

社内文書のSGMLによる標準化とイントラネットによる活用

3AC-5

森口 修 今村 誠 鈴木 克志

三菱電機株式会社 情報技術総合研究所

1. はじめに

文書を計算機上で作成する機会が日常的になったが、非定型の社内業務の場合、作成した文書を結局紙に印刷してから業務フローに乗せているのが現状ではまだ多い。しかし、文書を電子化した状態のままネットワーク経由で最後まで利用することで、社内業務を効率化するという要求は高い。そのためには、文書を標準化して必要な内容を文書から機械的に抽出する仕組みが必要である。その解決策を提案するものとして CALS (Commerce At Light Speed)の構想がある。CALSでは、文書の形式を SGML (Standard Generalized Markup Language)という文書記述言語で統一し、機械可読なマークアップタグを計算機により解析することで、所望の文書処理を行なう方法を示唆している。

本稿では、まず社内業務で必要となる文書管理機能について分析する。次に、文書管理機能を実現するために必要となる SGML ツールを作成し、SGML ツールを用いて社内業務における文書管理システムを構築する方法について述べる。最後にまとめとして、社内業務文書管理システムに期待できる効果と今後の課題について述べる。

2. 文書管理の要求分析

社内業務では、文書の検索機能を基本として、文書の構成管理、履歴管理、アクセス権制御、ワークフローやコラボレーションなどの機能が要求される。これらの機能を実現するためには、文書を管理するための情報として文書属性(書誌情報)を文書に付与する必要がある。しかし、市販の文書管理ツールを利用する場合、ツール毎に提供される機能とそのため文書属性が限定されるので、その範囲の機能しか実現できない。

そこで文書を SGML とすれば、実現したい機

Practical Use of interoffice documents by standardization on SGML and Intranet

Osamu MORIGUCHI, Makoto IMAMURA, Katsushi SUZUKI
Mitsubishi Electric Corporation.

5-1-1 Ofuna, Kamakura, Kanagawa 247, JAPAN

能を先に分析し、そのために必要な文書属性を文書内に記述するための DTD(Document Type Definition)を後から設計することにより、社内業務における様々な要求に応えることができる。さらに、SGML 文書から文書属性を機械的に抽出することができれば、文書作成と同時に業務を進行することができる。

3. SGML ツールの作成

SGML 文書処理する方法として、文書構造操作言語を定義するもの[1]があるが、記述を容易にするために記述に制約を課すことにより複雑な処理が記述できなくなる。ここでは SGML ツールと呼ぶ SGML 文書を操作するための API(Application Programming Interface)を作成し、2章で要求機能として述べた、文書から属性を機械的に抽出する機能を SGML ツールを利用して実現する。

SGML ツールは、SGML 文書の任意の要素および要素属性にアクセスするための関数群からなる API であり、その代表的な関数を表 1 に、要素および要素属性からなる文書構造木を図 1 に示す。

表 1. SGML ツールにおける関数と機能

群	関数の機能
解 析	SGML 文書実体から文書構造木を作成
	文書構造木を開放
要 素	要素を入力し上位要素を出力
	要素を入力し下位要素リストを出力
	要素リストを入力し先頭要素を出力
	要素リストを入力し先頭以外の要素リストを出力
属 性	要素を入力し要素属性リストを出力
	要素属性を入力し属性値を出力
	要素名で検索し要素リストを出力
検 索	要素名と属性値で検索し要素リストを出力
	要素内容で検索し要素リストを出力
実 体	外部実体名を入力し外部実体を出力
	外部実体を入力し外部実体識別子を出力
記 法	記法名を入力し記法を出力
	記法を入力し記法識別子を出力

SGML ツールを利用すると、要素・要素属性・

要素間の階層関係による条件分岐や、複数要素の繰り返しによるマトリックスや、or 接続子、opt,plus,rep 出現標識、要素間の ID 参照などを含む複雑な構造の SGML 文書処理にも対処できる。

