

文書管理システム

まえがき

文書管理とは、一般的には組織における大量の文書を効率的に保管し、必要なときに取り出して有効活用、あるいは不要な文書を正しく廃棄するなどのための手法であって、これまでも様々な取り組みがなされてきた。ファイリング部材と書架に対する工夫や、作成される文書書式への標準化の取り組みとラベリング規約、あるいは図書館に見られる書誌目録カードの管理などがそれにあたるが、それらの多くが紙または冊子をベースにした原本(性)重視の保管と管理であった。また、単にファイリングという観点からは、マイクロフィルムや光ファイルといった大容量ファイルシステムによる省スペース化と取り出しの高速化がはかられてきたが、いずれも専用の設備を必要とし、高価であるとともに他システムとの広い互換性を欠くことから、文書管理のための標準的手法とされるには至っていないというのが現状であろう。

一方で、コンピュータ技術の著しい発展に伴う広範なオープンアーキテクチャの出現と、ネットワークを始めとした単位情報あたり的大幅なコスト低減がもたらしたものは、コンピュータネットワークによるマルチメディア(文字・数値・映像・音声・印刷・磁気記録・光学記録・有線・無線など)の融合とそれらへのデジタル処理技術の進歩、そして官庁や企業組織から個人に至るまでの迅速でシームレスな情報伝達の実現であろう。また、旧来のコンピュータが得意であったコードと数値による大量一括処理から、情報個々の処理へとそれらのリアルタイムでの取り扱いの機会が広がっていく中で、文書その情報の一つひとつになぞらえる事により、文書管理の世界も大きくその姿を変貌しつつある。知識が情報に、情報から知識へ、さらに文書になっていない暗黙知(個人や現場がもつ暗黙的な知識)を形式知(文書などに表現された情報)へ変換し、共有することによってさらに新しい知識の発掘が行われる。

こうした流れは単に効率的な文書の保管と管理にとどまらず、文書の中にある情報の再利用、再々利用を容易にするとともに「データマイニング」の手法としても大いに論じられるようになった。さらに先進的企業ではこういった文書管理システムを「ナレッジ・マネジメント・システム」という位置付けのもと、知的経営の一環として採用する動きもある。そしてこれらのベースには、テキスト処理技術、画像処理技術、データ圧縮技術はもとより、SGML (Standard Generalized Mark-up Language)やXML、メタデータ記述手法、全文検索技術などを背景に、それら要素技術の標準化が幅広く推し進められていることが大きな前提となるであろう。このような技術革新と時代の流れを睨みつつ、本書では電子情報による次世代の文書管理システムについて、その動向と可能性を探って見ることとした。

目次

まえがき	I
目次	II
1. はじめに.....	1
1.1. SGML を取り巻く現状	1
1.2. XML の出現とそのインパクト	2
1.3. 全文検索技術.....	4
モデルケース：新聞社A社.....	4
モデルケース：博物館・図書館	5
1.4. CALS での取り組み	5
(1)宇宙 CALS	5
(2)航空機 CALS.....	5
(3)自動車 CALS.....	5
(4)鉄鋼設備 CALS.....	6
(5)電子機器・部品 CALS	6
(6)プラント CALS.....	6
(7)ソフトウェア CALS.....	6
(8)船舶 CALS	6
(9)電力 CALS	6

1. はじめに

ネットワークインフラとマルチメディア技術の急速な発展により、コンピュータで取り扱う事の出来る情報は従来の数値・統計型から、テキスト・音声・画像・動画と著しい進歩をとげた。中でも、インターネットを介したオープンプラットフォームでの情報交換は、世界中の公開サーバーのあらゆる情報へ、我々を瞬時に導いてくれる。あるものは華麗な GUI を展開し、あるものはサウンド付きのダイナミックな情報を与えてくれる。端末機器となるパソコンはここ数年で飛躍的な高性能化と低価格化、標準化が進み、企業でも家庭でもずいぶんと手に入れやすくなった。

