

SGMLの動向と応用

池田 実

東芝アドバンスシステム株式会社

SGML担当

1. はじめに

本稿では新たな動きが見えてきたSGMLについて、主要なSGMLの会議1つであるSGML '96を通してその動向を報告する。特に、その中からこれからのSGMLの利用にとって重要と思われるマイクロドキュメントとXMLを取り上げ概説する。

これまでSGMLは、ドキュメントの電子的な表現あるいは交換の標準としてその重要性が認識されてきたが、最近ではSGMLの構造的な活用したドキュメントとデータ処理システムを統合するような高度なアプリケーションも生まれてきている。最後にこれらの典型的な高度な応用システムの1つの事例を紹介する。

2. SGML '96

これまでSGMLの推進や普及に努めてきた団体の一つにGCA (Graphic Communication Association) がある。このGCAが主催する会議としてはSGML Asia-Pacific ('96はシドニ)、SGML Europe ('96はミュンヘン)、また世界レベルの会議としてSGMLxx (xxは開催年次)があり、それぞれ年に1回開催されている。昨年のSGML '96は第10回目の会議とういこともあり、ある種の節目を感じさせる内容となった。本稿では、この会議を通じて著者が感じたSGMLの動向を報告する。

2.1 会議概要

SGML '96は、ボストンで11月18日から5日間開催された。会議の参加人数は1400名で、うち日本からの参加は展示会のみ参加者を含めて30名弱(推定)であった。ちなみにSGML '88の参加者は、たった52名だったそうである。

2.2 会議内容

本会議の参加者の大勢(80%~90%)はSGMLのユーザ企業や団体に属する人たちで、印刷・出版や製造業を初め、その他の企業、軍、政府関係などあらゆるジャンルにわたる。日本におけるこの種の会議や展示では圧倒的にベンダー中心になるのに対して、欧米ではユーザが主体となりSGMLの利用を進めている点が根本的に異なる。SGMLは、利用者側の参加なしには成功できないものであることを考えると、この相違には大きな意味があるように思える。

また、昨年まで日本からの会議への参加者は数人であったが、昨今のCALSブームを受けてか、本年は30名ほどと急増した。その構成は、殆どがツールやサービスなどのベンダーであった。

会議では、Newcomer、User、Expert、Business Mgt、Case Studyの5つのトラックが平行に持たれた。この種の会議ではおなじみの深夜におよぶ討議や昼食時のBOF (Births Of a Feather)テーブルでの情報交換も活発に行われた。その他にペーパーセッション

ンが100ほどあった。会議での発表件数は80編ほどで、日本からの発表も3件ほどあった。参加者、セッションや出展数のいずれをとっても過去最大のSGML会議となった。

2.3 ツール展示

約47社の展示があった。特に目新しい製品はなかったが、予定を含め日本語を扱えるツールが増えてきた。ArborText、Frame+SGML、Near&Far Author、Grif、Baliseなどが新たに日本語に対応する(含予定)。

また、多少新しい傾向としては、製品の展示以外にサービス、書籍、教育エンジニアリングなどベンダーの種類に幅がでてきたことがあげられる。

3. SGMLの動向

3.1 全体的傾向

欧米を中心にして既にSGMLが広く実用に供されているためか、全体に落ち着いた雰囲気、より現実的な話題が多くなった。その一方で、SGMLは伝統的な出版ないしは情報交換に対する応用から、より広い領域への応用と高度化が試みられるようになったことが、強く印象に残った。また、今回の会議が今後のSGMLのあり方を左右するターニングポイントになるのではないかとの印象も受けた。

当然ながらWWW(HTML)の普及などが、SGMLにとっても刺激となっていることは紛れのない事実であり、これまでどちらかと言うと完全性を目指していたSGMLが、より使いやすいSGMLを目指す方向に転換する、あるいは分派して行く可能性を感じた。

4. 技術動向

今回の会議で最もホットな話題は、マイクロドキュメントとXMLであろう。次にこの2つを取り上げ概説する。また、その準備としてHTMLとSGMLを概説し比較する。

4.1 XML(eXtensible Markup Language)の位置づけ

このXMLは、その名称からはHTMLの拡張のように感じられるが、実際にはむしろHTMLよりSGMLにはるかに近い仕様案である。

XMLの発表はHTMLの規格を制定しているW3C(World Wide Web Consortium)内のグループからの提案であるが、検討メンバーの多くはSGMLコミュニティーの人間で、その点から言えば当然の結果かもしれない。

XMLが現時点で全面的な賛同を得ているとは言えないが、今回の会議でそのドラフト仕様が公開され、基本的にSGMLのサブセットであることが確認された。その結果、HTMLのようにSGMLと別種の規格案ではないことが明確になり、急速に関心と同意を集めている。しかし、このXMLが事実上の標準となり得るどうかは判明するまでにはもう少し時間を要するであろう。

