

SGML 文書構造エディタ

豊田 清 熊谷 恵理 坂東 達夫

松下電送(株) 技術研究所

要旨

SGMLの規格に従った文書は通常のワードプロセッサやテキストエディタなどでも作成できるが、作成者は文書の論理構造を理解し、SGMLに精通する必要がある。作成した文書に構成要素の開始、終了を示すマーク付けのエラーも発生しやすい。

そこで、SGMLについてほとんど知識がなくとも、SGMLの規格に従って文書を誤りなく作成することができるSGML文書構造エディタを開発した。本エディタは文書構造の把握、マーク付けのエラー防止を考慮して、作成する文書の構成要素をあらかじめ画面に表示することを特徴としている。

SGML Document Structure Editor

Kiyoshi Toyoda Eri Kumagai Tatsuo Bando

Engineering Research Lab., Matsushita Graphic Communication Systems, Inc.

Abstract

In order to prepare a SGML-coded text using a conventional word processor or text editor, the user must have thorough knowledge of the SGML, and understand the structure of the document. To solve this problem, an SGML structured editor has been developed. The editor gives guidance to the user how to mark the document correctly. It displays the marks of the structure elements before the user inputs the text so that the document structure is understood and mark-up errors are avoided.

1. はじめに

欧米の公的機関や企業では、膨大な量の文書を編集、検索、配布など統合的な観点から効率良く処理するために、SGML(Standard Generalized Markup Language)¹⁾を文書記述言語として導入する動きが活発になっている。特に、米国国防総省の CALS (Computer Aided Acquisition and Logistics Support) プロジェクトは、設備、兵器に関する技術文献、手引書、規格書などの記述に SGML を導入し、無駄な資料の廃棄や検索誤り等を無くして大幅なコストの削減を実現しようとしている。このプロジェクトにより米国の SGML 関連の製品市場²⁾は急速な立ち上がりを見せている。

SGML は 1986 年に ISO (国際標準化機構) によって標準化され、日本では 1991 年 3 月末に SGML-JIS 原案が作成され、現在審議中である。最近では SGML 懇談会³⁾等の活動により国内でもその有効性が認識されてきており、電子文書処理の効率化の核心技术として、文書の統合的な処理を必要とするシステムを運営する人達など各方面からの期待が高まっている。

SGML で記述された文書を誤りなく作成するには専用のエディタが必要となる。そこで、筆者らは日本語文書の文書型定義 (DTD: Document Type Definition) にそって文書を作成することのできる SGML 文書構造エディタ (以下、SGML エディタと呼ぶ) を開発した。⁴⁾本稿ではその特徴及び統合的な文書処理への応用例について述べる。

2. SGML 文書の基本構成

SGML 文書は SGML 宣言、DTD 及びタグ付き文書の 3 パートによって構成される。SGML 宣言は SGML 文書を作成する際のコード系、タグコードの定義等を行う。コード系は文字を表現するための符号化規則であり、処理システムの種類や国によって異なる。タグコードはそのマークを定義するビット列である。DTD は文書の論理構造、つまり、論題、著者、要旨、本論のような文

```
<!DOCTYPE 論文 [
<!ELEMENT 論文 -- (論題、著者、要旨、本論、参考文献?)>
<!ELEMENT 論題 -- (#PCDATA)>
<!ELEMENT 著者 -- (著者名、所属)+>
<!ELEMENT 著者名 -- (#PCDATA)>
<!ELEMENT 所属 -- (#PCDATA)>
<!ELEMENT 要旨 -- (段落)+>
<!ELEMENT 本論 -- (節)+ (図|表|脚注)>
<!ELEMENT (図|表) -- EMPTY>
<!ELEMENT 脚注 -- (#PCDATA)>
```

第1図 DTD (文書型定義)

```
<論文>
<論題>
SGML文書構造エディタ</論題>
<著者>
<著者名>
豊田 潤</著者名>
<所属>
松下電送株式会社</所属>
</著者>
<要旨>
<段落>
</段落>
</要旨>
<本論>
<節>
<節題>
緒言</節題>
<段落>
```