一方で 1994 年に閣議決定された「行政情報化推進基本計画」に基づき、政府は行政情報の電子化を進めてきた。SGML の技術に着目した白書等の電子化と公開、あるいは霞ヶ関 WAN での各省庁の行政情報クリアリング(所在案内)システムと省庁間の電子文書交換システムの整備、そして総務庁が中心となって推進する「総合的な文書管理システム」の整備などがそれにあたる。また内閣総理大臣の発案による「ミレニアム(千年紀)プロジェクト」は、ワンストップ・サービスをも含めた「スーパー電子政府」の実現を目指す。1999 年春には「情報公開法」も成立し、まさに文書管理システムのメガトレンドはいよいよ大きく動き始めた。

1.1. SGML を取り巻く現状

コンピュータによる情報共有の技術として SGML が国際規格となって十余年が経過したが、欧米では広く一般に普及し、(米国国防総省での軍事資料から民間でもマニュアルや技術文書、論文などに広く利用され)各種ツール群も整備されつつあるのに比べ、日本国内では今一つ「特殊な技術」という認識から脱し得ず、その普及範囲もやや限られた分野にあった。以下は先進的な各分野での事例である。

まず学術分野では古くは学術論文への採用に始まり、主に学者や研究者の評価を得て論文や紀要のデータベースが作られた。

文部省大学共同利用機構の学術情報センターでは、著作権の及ばない文献や著作者の了解を得た文献を中心に学術情報データベースを SGML にて構築し、NACSIS-IR として広く研究者の間で共同利用している。

国文学研究資料館では同様に古典文学(和歌、落語の小噺、和古書など)や歴史史料、あるいは同館内にある目録資料などを順次 SGML に変換して、研究用途に利用している。

行政機関での SGML の採用は、前述の「行政情報化推進基本計画」をもとにした行政改革の一環であり、

- ①白書等の統一的な仕様による情報公開、
- ②許認可申請、届出等手続きの迅速化・高度化、

- ③行政の事務処理の効率化・高度化、
- ④意思伝達の高度化、
- ⑤行政情報の省庁間総合利用、
- ⑥行政の情報化・高度化によるコスト削減（小さな政府）

などである。具体的には情報処理振興事業協会による「行政情報検索提供実験」の成果を踏まえ、総務庁主導で白書等の SGML 化が進められていることと、厚生省への「医薬品・医薬部外品・化粧品、医療用具」の許認可申請とその審査システム、特許庁の「特許情報システム」などがある。

CALS に代表される民間企業の動きは、一時期業界内での業務の連携強化を目標に SGML による文書仕様の標準化が提唱されたが、景気が低迷するなかで新たな投資意欲を掻き立てるには至っておらず、ただそれでも工業製品の輸出企業などで製品マニュアルやドキュメント類を SGML で提供するほか、一部積極的な企業で組織のスリム化や業務推進のコストダウンをはかって独自の SGML 戦略を実現するケースもある。

そうした中で、SGML に携わる技術者は相変わらず不足し、その必要性を認識している大手システムベンダーも、この事業領域としての取り組みに相当の期間を費やしてきた。技術的にも、常に日本のコンピュータ業界を悩ましてきた「日本語の取扱い」という問題がここでも立ちはだかつており、海外の SGML ツール群が 2 バイト文字を取り扱っていなかった事もあって、それらの導入を妨げてきた。特に SGML 文書の閲覧ソフト（ブラウザ）や検索ソフトの日本語対応版が少なく、SGML のメリットが一般に理解されにくい環境にあったことも事実であろう。また、あらゆる SGML に必須の DTD（Document Type Definition - 文書型定義）を作成するために、文書の構造解析と SGML に精通した技術者が必要であり、その意味でも SGML がわかりにくいと言われてきた所為であろうか。