4.2 HTML概説

HTMLは、WWW上のプロトコルの一種で、タグがブラウザ上でのドキュメントの表示形式を決定する。HTMLドキュメントの作成者は、ドキュメントのタイトルなどのドキュメント内の

各要素に対して、予め決められているタグのセットのなかから特定のタグを選択して指定する。この場合のタグは、ハイパーテキストリンクなどを除き、画面上の表示形式を決める働きしか持たない。

また、HTMLのタグの表記と意味は規格化されており、ユーザが自分の都合で変更したり、あらたに追加したりすることはできない。このタグの仕様は、W3Cの責任で改訂されているが、実際にはメーカ毎の方言が存在したり、頻繁にバージョンが上がったりして長期的な観点から見ると安定性に欠ける。結果として、現在のワープロにおける互換性の欠如にも似た状況を招きかねない。つまり、HTMLドキュメントを長期にわたって保存したり、利用するときの媒体として採用するのは危険だと言える。

4.3 SGML概説

SGMLの特徴は、ドキュメント中で利用するタグの種類や相互の関係を、ユーザが定義できる点にある。SGMLではデータベースの設計に似て、予めユーザがその構造を設計し、規格で定められた言語を利用して記述しなければならない。これをDTD(document Type Definition)と呼ぶ。ここにSGMLの長所と欠点が併存する。

つまり、HTMLでは、タグの種類や構造(相互関係)は予め他から与えられるため、ユーザは自分の仕様や要求にそった構造化ができない代わりに、設計という面倒な準備作業が不要である。それに対してSGMLでは、絶対に設計が不可欠となるため、実作業は複雑になりがちである。逆にその難解さを補って余りあるのが、論理表現の高さである。

現在SGMLの応用が進んでいる分野として飛行機業界と自動車業界が良く知られているが、その両者のドキュメントのDTDは全く別の構成を取る。その相違は、単純な見かけや記述内容の相違が原因する以外に、仮に同じ整備用のドキュメントであっても、対象物の性質や利用環境の違いなどを反映した結果であって、合理性を追求すればするほど独自の物となるのが普通である。

このように、SGMLでは記述内容や用途、ドキュメント設計者の意図などをDTDを通してドキュメント構成に反映することができる。この点はデータベースの設計と応用の関係に極めて似ている。これに関しては最後の例の中で説明する。

4.4 XML概説

XMLは、SGML中の機能から難解でかつまれにしか利用されない部分を削除して、SGMLをより簡便に利用できるようにしようとするものと見ることができる。

SGMLの規格には、DTD設計のための言語規格以外に、その言語仕様を変更するメタな仕様が存在する。これをSGML宣言(SGML Declaration)と称する。これを使うと通常であれば『<』と『>』でタグを表記するところを、全く別の文字で表わすようなことができる。このようなメタ機能は強力である反面、複雑すぎて混乱を招く危険がある。例えば、タグを別な文字に置き換えたSGMLドキュメントは、それを見ただけでは何がタグであるのかさえも解らないことになる。その他にSGML宣言では、タグの省略や文字コードの置き換えやドキュメント中に出現できる各種の名前のネーミングルールの指定などがサポートされている。

XMLがSGMLと基本的に異なる点は、このSGML宣言を固定にしてしまったところにある。つまり、ユーザはSGML宣言を行うことができず、予めXML規格で定められた宣言に

従うことになる。

これは、HTMLとSGMLの差違にも通じる点であるが、実質的な制約はずっと少ない。図1はHTMLとSGMLおよびXMLの関係を模式的に示した物である。

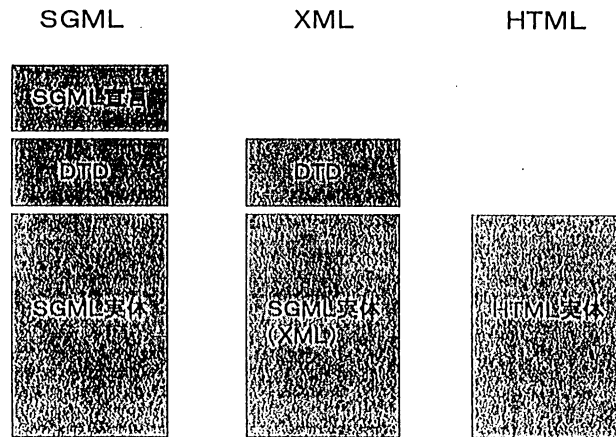


図1. SGML、XMLとHTMLの関係

5. マイクロドキュメント

SGMLでは、DTDは対象とするドキュメント全体を規定する物として存在する。従って簡単なDTDであれば、ドキュメント全体を容易に把握できるので問題はないが、現実に使われているDTDは相当に複雑で、SGMLの入力や編集は容易でなく、利用者に大きな負担を強いることになる。実はこの点が、SGML導入の最大の障壁となっている。

