

## 文書内容の操作に見るマークアップの効果

山口琢<sup>†1</sup> 小林龍生<sup>†2</sup>

<sup>†1</sup> 株式会社ジャストシステム

<sup>†2</sup> ジャストシステムデジタル文化研究所

**概要** 小説や報告書といった通常の意味での文書に施す電子的なマークアップの効果、実装システムを例題にして議論したい。われわれは、文書への電子的なマークアップが、書き手の創意工夫や、読み手の発見を促し、理解を深める助けになると考えている。これは、文書をXMLで記述することでレイアウト・組版作業を自動化し効率向上することや、商品情報データベースの拡張性を高めることなどは別種のメリットであろう。

**キーワード** XML、マークアップ、タグ、文書

### Consideration on the Effect of Markups in Writing, Reading and Operating Narrative Documents

YAMAGUCHI Taku<sup>†1</sup> KOBAYASHI Tatsuo<sup>†2</sup>

<sup>†1</sup> President Office, JustSystems Corporation

<sup>†2</sup> Justsystems Digital Culture Research Center

**Abstract** This paper discusses the effects of electronic markups, made on narrative documents such as novels and historical tales, or study reports with some experimental implementations. The authors are confident that electronic markups made on documents promote creative ideas of good writers and discovery and deep understanding of readers. There are different types of merit from those of automatic layout of documents or the expandability of the product information database.

**Keyword** XML, markup, tag, document

#### 1. はじめに

われわれのテーマと取り組み

われわれは、新たな文書の《読み》《書き》の可能性をひらきたいと考えている<sup>[1]</sup>。そのために、人々が意識的・無意識的に行っている《読み》方、《書き》方を見だし、それを《読み》《書き》ツールとして実装している<sup>[2] [3]</sup>。

この取り組みは発想支援にも関係している。文書の《読み》《書き》に試行錯誤や発散的・収束的思考を導入することで、新たな《読み》《書き》の可能性がひらかれるのではないかと<sup>[4] [5]</sup>。

この取り組みは知識共有にも関係している。《読み》方や《書き》方の一部を、人々の頭の中からICT環境に《外在化》し電子的な手段として実装すると、知識共有を促進するのではないかと<sup>[5] [6]</sup>。

この報告の構成

この報告では、XMLの観点から、われわれの取り組みの着眼点について詳しく論じて、課題、アプローチ、実装例を紹介する。

われわれは、文書を、要素の入れ子として構造化することが、文書の新たな《読み》《書き》の可能性をひ

らくために有効であると考えている。そこで、われわれは、文書データの記述にXMLを採用して、《読み》《書き》ツールの実装にはxfy<sup>[8]</sup>を採用している。

#### 2. XMLとマークアップ

XML<sup>[7]</sup>では文書をテキストで表現し、ある一部分を開始タグと終了タグで囲むことで、その部分が文書の中で意味を持つ要素であることを示す。このように文書全体の一部分を区切ってマークをつけることをマークアップ(markup)と呼ぶ。XMLでは、タグによってマークアップを示す。マークアップはXML以前から電子組版システムなどで利用されてきた。HTMLも文書をマークアップで表現する。

このようなマークアップの意味や構造を具体的かつ厳密に設計して、データをマークアップで記述して、高い信頼性でもって高速かつ大量にプログラムで自動処理する取り組みは、従来から行われてきた。また、そのような意味や構造を標準化して、システム間で相互運用を可能とする取り組みも、やはり従来から様々に行われている。XMLは、このようなデータ記述形式として定着している。

これに対しわれわれは、文書の区切りそのものが、対話的な編集の操作性を高める点に着目している。マークアップが抽象的であっても、あるいはシステムがマークアップの具体的な意味に踏み込まなくても、そのマークアップを役立てることができると考えている。

このような視点での技術開発が、これまで不足していたのではないだろうか。

両者の取り組みは、相互に補い合うものである。なぜなら、文書のどの部分をマークアップするかということは、そのマークアップにどのような意味を持たせるかということと不可分だからである。抽象化を外れて、無意味で一貫性を欠いたマークアップでは、利用者に効果的な編集機能を提供することはできない。