1.2. XML の出現とそのインパクト

1996 年秋に W3C（World Wide Web Consortium）の XML ワークグループより、XML（eXtensible Markup Language）の草案が起草され、1997 年 5 月にはスペインのバルセロナでの SGML/XML 国際コンファレンスで発表、同年 12 月のワシントンでのコンファレンスにおいて、XML を W3C の Proposed Recommendation とし、1998 年 2 月 10 日付けでレビューを終わった XML Ver.1.0 を発表した。これを受けて日本規格協会でも XML 仕様の翻訳と日本語解説を加え、同年 5 月には JIS の TR(テクニカルレポート)として公開した。

インターネット時代の到来とともに、Web 上で出版する手法として HTML（Hyper Text Markup Language）が考案され、初心者にも使える手軽で安価な手法として重宝されてきた。HTML は「世界最大の SGML アプリケーション」と言われてきたが、HTML の機能だけで企業や組織間の文書をやりとりすると、いろいろな障害が発生する。これは HTML

が Web 上で配信された文書の表示フォーマットを決めるだけの機能しか持ち合わせていないため、文書の構造化ができない事にある。そのため、配信された文書を再利用するにも大きな障害となっていた。

Web 上で SGML 情報を配布し、その情報を効果的に再利用するために XML は開発された。XML と SGML は、使用用途により異なるが、根本的な部分ではほとんど同じものとして考えられる。現代の Web 上で構造化情報を配布する際に不要な SGML の機能を削除し、ネットワークに相応しい制約を加えた XML の仕様は、紛れもなく SGML でありながら、ネットワークの負荷を軽減し、スタイルシートやブラウザなどの XML ツール群も整備されつつあることから、あらゆる分野で大きな期待を集めている。文書管理や知的財産管理など、これまで SGML が必須と言われてきた世界でも、XML とネットワークによる情報共有の実例が数多く出始めている。

XML はインターネット時代の文書/データ記述言語であり、SGML のサブセットとして XML が提供する階層構造でほとんどの情報を表現することができる。SGML 同様に可視的なタグの記述が可能なことや、国際文字セットへの対応、強化されたハイパーリンク機能などが各方面での評価を得、IT(Information Technology)企業各社がその採用に動き出した。マイクロソフト社は IE4.0 および Office で XML をサポート、Netscape 社も公開ソースコードに XML パーサを組み込み、サンマイクロ社の Java は XML Library を提供、オラクルの RDB Oracle8i にも XML パーサを搭載、などである。

EC/EDI の世界が最も具体的な XML アプリケーションであろう。むしろ XML を待ち焦がれていたはずである。従来の VAN や専用線と専用ソフトを使った EDI システムは、コストや柔軟性、可搬性を犠牲にせざるを得なかった。世界中につながるインターネットをインフラとして注目してはいるが、SGML ではロジックが重くなるからだと思われる。XML の本格化をみて Verisign や Entrust、Baltimore、Netscape、RSA、Microsoft、E-Certify、Xcert などの公開暗号鍵を使った電子認証サービスが始まり、これらに対応した KyberPASS(カイバーパス)という統合セキュリティシステムも世に出て、CALIS が目指した EC(Electric Commerce—電子商取引)時代がついに到来した。

