面の開発方法はメジャーなワープロのスタイルシートを利用してSGML文をそのスタイルにマップさせてワープロで印刷することである。

SMSL/HyTimeを意識した運用環境の構築等となると、更に先の話しになってしまう。

## 3. SGML普及の兆候

### 3.1 SGMLを先導するCALs

米国国防総省の主導で始まり、現在は商務省も一緒になって推進しているCALs(Computer-Aided Acquisition and Logistic Support)はSGMLを文書標準として強力に推進している。ISOで未完の関連標準については、CALs運用のための規格を策定している。この影響で航空機、自動車、通信機、薬、半導体等の各業界が技術文書にSGMLを採用し、一方では、有価証券報告書等にもSGMLの適用が開始されている。

### 3.2 各種マニュアル/辞書/特許へのSGMLの適用

マニュアルやCD-ROM化辞書、更には特許にもSGMLの適用が始まろうとしている。今後は、関連分野での適用範囲が増大しよう。

### 3.3 SGMLの大衆化を先導するMosaic/HTML

最近のインターネットの普及の状況は爆発的で今年末には利用者が1億人にもなるといふ。この中でドキュメント参照の標準ツールとして一挙に脚光を浴びてきたのがMosaicであり、この文書の記述形式がHTML(Hyper Text Markup Language)でこれはSGMLの一記述形式である。

#### 4. 情報通有とSGML

Mosaicはインターネット上で情報通有するための主要ツールとして躍り出ようとしているが、参照のみのユーザはSGMLを意識する事はないが、文書を加工してMosaicに対応する場合には現状では何らかの形でSGMLを意識する必要がある。

一方Nifty Serve等の通常のパソコン通信でのメールのやりとりとか、電子会議室での発言等はここ当分は平文であろう。しかし、これらの文にも簡易タグによる構造化の概念を導入することは可能である。もちろん構造化した平文をSGML文に変換するのは容易であり、その逆も簡単である。通信文書を構造化することから得られる利点は文献1)に良く記述されている。

構造化文が良いといっても、いきなりSGML文書をメールとして飛び交わすには、SGMLを意識した相当に使い勝手の良い通信ソフトの提供が必要であり、これを一挙に広めるには時間が掛かると思われる。

構造化平文とSGML文を流布させるとすると、文書の倉庫としての格納媒体にはどの形式を用いるのが妥当なのであろうか？

データベース的価値が高いものは汎用的な検索言語の点から、マニュアルは版数管理等の点から、SGML文書をデータベース的な形式で格納した方が良いと思われる。

一方、個人的な利用では、平文の延長線で文書を管理すれば良いので、格納形式は簡易タグによる構造化文で良いのではないであろうか。わざわざSGML文書形式で格納することはない。

#### 5. まとめ

情報通有をスムーズに実現するには、文書の生成・加工・配付・受理・共有・検索・整理を効果的・効率的に達成することが必要で、そのためには標準化された構造化文書を通有させることが必須である。このことを良く理解しているCALISはトップダウン指向でSGMLを先導し、Mosaicは同様の理解の元にHTMLを採用、SGML文化の大衆化に寄与し始めている。この勢いを加速するためには、現在問題になっているSGML文の入力問題を解決する考えかた、簡易ツールを早期提供する一方、未完の関連規格の早期策定とこれに対応するツールの早期提供が必要である。

#### 文献

- 1) Thomas W. Malone, Kenneth R. Grant, Kum-Yew Lai, Ramana Rao, David Rosenblitt: MIT  
「Semistructured Messages Are Surprisingly Useful for Computer-Supported Coordination」 ACM Transactions on Office Information System  
Vol. 5, No. 2., April 1987
- 2) 吉岡誠編『SGMLのススメ』オーム社 1993年
- 3) SGML懇談会 実用化WG『実践SGML』規格協会 1992年