以降、この報告では、文書の区切りそのものを、「マークアップそのもの」と呼ぶことにする。

### マークアップとその意味

マークアップの意味に基づいて処理するのは別に、マークアップそのものが役立つとはどういうことか。HTML と対比することで XML を導入する場合を例に、この着眼点を説明しよう。

XML を学び始めたとき、HTML と対比して、[図 1](#) のように XML の長を理解した覚えはないだろうか？(あるいは、HTML から XHTML への移行に際して。)

- ・ HTML では、個々の要素の表示を指定することで、ブラウザウィンドウの幅などに応じて、文書全体を自動レイアウトできる。XML では、個々の要素の意味を指定することで、自動レイアウトだけでなく様々な目的に応じて、文書を機械的に自動処理することが可能である。
- ・ HTML では、あらかじめ決められたタグしか用いることはできない。XML では、新たなタグを導入して、本来記述したかった意味や構造を表現できる。

このような理解そのものは正しい。しかし、このように理解したときに、併せて次のように思い込んでいないだろうか。

- ・ その内容の意味を表現するタグをつけて、その意味に基づいて自動機械処理するのでなければ、文章をマークアップすることにさしたる効果はない。

われわれの着眼点は、しばしばこのようにして見過ごされがちである。すなわち、われわれは、そのようなマークアップでも効果があると考えている。その意味に特化した機能がシステムに実装されていなくても、利用者が意味を理解できるのであれば、そのマークアップに基づく効果的な機能を提供するシステムがありうると考えているのである。

### マークアップを利用する機能

マークアップに基づく効果的な機能の例をあげよう。ここでは、身近な例をあげるために、XML から離れて、従来からあるエディタ、ワープロやアイデアプロセッサで実現されている機能を見る。

### シングル・ダブルクリックで選択・編集

特定の箇所ですぐにシングルまたはダブルクリックするだけで、単語、行または段落を選択できる。そして、それらをドラッグ&ドロップでコピーや移動できる。

### 表の行・列・セルの編集

行・列・セル単位で追加・削除・移動できる。

### アウトライン編集

章・節の追加・削除・移動。

これら機能は、利用者にとって次のような効果を持つと考えられる。

### 小さな・簡単な操作で大域的・複雑な編集

小さな・簡単な操作で、文書の内容を大域的・複雑に書き換えることができる。一文字の挿入・削除と比べると、段落や章など、より大きな単位を対象に操作できる。また、コピーや移動の先は元の位置と連続してなくてよい。さらに、内容を表に整理すると、行・列操作によって、段落や章の操作以上に複雑な編集が行える。

### 試行錯誤

大域的・複雑な編集も小さな・簡単な操作で行えるので、文書内容編集の試行錯誤が操作的にも心理的にも容易になる。

ここで留意すべきは、これら効果は、マークアップの意味に基づくというよりも、「そのように区切られている」というマークアップそのものに由来する、ということである。

すなわち、利用者は章その意味に基づいて移動するが、その操作を支援するワープロがその章の意味を知っているわけではない。

[図 1](#) の表を [図 2](#) のように、含まれるイベントを増やして少し複雑にしてみよう。

タイトル	第 66 回 DD 研	第 67 回 DD 研	第 68 回 DD 研
日時	2008 年 6 月 6 日	2008 年 7 月 24 日(木)、25 日(金)	9 月 26 日
場所	大森ベルポート	北海道大学	大森ベルポート

図 2 複数のイベントを含む表

図 2 で、表編集機能を実装したエディタであれば、2 番目の行を選択して、複数のイベント情報から日時部分を一括して削除することができる。このとき、2 行目が日時であることを知っているのは利用者であって、エディタではない。それにもかかわらず、簡単な操作で日時を削除することができる。

### マークアップそのもの

マークアップそのものと呼んでいるものは、その意味と独立ではない。例えば [図 1](#) の HTML の表を [図 3](#) の

## HTML の場合

## XML の場合

ソース

```
<table>
<tbody>
<tr>
<td>タイトル</td>
<td>第 66 回 DD 研</td>
</tr>
<tr>
<td>日時</td>
<td>2008 年 6 月 6 日</td>
</tr>
<tr>
<td>場所</td>
<td>大森ベルポート</td>
</tr>
</tbody>
</table>
```

```
<event>
<summary>第 66 回 DD 研</summary>
<dateTime>2008-06-06</dateTime>
<location>大森ベルポート</location>
</event>
```

