

# SGMLとWWWによる情報活用の 一実現形態

3Q-4

山口恵吾 速水治夫 岩切高明 大林恵次  
NTT情報通信研究所

## 1.はじめに

定型的なドキュメントの記述方法としてStandard Generalized Markup Language<sup>1)</sup>(以下SGMLと呼ぶ)が注目されている。SGMLではドキュメントの論理構造をタグで表現できるため、プログラムによる自動編集が容易である。一方、広く不特定多数への情報発信としてWorld Wide Web(以下WWWと呼ぶ)による方法が脚光を浴びている。

本稿では、SGMLで記述された大量のドキュメントから必要な部分を抽出し、WWWで発信するシステムを試作したので報告する。

## 2.従来の問題点

WWWで提供する情報は、Hyper Text Markup Language(以下HTMLと呼ぶ)で記述する必要がある。このため、蓄積されているテキスト情報を提供する場合には、情報発信する部分を抽出しタグと呼ばれるマーク付けを行わなければならない。従来、このような作業はすべて人手により実施していた。

これらの作業を自動化するには、必要情報だけを抽出し、HTMLへの変換が必要であるが、ワープロ文書等のドキュメントは文書の論理構造記述を持たないため、抽出・変換処理の自動化は非常に困難であった。

## 3.SGMLの利用

### (1)処理方式概要

先に述べた問題を解決する方法としては、ドキュメントが何らかの構造記述を持つ必要があり、SGMLの効果的な適用法を検討した。

図1は、SGML文書インスタンスからHTMLを生成する機能の概略を示したものである。

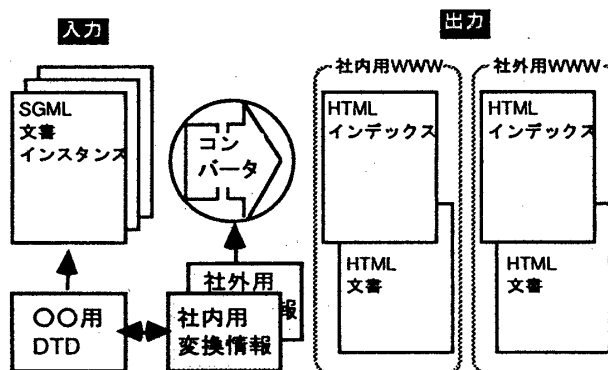


図1 抽出・変換方式の概略

SGMLは、文書型定義(Document Type Definition; 以下DTDと呼ぶ)により論理構造(表題、著者、作成日、あらまし、本文など文書内の意味)を定義することができる。DTDは、用途対応に作成され、SGML文書はこのDTDに従って複数のインスタンスが生成される。したがって抽出・変換処理を固定的に作成せず、プロシージャと変換情報(処理指定)に分け、前者をコンバータとして作成し後者をDTDに対応させて定義することにした。この方法により、1つの変換情報により複数のSGML文書インスタンスを処理することができる。また、1つのDTDに対して複数の変換情報を指定することにより、例えば1つのSGML文書インスタンスから社内用、社外用のHTML文書を作成できる。

コンバータは、SGML文書インスタンスを入力とし、HTML文書およびHTMLインデックスへ出力する。HTMLインデックスは、すでにホームページからリンクされたファイルでありHTML文書出力時に変換文書のタイトルとリンク先を追加する。

### (2)抽出・変換処理例

図2は、変換情報ファイル(TYPE1.CNF)をもとにSGML文書インスタンス(07-201.SGM)からHTML(07-201.HTM)を生成する様子を一例として示した

An Information Delivery System Using SGML and WWW

Keigo Yamaguchi, Haruo Hayami, Takaaki Iwakiri, Keiji Ohbayashi

NTT Information and Communication Systems Laboratories  
1-2356 Take, Yokosuka, Kanagawa, Japan

