

SGML 適用による電子化の実践  
—SGML システムの構築及び文書の SGML 化における諸問題—

笠原 健成

kasahara@soft.hitachi.co.jp

株式会社日立製作所 ソフトウェア開発本部

SGML(Standard Generalized Markup Language)は、文書データの交換や多角的な利用のためのプラットフォームとして注目されている。SGML を適用するということは、文書データや作業方法を標準化することを意味する。本稿では、SGML システムの構築及び文書の SGML 化の過程で発生する諸問題や留意すべき事項について、製品マニュアルを開発する観点から述べる。

Building SGML Documents:  
Problems and Solutions When Building SGML Systems and  
Applying SGML to Documents

Takenari Kasahara

kasahara@soft.hitachi.co.jp

Hitachi, Ltd. Software Development Center

Standard Generalized Markup Language (SGML) is a popular platform used to exchange and diversify document data. To apply SGML, work procedures and document data must be standardized. This paper describes problems which occur when building SGML systems and applying SGML to manual development.

## 1. はじめに

SGML は、文書データの交換、複数メディアへの出力などに対応するためのプラットフォームとして注目されている。この前提となる、いかに効率良く文書を SGML 化するか、また SGML 化されていない既存の文書をどのように文書 DB に取り込むかといった問題は、システム全体の成否に大きく関わる。文書を適切かつ体系的に DB 化できるかどうかは、SGML システム自体の可用性を決定するといっても言い過ぎではない。しかし、このような文書の DB 化にはクリアしなければならない様々な問題がある。

本稿では、SGML システムの構築及び文書の SGML 化の途上で発生する諸問題や留意すべき事項について概観する。

## 2. 適用対象による要件の相違

SGML を適用する文書を次の観点で分類した場合に、システムに要求される事項を大まかにまとめる。

### 2.1 文書の用途及び流通の範囲

社内でのみ流通する文書と、製品化又は公的な配布を目的とした文書(及びその中間生産物の文書)とに大別できる。

#### (1) 社内文書

社内で運用されている帳票類や、業務上の手順・手引書、製品開発のための仕様書などがこれにあたる。これらの文書は、出力媒体や文書交換先も事前に確定することが比較的容易である。また、業務遂行の円滑化が目的であり、その妨げとならない範囲で各メディアへの出力品質の巧拙はさほど問われない。

#### (2) 製品・公的配布文書

製品マニュアルや学術的な刊行物などがこれにあたる。これらは、将来的に発生するであろう様々なメディアへの出力や文書交換の

ニーズに対応できるように、システム構築、SGML 化の段階での配慮が必要である。さらに、各メディアへの出力品質はある程度水準が要求される。また、社内文書である製品開発のための仕様書などは、特にマニュアルでは開発の素材となるため、社内文書との連携・流用の枠組みについても検討する必要がある。

### 2.2 新規作成文書と既存文書

SGML システム構築後に新規作成する文書を SGML 化する場合と、既存の非 SGML 文書をシステムに取り込む場合とでは作業上の留意点が異なる。

#### (1) 新規作成文書

文書構造(DTD)を先に定義した上で文書を作成できるため、SGML 化は比較的容易である。また、SGML システムを適用する文書種別を新規に追加する場合でも、システム側で許容できる文書構造の範囲は既に明確なため、両者のすり合わせはしやすい。さらに、出力先のメディアを考慮した文書作成の方法を事前に規定することもできる。

ただし、定義した構造に合わせて文書を作成してくれるように作業者に徹底するには、システムとしての仕掛け作りや事前の P.R.・教育が必要である。

#### (2) 既存文書

対象となる文書に関する綿密な構造分析が必要である。その際には、対象範囲に漏れがないようサンプルを大量かつ広範囲に確保しなければならない。また、SGML 化の際にはシステムに取り込みが困難な構造や記述の修正が多量に発生する可能性がある。

### 2.3 本稿が扱う範囲

本稿では、特に製品化が予定されている文書で、かつベースとなる文書が存在する場合、

具体的には製品マニュアルに焦点を絞って記述する。また、SGML 化の途上で、対象の文書と密接に関わる社内文書や開発中に派生する中間文書についても多少言及する。この範囲外の文書については対象としていないが、本稿の内容のかなりの部分はそれらの文書にも適用できると考える。