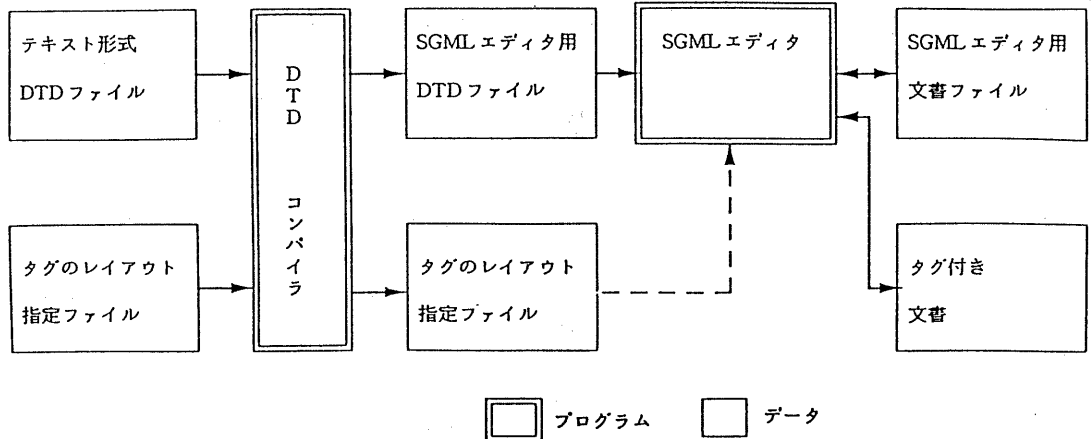
SGMLとは、タイトル、著者、記事などの

第2図 タグ付き文書

書を構成する要素の名前と出現順序等を記述しており、出版社等によって異なる。例えば、本論文の文書構造を SGML で記述したものが第1図の DTD であり、構成要素を SGML にしたがってマーク付けしたものが第2図のタグ付き文書である。タグは文書の各構成要素の開始、終了を示すマークである。

3. DTD コンパイラと SGML エディタの関係

SGML エディタは文書に応じた DTD を入力し、この DTD にしたがって文章を作成することによってタグ付き文書を出力する。ただし、テキ



第3図 DTD コンパイラとSGML エディタの関係

スト形式の DTD をそのまま入力するのではなく、SGML エディタで効率良く処理するために、DTD に独自の変換を行ったファイルを入力対象としている。この変換を行うためのプログラムを DTD コンパイラと呼んでいる。DTD コンパイラと SGML エディタの関係を 第3図 に示す。ここで、SGML エディタ用文書ファイルとはタグ付き文書を SGML エディタが効率良く処理できるようにフォーマット変換した独自のファイルである。ユーザは通常このファイルを使って編集を行い、他の SGML 処理系とデータ交換する時にはタグ付き文書を用いる。

SGML 宣言はシフト JIS コードを扱えるように、ISO のひな形となる典型的な SGML 宣言を修正したものを前提としており、SGML エディタに SGML 宣言を入力する必要はない。ただし、タグ付き文書を他の SGML 処理系とデータ交換するためには、この SGML 宣言が必要となる。

4. 特徴

カナ漢変換、あるいは文書入力の基本的な機能についてはワープロと同様であるが、タグ付け

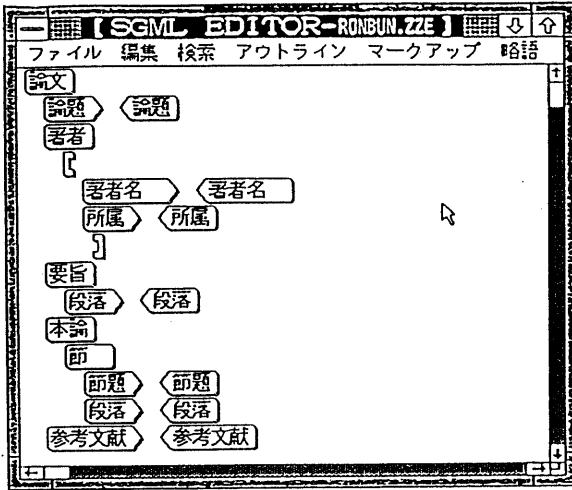
のエラー防止、わかりやすい画面構成などを考慮して以下の特徴をもつ。

1) DTD に対応した要素内容のタグが画面にガイダンスとして表示される。

例えば 第1図 の DTD を SGML エディタで読み込むと 第4図 の画面が表示される。DTD に記述された要素の中で必須要素とオプション要素が表示され、OR 結合子 | で結ばれた物は先頭の要素が表示される。グループとして扱う要素は [] で囲んで表示する。