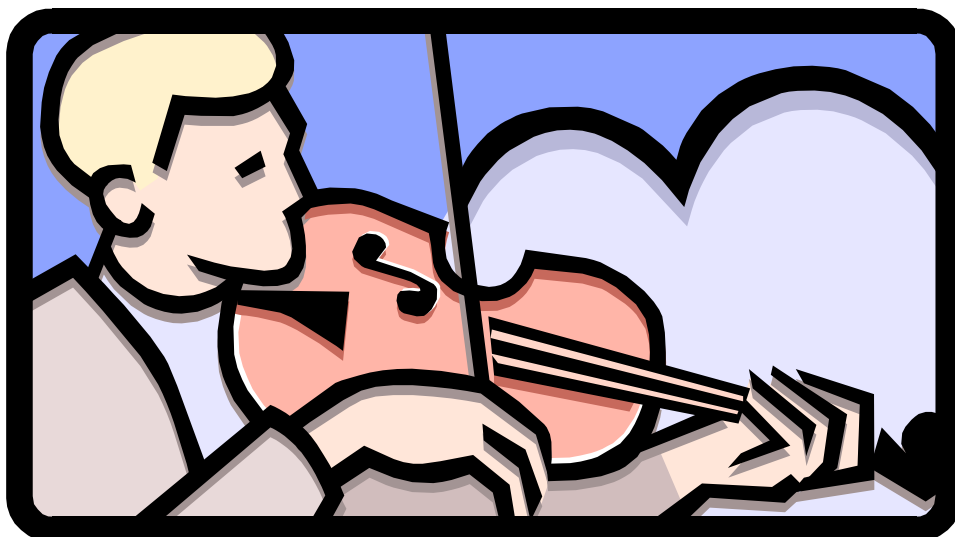
この他にも様々な XML アプリケーションが出始めている。電子地図(GIS)の流通を図るための G-XML、健康診断のデータを転勤先でもネットワークで取得できる HDML、携帯電話向けの XML である WML(WAP Mark-up Language)、マルチメディアを定義する SMIL Boston、MathML(数式)、CML(化学式)、インターネット TV(World Wide Vision)のための XML によるコンテンツ記述モデルである MAPS(Multimedia Architectural Platform Specification)など、日本と世界のあらゆるところで XML が動き始めている。

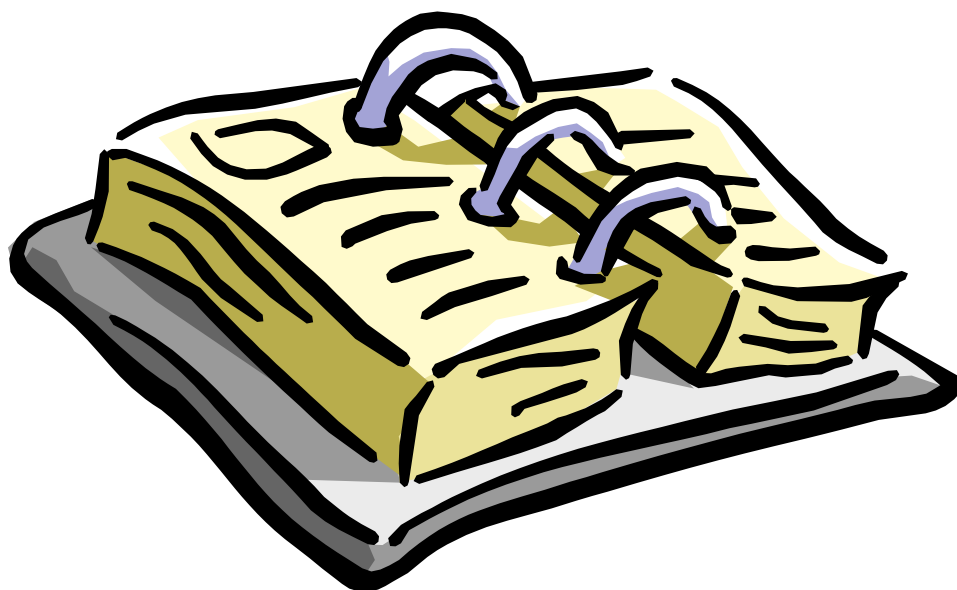
1.3. 全文検索技術

従来のコンピュータ処理に比較すると、Web 検索や電子図書館などの文書型データベースは追加型がほとんどであり、その効果的な処理には「全文検索」が必須となっている。特に膨大な蔵書の中から目的の書物を探し出すにあたり、タイトルや著者、出版社などが予め分かっていたらよいが、そうでなく「こんなことが書かれている本・・・」といった気軽な検索には、書物の中身がテキスト化されていて、全文の中から目的の文章を探し出す必要がある。以下にその必要性を図示してみた。