ものである。

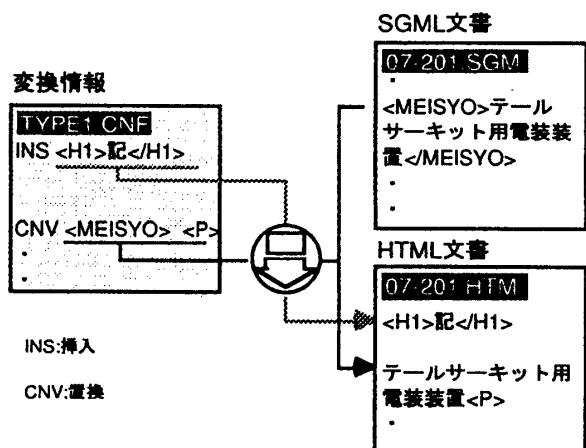


図2 抽出・変換処理の一例

TYPE1.CNFファイル一行目のINS文は、変換元に存在しない情報をHTMLへ出力するための挿入命令を示す。二行目のCNV文は、<MEISYO>タグを07-201.SGMから検索しその情報を<P>へ変換してHTMLへ出力するための置換命令を示す。

#### 4.提案方式の実現例

##### (1)システムにおける本方式の位置づけ

図3は、既存のドキュメント管理システム<sup>[2]</sup>に本提案方式を組み込んだ場合の構成である。

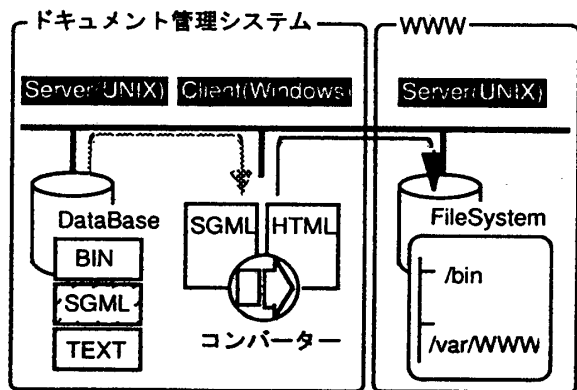


図3 提案機能を組み込んだ場合の構成

ドキュメント管理システムは、クライアント・サーバ型のシステムで構成され、サーバのデータベースでドキュメントを管理する。クライアントでは登録や参照などの操作で、サーバ上のドキュメントを自由にアクセスできる。

ドキュメント管理システムのSGML文書インスタンスを利用するためには、クライアントの操作が必要であり、今回この部分に提案した機能を実装した。

##### (2)クライアントの構成と処理の流れ

図4は、実装した機能の構成と処理の流れを示したものである。

ドキュメント管理システムには、Windows APIを利用できる機能があり、これを利用した。具体的には、試作機能をDynamic Link Library(CONV.DLL)としてWindowsへ登録することにより、ドキュメント管理システムの一部として動作できる。

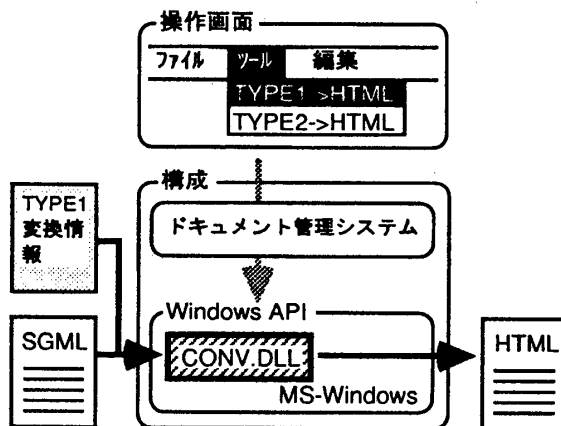


図4 構成と処理の流れ

処理の起動は、参照などの操作をおこない、SGML文書インスタンスをクライアントに呼び出す。ここで、今回試作した変換情報をメニューから選ぶとSGML文書インスタンスからHTMLの生成を行いNFS機能を利用してWWWサーバのホームディレクトリへ書き込む。

##### 5.今後の予定

今回試作した機能を用いた結果、複数のSGML文書インスタンスからHTMLを生成しWWWサーバへ登録でき、文書インスタンスの修正も即座に反映できることを確認した。

今回の出力はHTML文書の出力とインデックスへの追加という固定的な範囲に限定した。今後は、本方式の定量的な評価を行うとともに、変換情報ファイルを容易に記述する方法、および柔軟な出力についても検討する予定である。

##### <参考文献>

- [1]吉岡誠編著:"SGMLのススメ",株式会社オーム社,1993
- [2]山口恵吾他:"ドキュメント管理システムにおけるクライアント間同期方式の一提案",信学総大D-243(P241), 1995.3