### 3. SGML 化作業の各過程での問題

SGML システムの構築、及びそれに続く文書の SGML 化の過程を次のように分け、それぞれについて考慮すべき事項をまとめる。

- 文書構造の分析
- DTD の設計
- ツール類の整備
- 文書 DB の構築
- SGML 化実作業

#### 3.1 文書構造の分析

##### (1) 広範囲のサンプル収集

分析の対象とするサンプルは多く、また偏りなく収集する必要がある。マニュアルで考えれば、概説書、手引書、操作書、リファレンスなどでは、それぞれ出現する構造や要素はかなり異なる。各形態の文書を縦断的かつ横断的に分析し、共通に見いだせる構造・要素やそれぞれに固有なものを明らかにしなければならない。

##### (2) 例外要素の明確化

収集した構造・要素が対象文書の大半で出現するのか、それともごく一部だけに例外的に出現するのかを充分見極める必要がある。例外的な構造・要素も漏れなく DTD に取り込もうとした場合、文書構造が巨大かつ煩雑なものとなり、SGML 化後も処理に不具合が生じる可能性がある。

##### (3) 表現と意味の区別

出力体裁や文字修飾が同じであっても、そ

れらが文書内で同じ意味合いで使用されているとは限らない。例えば、同じ太字で出力されている字句が、一方は単なる強調表現であり一方は巻末の用語解説で説明される言葉を示しているといった場合がある。表現形式にとらわれず、それが示す意味を抽出する必要がある。

##### (4) 他言語化の考慮

本稿の主旨からは多少外れるかもしれないが、サンプルの文書が他言語に翻訳されている場合には、翻訳済みの文書も併せて分析することも必要かもしれない。同じ内容を記述していても、言語が異なれば出現する構造や要素が微妙に異なるはずである。そのような相違は各言語に最適な記述方法を反映しているが、構造の分析及び DTD の設計で極力共用できるようにすり合わせをしておくことが、SGML システム適用後に発生する文書の他言語化には効率的である。

#### 3.2 DTD の設計

##### (1) 交換先・変換先の考慮

SGML 化した文書は、コンバートによって他 DTD に対応することになるであろう。想定されるデータ交換先の DTD や各業界で広く流通している DTD は前もって十分に分析し、それらへの変換の可能性について検証する必要がある。

##### (2) 広範囲への適用の指向

マニュアルのような文書は、ターゲット製品の開発者によって作成される機能仕様書などが開発のベースとなる。また、マニュアル作成の初期段階では企画書や設計書などを作成するはずである。後工程での作業を効率良く行うためには、そのような開発の素材となる文書も極力一元的に管理できる DTD としての枠組みが重要であろう。

また、マニュアルが適用される製品機種や

バージョンに応じて書き分けが必要な記述は DTD に要素を設け、極力一つの SGML データで対応できるようにするなどの工夫が必要である。

### (3) タグセットの使い分けの検討

文書構造の分析の際に抽出された構造・要素をどこまで DTD として実装するかは充分検討しなければならない。すべてを実装しようとしたり、例外的な要素までも含めてしまうのは SGML システム自体の使い勝手を低下させることになりかねない。文書のタイプに応じて DTD を複数用意して使い分けることや、既存の非 SGML 文書の取り込み専用のタグセットを明確にすることも検討すべきである。本件については、SGML システムの実運用の際にも、教育などによる徹底が必要である。

## 3.3 ツール類の整備

### (1) データ変換

SGML 化された文書を紙、CD-ROM などのメディアに出力するには、文書データのコンバートが必要となる。各メディアに対応したオーサリングツールはこの機能を包含しているものもあるが、HTML や他 DTD などへの変換は独自にコンバータ又はコンバートのスクリプトを用意しなければならない。コンバートの精度が完璧でない場合には、後工程での作業内容やその手順を明確にして手作業を少しでも減らすことが重要である。