現在、半導体業界 (Pinacles) や自動車業界 (J2008) あるいは航空業界 (ATA2100) さらには薬品業界などでドキュメント情報の交換標準としてDTDが規定されつつあるが、いずれも100とか200とかのエレメントを構成要素とする複雑なDTDである。

古くから利用されてきたデータベースのアプリケーションでは、ユーザやオペレータが一度に数100項目も入力したり変更することなどまずあり得ない。つまりデータベースでは、データベース全体の構造とアプリケーションから見えるデータ構造とが分離できるようになっている。これによって、ユーザはその時点で必要な情報要素だけを相手にすれば済み、見通しが良く簡便な利用が可能となる。

残念ながら、現在のSGMLはデータモデル的に見て、上記のようなデータベースのビューに該当する概念が存在しない。

SGMLが紙ベースのドキュメントから離れ、コンピュータ上のダイナミックな情報資源として利用されるにつれ、この欠点は大きな問題となってくる。例えば大量のドキュメントを大きなデータベースに貯え、検索サービスを行ったときに、ユーザが最終的に必要とするのは、該当の巻号のドキュメント全体ではなく、その中の特定の内容記述である。このようにユーザが求める情報のドメインは場合に応じて自在にに変化するのが本来であり、固定的なドキュメントの記憶構造ではこうした要求に対応できない。つまり、元来のHTMLやSGMLあるいはXMLを素朴にファイル上に実現しただけでは、こうした要求に応えられない。

マイクロドキュメントは、前述のような要求に応えるために考えられた物で、SGML(あるいはXML)を論理的、あるいは利用上の小区画に分割し記録・管理する。これにより、ドキュメントをドキュメント全体でも、あるいは分割した小区画単位にでも自在に検索したり、検索結果を再構成したりする機能が提供できる。次の図は、マイクロドキュメントと従来のSGMLファイルを比較した物である。

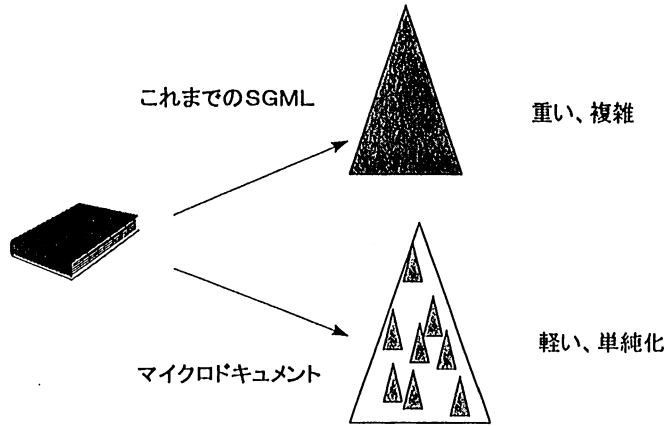


図2. SGMLとマイクロドキュメントの関係

次にマイクロドキュメントを提唱しているOmniMark社のデモシステムを例にマイクロドキュメントの実現例を見てみる。このデモシステムは、薬品情報の提供を想定して構成されている。

薬品情報は1つのドキュメントが特定の薬品の情報を表わしている一方で、その内部には論理的に独立した複数の小さなドキュメントが含まれている。マイクロドキュメントではこの小部分(マイクロドキュメント)をそれぞれ独立したDTDで記述する。

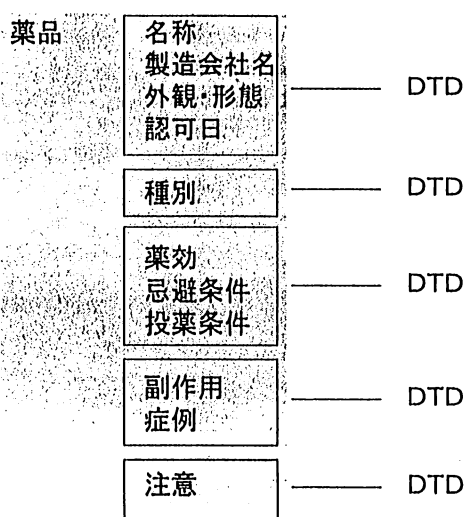


図3. マイクロドキュメント

このシステムを使って薬品の検索を行うと、『XXと組み合わせるといけない薬を見つけ、その副作用と症例だけを表示する』ようなことが可能となる。つまり元のドキュメント全体を単位とする操作に加え、各DTD単位のドキュメントを要素とする操作や、検索結果を組み合わせると元のドキュメントとは構成の異なるドキュメントを動的に構成するようなダイナミックな処理が可能となる。