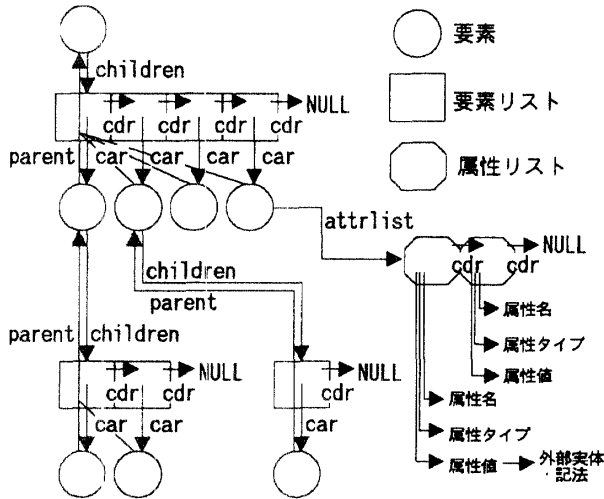


図1.要素と要素属性の内部構造

4. 社内業務文書管理システムの構築方法

データベースおよびイントラネット上の Web サーバに加えて、SGML ツールを利用した社内業務文書管理システムの構成を図2に示す。従来からこのようなシステムは提案されているが[2]、このシステムは、SGML 文書からの文書属性抽出、文書属性の一元管理および文書実体の分散を特徴とし、文書属性の自動登録、文書の一括検索、イントラネットによる文書アクセスが可能である。

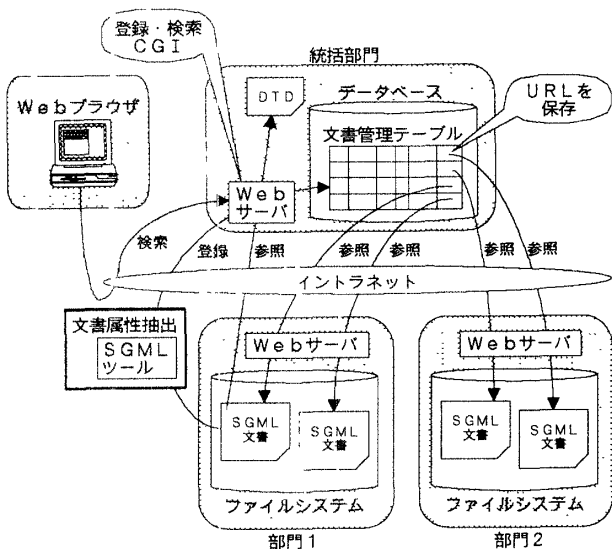


図2.社内業務文書管理システムの構成図

4.1 SGML 文書からの文書属性抽出

SGML 文書から検索のインデックスとなる文書

属性を抽出するため、SGML ツールを利用する。文書属性抽出は、まず SGML 文書を表1の関数により解析し、図1に示すような文書構造木を作成する。次に文書構造木を走査しながら、データベースに登録すべき文書属性を抽出する。抽出した文書属性と文書の URL を統括部門 Web サーバの文書登録 CGI によりデータベースに登録する。

4.2 文書属性の一元管理

検索のインデックスとなる文書属性は統括部門のデータベースで一元管理し、複数のリポジトリにある文書を統括部門の Web サーバの文書検索 CGI(Common Gateway Interface)により一括して検索する。さらに、SGML 文書からの DTD の外部実体参照は統括部門 Web サーバへの URL とし、標準 DTD を統括部門で一元管理する。

4.3 文書実体の分散

各部門が作成した SGML 文書実体は、当該部門のリポジトリに分散して保存する。各部門に Web サーバを置くことで、イントラネット経由での他部門からの URL(Uniform Resource Locator)による文書アクセスを可能とする。

5. まとめ

社内業務文書管理システムの構築について述べた。システムの特徴は以下の3つである。

- (1) SGML 文書から文書属性を抽出し、文書属性の自動登録が可能。
- (2) 文書属性を統括部門データベースで一元管理し、文書の一括検索が可能。
- (3) 文書実体を各部門 Web サーバで管理し、イントラネットによる文書アクセスが可能

今後の課題は以下の3つである。

- (1) SGML ツール改良により DTD の変更に伴うシステム再構築作業を省力化
- (2) 文書の構文や内容を事前に検証することにより業務フローにおけるエラー文書を削減
- (3) 社外標準 DTD の取り込みや社外とのセキュリティ確保によるインターネットへの拡張

参考文献

[1] 角間：“SGML 文書の論理構造変換方式”，情報学会第 52 回全大,4J-8(1996)
 [2] 神山,佐藤：“CALIS 対応文書管理システム”，情報学会第 54 回全大,7S-6(1997)