### (2) タグ付け作業の支援

タグ付けの作業は、ワープロソフトや DTP による WYSIWYG な操作に慣れた作業者にとってかなり負担になる。この作業をどのように支援できるかによってシステム全体の効率はかなり左右される。

#### ● 文書の新規作成と改訂

ある程度まとまった量の文書を執筆する新

規作成の段階と、それを改訂する作業とで使用するツールを使い分けることが検討できる。例えば、新規作成ではワープロソフトなどで編集したデータをコンバートによって SGML 化するという仕組みが必要であろう。市販のワープロソフトや DTP では、スタイル属性を規約化された独自の形式のテキストファイルとして出力する機能を持つものもある。これを媒介とすれば、コンバートスクリプトの作成によって SGML データを生成できる。このような方法が利用できれば、SGML や DTD について知らない作業者が作成した文書もシステムに取り込みやすくなる。また、改訂作業は、データの保全、作業の容易さの観点から SGML エディタを導入することが現実的である。

#### ● ハイパーリンクの整合性の確保

参照元と参照先との整合性を管理する作業は、SGML 文書を扱う上で特にミスが発生しやすい部分である。このような整合性の確保は文書作成者自信が手作業で行うのではなく、タグ付けツールの機能として盛り込むべきである。

#### ● 出力結果の確認

SGML データの正確さの確認及び各メディアへの出力結果を簡易的に確認できることは WYSIWYG に慣れた作業者にとって重要である。

### (3) 開発ツール群の統合

各メディアへの出力やデータコンバートなど、作業者が扱うツールは、SGML 化によって増加する。また、扱うデータも SGML テキストやグラフィックデータなど様々である。これらのツールやデータを統合・管理するには、文書 DB と連携した作業上の統合環境が必要である。これには次のような機能が要求される。

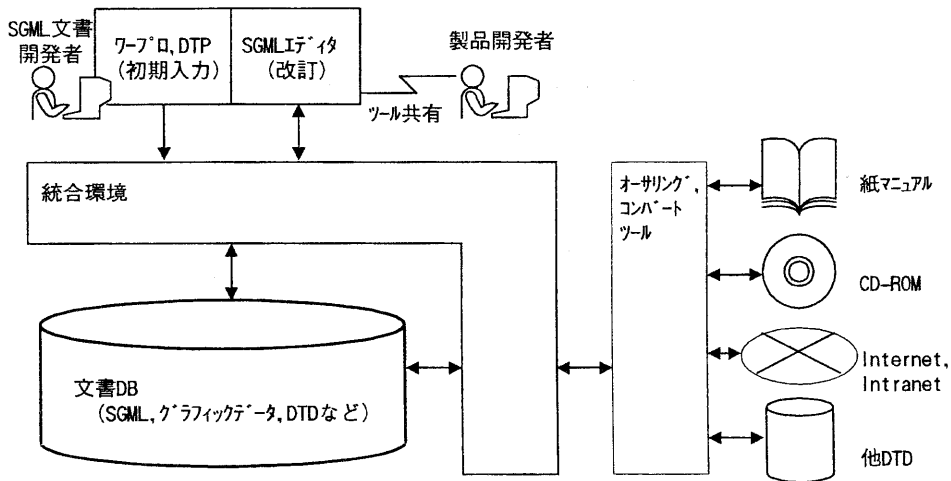


Fig. 1 SGML ツールの構成例

- 文書データの構成の管理(文書内, 文書間)
- 各ツールの制御
- 文書 DB へのアクセス制御
- ワークフロー管理

このような観点でまとめた SGML ツールの構成例を Fig. 1 に示す。

### 3.4 文書 DB の構築

SGML データを管理する DB には次に示す機能が必須と考える。

- データへのアクセス制御
- バージョン管理
- 差分管理
- 世代回復
- SGML データの特定の構造・要素をキーとした情報検索・取り出し機能
- 関連するデータ間(例えば, SGML テキストとそこから派生するデータ)の整合性の管理
- SGML データと DTD, 使用ツールとの関連付けの管理
- SGML データのコンポーネント化への対応