しかしながら、元のドキュメントとマイクロドキュメントの集合として表わされるドキュメントが論理的に同一であることを証明したり、検証する手段がまだ存在しない点に問題を残すが、実用の観点から言えば、マイクロドキュメントがSGMLの利用に大きな進展をもたらす可能性があると言えよう。

ドキュメントの利用形態は、これまでの出版のように発行側から一方的に受け手に流れる単純なモデルでなく、インタラクティブな利用が急増すると見られている。このマイクロドキュメントの登場は、その傾向を一層加速するものと思われる。また、これによりSGMLとインターネットの連携も一層容易になった。

6. SGML利用の最前線

最後にSGMLを利用したドキュメントシステムと既存の情報処理系が連携する応用例を取り上げる。

これまでのドキュメントは、観念的には情報の一種として捕らえられていても、現実には他の情報系とは独立で、連携することはほとんどあり得なかった。現在ではドキュメントはその作成過程でワープロ等で電子化されるのが一般的で、そのために特別な費用が発生するわけではない。このような状況の下で、ドキュメントを本来の情報として活用し、情報の共有や活用を図ろうとするのは自然な流れである。

SGMLはそうした目的には最適な機能を提供する。つまり、タグを用いて整然と構造化されたドキュメントは、整形して印刷すれば従来のドキュメントその物だが、電子的に保存され

たドキュメント自体はコンピュータがハンドリング可能な電子情報でもあるからである。つまり、SGMLはドキュメントに『電子情報』の性質と『いつでも印刷や表示可能なドキュメント』の両方の性質を与えることができる。

次の例はそうした未来指向のドキュメントシステムの例である。次の図で、左上は電子ドキュメントの世界で、修理マニュアルがSGMLで書かれている。また、同時にこのマニュアルには該当の修理を行うために必要な道具や交換部品あるいは必要な技能資格などが記されている。また、それら必要な資源には特定のタグが付され、検索し記述内容を自在に取り出せるようになっている。

図中の番号は実際の修理開始に至るまでの手順で、1. は修理マニュアル、2. は、上記の資源を管理している基幹系のリアルタイムコンピュータシステムを表わし、3. は結果として発行される修理の手配書を表わしている。

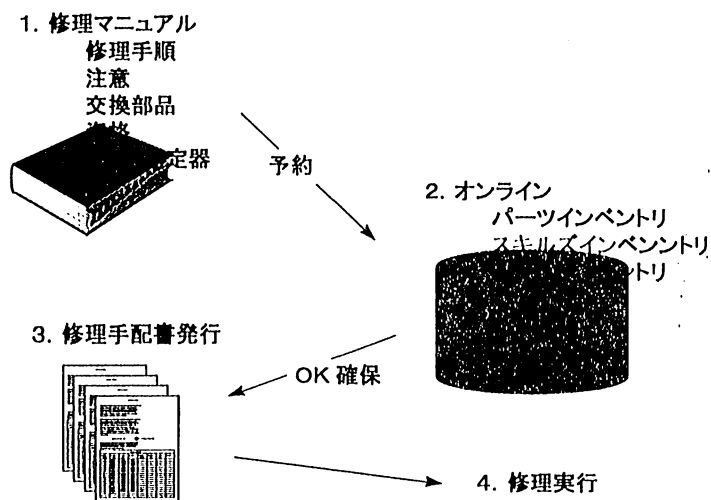


図4. SGMLと異種情報の連携

この例では、1のSGMLドキュメントを技術者が検索して修理すべき個所を特定し、その場所を2のオンラインシステムにネットワークを通じて知らせる。2のインベントリシステムは、SGMLドキュメントから予め決められたタグを使ってその内容を取り出す。その内容に基づいて実際の予約を行い、完了後、3の手配書を発行する。この手配書には、1中の手順などの記述内容と、2で割り当てられた資源の所在場所などが組み込まれ、動的に発行される。

この一連の作業は、銀行のオンラインシステムなどと同様にネットワークを通じてリアルタイムに処理される。つまり、この例ではSGMLドキュメントと情報処理系がネットワークを通じて情報連携しているわけである。

7. 最後に

率直に言ってSGMLの導入は簡単でない、が、同時に新しい世界を築くインフラともなりうるもので、『静かな革命』をもたらす可能性を持つ。ことに間接部門やオフィスワークの生産性の悪さが、決定的なウィークポイントになりつつある日本の企業にとってSGMLは、欧米以上に重要であるのかもしれない。

参考文献

- (1) Extensible Markup Language(XML) Version 1.0, World Wide Web Consortium, Working Draft, November 1996
- (2) Standard Generalized markup Language (SGML) ISO 8879
- (3) SGML '96 Conference Proceedings