2) アウトラインプロセッサの機能を有しており、章や節の作成、編集ができる。

DTD に定義した要素の階層レベルに対応してタグの表示位置をかえており、ユーザが文書の階層構造を理解できるようにしている。さらに、階層レベルに応じて表示、未表示を選択することができるので、文書のアウトラインだけ表示して編集することができる。第5図 に示す網掛けされたタグは、さらに下位レベルの要素を持っていることを示している。このタグをマウスでダブルク

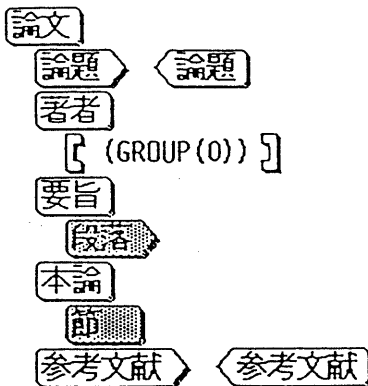


第4図 全階層レベルのタグ表示

リックすれば、下位の要素が展開表示される。

3) DTD がツリー構造で示され、文書全体の構造を文書作成中に確認できる。

編集中の文書の最上位要素から、現在のカーソル位置の要素及びその要素を構成する要素をグラフィカルなツリー構造で表示する。(第6図)



第5図 階層レベル2を指定したときのタグ表示

4) SGML 文書を LaTeX コマンド付き文書に変換できる。

DTD で定義されている文書の各要素が、LaTeX のどのコマンドに対応しているかを記述した変換ファイルを SGML エディタの入力として、LaTeX コマンド付き文書を出力する。対応付けは SGML エディタ独自の簡易言語を用いる。

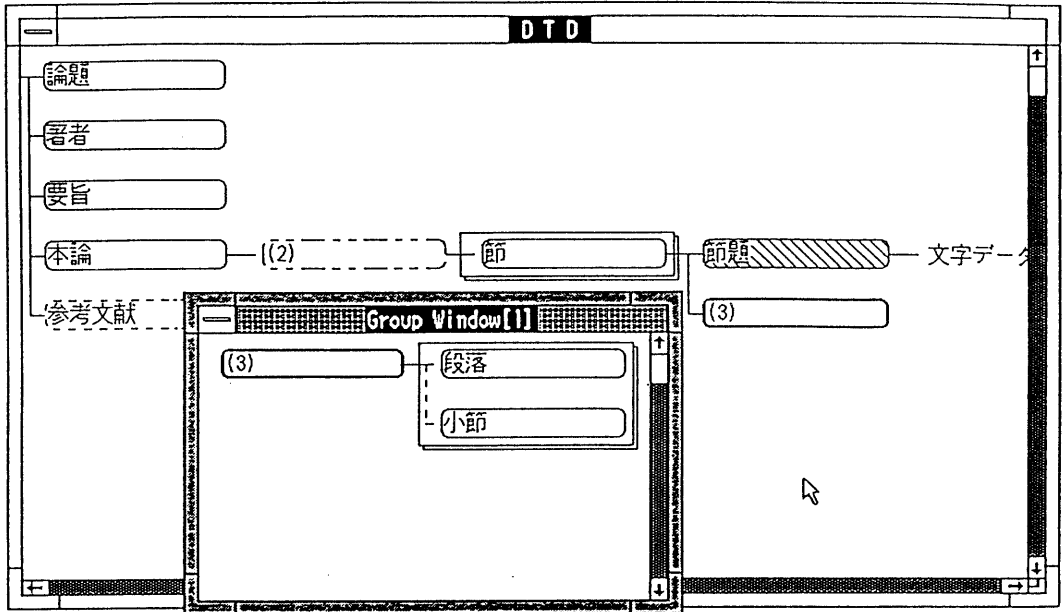
5) タグ付き文書を入力して、編集できる。

タグ付き文書を SGML エディタ独自のフォーマットに変換して編集する。DTD は DTD コンパイラによって変換されたものを入力とする。SGML 宣言は SGML エディタで扱えるものを前提としている。