- オンライン出荷・配布との連携

また, 複数メディアへの対応を念頭においていた場合, 元となる SGML テキストやグラフィックデータに加えてそこから派生するデータなど, 管理しなければならないデータの総容量は非常に大きくなる。保持すべきデータを厳選するとともに, データ圧縮や差分による世代管理についても検討すべきである。

### 3.5 SGML 化実作業

実作業の開始に先だって, 作業方法を手順書としてまとめる必要がある。そこでは, DTD やタグの解説やツール類の使用方法に加えて, 既存の非 SGML 文書の取り込み方法について明確に規定しなければならない。非 SGML 文書のタイプ別にこの作業の流れを模式的にまとめたのが Fig. 2 である。ここでは, このような既存文書を SGML 化する際の留意事項についてまとめる。

#### (1) 電子データ化

タイプ印刷によるマニュアルなどは, 次の

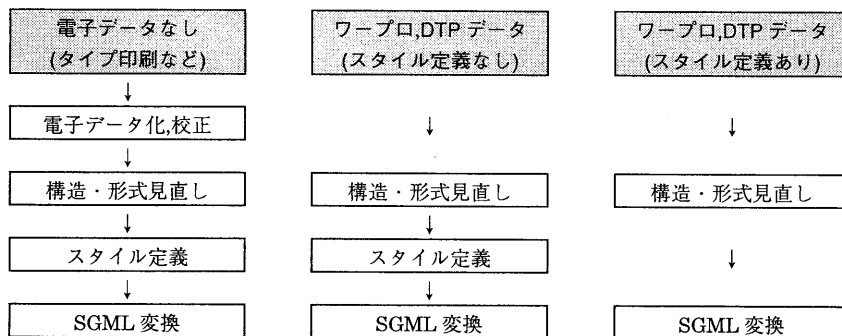


Fig. 2 既存文書のSGML化の流れ

どちらかの方法で電子データを作成することになる。

● 文字認識ソフトウェアの使用

スキャナ入力した原稿を認識してテキストファイルとして出力するものが一般的である。このようなソフトウェアは平均の認識率はかなり高いが、使用に際してはいくつか注意すべき点がある。例えば、図中の文字が本文と同列に認識されることを防ぐため、あらかじめ図の部分のマスキングしたり、表は認識後にセルの配列を修正したりといった作業が必要である。文書作成者が定義した外字も手作業ですべてチェックアウトする必要がある。また、何度も改訂を重ねて文字が不均一になっている文書は認識率が低下する。どのような条件で最良の認識率が得られるかを事前に充分調査する必要がある。

● オペレータによる入力

上級のオペレータであれば、入力後の文字校正も含めてA4版1ページあたり約15分で作業可能であるという。しかし、これはテキスト部分だけの話であり、図についてはスキャナ入力を実施するか又は新規にグラフィックソフトで作成しなければならない。

(2) 構造の見直し

SGMLでは、章、節、項といった構造に従って文書を作成することが要求される。外部とのデータ交換や他国語化が想定される場合には、特にこの点は重要である。しかし、既存のマニュアルは必ずしもそのようなDTDで許容される文書構造に合致しているとは限らない。構造を省略していたり、許容しないほど深いネストが使用されていたりする可能性がある。このような場合には、単なるリライト作業では対応できないことが多い。対象製品の仕様を熟知した者が作業するか、ベースマニュアルの作成者と共同で作業することが必要となる。

(3) 形式の見直し

最初から紙というメディアだけをターゲットとした文書やWYSIWYGな操作環境で作成された既存の文書は、そのままではSGML化が難しい場合がある。そのような記述は見直しや修正が必要である。想定される出力先の各メディア特性を充分考慮して極力共有化を図るのがよいと考える。

● 記号、外字

SGMLでは使用する記号は通常SGML宣言で定義する。ISO[1]の規格に従うとか、JIS[2]の規格に従うというようにその範囲を明確にしなければならない。範囲外の記

号や個別に作成された外字などを使用して  
いる部分は修正が必要である。別の記号に  
置き換えるか、文章による説明で言い換え  
るかをそれぞれの事例ごとに決定する必要  
がある。また、上付き・下付き文字や文字  
修飾などにも同様の注意が必要である。