こうした全文検索システムの中で、はじめから SGML を意識して開発されたカナダ OpenText 社の検索システム「OpenText」があり、以下にその簡単な手法と事例を紹介したい。

モデルケース：新聞社 A 社





1.4. CALS での取り組み

1996年4月に通商産業省の指導による「企業間高度電子商取引推進事業」の一環として、情報処理振興事業協会と9業種10プロジェクトによる「CALS 実証実験」が行われ、1997年度で2年間の実験期間を終えた。技術文書の交流や電子的な商取引など、文書管理に関わる部分について SGML の採用を含めた報告がなされている。以下は各プロジェクトの SGML に関わる活動の簡単な抜粋である。*

(1)宇宙 CALS

人工衛星製造の過程において、発注者と受注者・製造者に至る複数の企業間での技術情報共有と整備による生産性の向上がテーマで、「インターフェース管理図面」および「技術連絡書」についての SGML 化の実験を行った。

(2)航空機 CALS

米国ボーイング社での航空機開発段階における SGML を採用した CALS の実績を踏まえ、プロジェクトは進められた。このテーマは SGML による「技術文書の作成・管理実証実験」であった。

(3)自動車 CALS

多くの対米輸出があるこの業界は SGML によるドキュメント整備が必須であるといわれる。そういった中で自動車 CALS では、SGML に関連するテーマとして「サービス情報ネットワークシステム」「法規情報データベースシステム」の実証実験を行った。

(4)鉄鋼設備 CALS

業界が異なる多数の企業が参加するこのプロジェクトは、いかに業務連携を図ることができるかという観点で、「設備管理情報の標準化」「企業間業務連携モデル」という2つのテーマを取り上げ、具体的に SGML に関わる実証実験を行った。

(5)電子機器・部品 CALS

部品供給側である部品製造業社とそれらの部品を基に機器を設計する機器設計業者間でのタイムリーで効率的な部品情報交換のための標準化を目的とし、Pinnacles-DTD を改良した「E-CALS DTD」を作成して「提供情報の標準化」「情報作成・提供システムの標準化」の提唱を行った。この中では図面情報のリンク機能だけでなく、PDF ファイルに対するリンク情報も組み込まれている。

(6)プラント CALS

プラントの設計から運転・保守までの業務の効率化やプラントの安全性向上を目指し、SGML による総合実証「調達業務遂行のための標準化」を行った。これはプラント・データの交換モデルの確立を目的としたもの。

(7)ソフトウェア CALS

ソフトウェア業界での互いのソフトウェア開発情報や様々なノウハウを、セキュリティを確保しながら共有化できる環境を整備し、共同開発プロジェクトが業界内のどの企業とも効率的に実現できる環境を整備する。そのための SGML による各種文書作成要領・手順書の標準化に取り組んだ。

(8)船舶 CALS

船舶分野においても、「技術文書の交換機能の開発」「船舶設計図書の提出および審査・承認業務の電子化の実証」をテーマに、代表的な4種の技術文書（建造仕様書・注文仕様書・トリム計算書・横強度計算書）の DTD 設計および技術文書データベースの機能要件の研究を行った。

(9)電力 CALS

今回の実証実験でのパイロットケースとされたのが電力分野であり、火力発電所の発電プラントにおける保守業務をテーマに実施された。「技術連絡書」の標準化にあたり DTD 設計の考え方で意見が分かれたが(全業界をカバーするスーパーセットとするか、業界別の個別 DTD とするか)、SGML 実装の容易さを考慮して個別 DTD となったようである。

※参考文献 財) データベース振興センター：「データベース構築・流通および利用に関する基礎技術調査報告書」平成9年度通商産業省委託調査(1998)