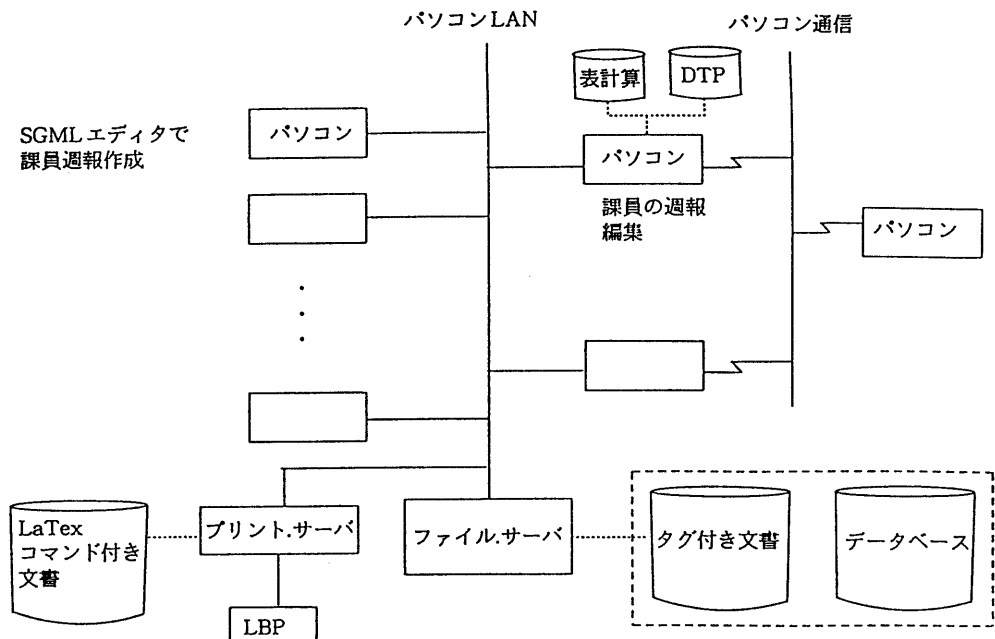
5. 応用

SGML 文書は論理的な構造情報を持っているため、様々な処理系の原情報として使うことができる。例えば第7図は週報の管理システムの例である。各課員は、パソコン上の SGML エディタで週報を作成し、タグ付き文書ファイル・サーバーに登録する。各課員の週報は成果、売上げなどの論理項目別にまとめられ自動的に課長の週報が作成される。課長はこれを SGML エディタで読み込み、編集して LaTeX コマンド付き文書に変換して印刷する。必要なら、DTP 用のファイルに変換し自分で編集したり、表計算用のファイルに変換して今週の売上げ合計を計算することもできる。さらに、ファイル・サーバーから必要な情報を検索することも可能である。

タグ付き文書はテキスト形式で書かれているので、パソコン通信で外部のパソコンへ送信し、受信者はそのまま文書を読むこともできる。受信者が SGML を処理するシステムを持っていれば、好みのフォーマットで印刷したり、表計算へ取り込むことも可能である。このように、SGML で文書を作成しておけば処理の範囲が広がり、将来のシステム拡張への対応が柔軟にできる。



第6図 DTDのツリー構造



第7図 SGMLエディタを使った文書管理システム

6 . 今後の課題

今後、SGML エディタを使った統合文書処理システムの実用化と評価を行い、さらに SGML の有効性を検証していく。また、SGML エディタに画像、図形を取り込む機能を付加すること、及び、表、数式の入力のユーザ・インターフェイスの改善などを進めていきたい。

参考文献

- 1)ISO 8879 Standard Generalized Markup Language (SGML)1986-10-15
- 2)CAL S Expo'89:Commercialization of CAL S Products,SEYBOLD Report on Publishing Systems, vol.19,num.9,January 29 1990 p15
- 3) S G M L 懇談会「今すぐ使える S G M L 処理ソフトウェア」標準化ジャーナル、Vol.2 0 1990.5.p47
- 4)豊田他、「SGML 文書構造エディタ」、National Technical Report Vol.36 No.5 Oct.1990 p89 ~ 95