● 表組

セルの関連が複雑な表は、それ自体 SGML  
システムで処理しにくい。ツールや出力先  
のメディア側が対応できるかだけでなく、  
タグ付け方法も非常に煩雑である。既存の  
表を見直すだけでなく、SGML 化を前提と  
した表の作成基準の決定が必要である。次  
に表の修正の例を一つだけ示す。

[修正前]

タイプ \ パラメタ	-p	-a	-x
High			
Normal			
Low			

[修正例]

タイプ	パラメタ		
	-p	-a	-x
High			
Normal			
Low			

Fig. 3 表組の修正例

● 数式、コマンドシンタックス

SGML では、複数行にわたる記述を実現す  
ることが難しい。分数を含んだ数式や次に  
示すようなコマンドシンタックスがそれに  
該当する。

```
RHOST [ HOST(ホスト名称)
        [, USER(ユーザ名称)]
        [, DELETE]
        LIST ]
```

Fig. 4 コマンドシンタックスの記述例

このような記述はグラフィックデータで表  
現することも考えられるが、電子マニュアル  
化した際に、

- ・検索機能での検出ができなくなる
- ・表示速度が低下する

といった問題が生じる。また、TeX による  
処理も可能であるが、そのためにはこれら  
の記述をすべてタグとして定義しなければ  
ならない。タグ付け方法は非常に複雑にな  
るであろう。コマンドシンタックスなどは、  
UNIX で用いられている「[, ], |」を使用  
した書き換えが可能かもしれない。

4. 体制作り

話しは前後するが、SGML 化を推進する上  
では、これまで述べてきた事項以外に必ず解  
決しなければならない問題がある。

(1) コンセンサスを得ること

SGML の導入は、作業の流れやアウトプ  
ットを大きく変えることになる。ワープロや  
DTP といった WYSIWYG での編集に慣れた  
作業者は、タグ付きのデータを扱うことに不  
信感を抱くかもしれない。また、紙メディア  
に特化した DTP や TeX によるきめ細やかな  
レイアウトに慣れている人々にとって、今ま  
で実現できていた出力体裁がなぜ SGML では  
できないのか、といった不満を感じるかもしれ  
ない。関係者には SGML の性質を事前に充  
分説明して理解を得ておかなければならない。  
恐らくこの作業は、予想以上に困難を伴うで  
あろう。複数メディアへの対応、文書交換手  
順の円滑化、SGML 化後の工期短縮をメリッ  
トとして挙げて P.R.すべきである。

(2) 人材の確保

SGML システムの構築、運用には、従来の  
マニュアル開発作業に要求されるのとは別の  
知識、能力が要求される。例えば DTD の設計  
一つとっても、ある程度のプログラミングの

知識は必要であろう。また、文書 DB の構築ではシステム設計のノウハウが要求される。このような人材を確保又は育成することはシステムの安定稼働にとって必須条件である。

### (3) 文書作法の変化

様々なメディアへの対応を目的とした SGML には、従来の紙ベースで培われてきた文書作法技術はそのまま適用できないかもしれない。オンラインヘルプの開発に従事している方は既に感じていることだろうが、メディアが異なれば要求される最適な表現技術は異なるはずである。一元化されたデータで複数メディアへの出力に対応する SGML に最適な文書作法についての研究や知見の蓄積が必要であると考ええる。

## 5. 終わりに

本稿では、SGML システムの構築及び文書の SGML 化における諸問題や留意すべき事項について述べた。SGML を導入することは、データやその出力形態を標準化するだけでなく、仕事の流れも含めて標準化することである。個々の実状やニーズに合わせて開発環境や手順を見直し・再構築し、諸問題に対して柔軟に対処していくことが重要であると考ええる。

## 6. 参考文献

- [1]ISO/IEC 8879,"Standard Generalized Markup Language (SGML)",15/Oct/1995.
- [2]日本規格協会,"JIS ハンドブック 情報処理用語・データコード編",1994.