内容の意味に基づいた自動機械処理

できない、あるいは貧弱。  
「2008 年 6 月 6 日」が日時であることは、これを見た人間には分かるが、この情報を処理するプログラムには分からない。

できる。  
「2008-06-06」が日時であることは、dateTime という名前のタグによって示され、プログラムはこれを日時として処理することができる。

人間用に表示

ブラウザでそのまま表示できる。例えば:

タイトル	第 66 回 DD 研
日時	2008 年 6 月 6 日
場所	大森ベルポート

様々に変換・加工して表示できる。例えば:

- ・ HTML と同じ表形式に表示
- ・ カレンダー上にマップして表示

日	月	火	水	木	金	土
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

図 1 HTML と対比して XML のメリットを理解する例

ように、table、tr など全てのタグ、すなわちマークアップの名前を、同じ acb に書き換えたかどうか？これでは、これを表として、行・列・セルの編集機能を利用者に提供するには、多少込み入った処理と UI を伴うことになるだろう。

```
<acb>
<acb>
<acb>
<acb>タイトル</acb>
<acb>第 66 回 DD 研</acb>
</acb>
<acb>
<acb>日時</acb>
<acb>2008 年 6 月 6 日</acb>
</acb>
<acb>
<acb>場所</acb>
<acb>大森ベルポート</acb>
</acb>
</acb>
</acb>
```

図 3 意味を大幅に捨象したマークアップ

すなわち、われわれがこの報告で《マークアップそのもの》と呼ぶのは、次のようなものである:

- ・ 編集操作の観点から規定されるデータ構造である。
- ・ その編集操作を提供するプログラムは、利用者が了解しているよりも、かなり抽象的な意味に基づいて処理する。
- ・ 要素の区切りは明確である。
- ・ その処理は、表組みや箇条書きなど、伝統的なページや文書の組み方を利用することで、利用者にとって分かり易いものとなる。

ここで、かなり抽象的な意味と言っているのは、次のようなものである:

- ・ 情報には種類があつて、それらは互いに区別される。
- ・ ある情報が、複数の要素情報の集合として表現されている。
- ・ それら要素集合の順序に、意味があつたりなかったりする。
- ・ 要素情報の種類は同じだったり、異なったりする。
- ・ それら要素情報は、さらに複数の要素情報の集合として表現されている。

#### 入れ子マークアップの特徴

マークアップそのものと呼んでいるものは、その意味と独立ではないが、別の概念である。最後にそれを示そう。

XML のマークアップは、入れ子構造になっているため、要素の部分的なオーバーラップを表現できない。図 4 では、虹の七色について、小林が暖かい色の範囲を、山口が冷たい色の範囲をマークアップしている。緑色が両方に含まれているが、XML ではこのようなオーバーラップは許されない。

<暖 by="小林">赤、橙、黄、<冷 by="山口">緑</暖>、青</冷>、藍、紫

図 4 オーバーラップする要素

すなわち、どのように区切るかということ、その区切りにどんな名前を付けるのか、どんな意味があるのかとは、別の概念である。

どのような区切りを許すかは、われわれが着目するような処理系の機能やわかりやすさに影響する。例えば、表でセルの結合が可能なエディターの場合、そのような表で行・列の削除や移動したときの振る舞いに、違和感を覚えたことのある利用者は多いだろう。

#### 3. 課題

以上の着眼点から、われわれの課題は、次のような条件を満たす文書編集機能を発見することである:

- ・ 利用者が了解しているよりも、かなり抽象的な意味に基づいて処理する。
- ・ 直感的に操作できる。
- ・ 簡単に操作できる。手数が少ない。

これによって、利用者には次のようなメリットがあるだろう:

- ・ 小さな・簡単な操作で大域的・複雑な編集できる。
- ・ 容易に試行錯誤できる。

このような文書編集機能が効果を発揮するのは、試行錯誤の場面であろう。文書の内容を試行錯誤しながら書いている段階では、各要素の意味は不確定である。本質的に、十分な意味づけができない要素を編集する場面でこそ、われわれの着眼点が生きるはずだ。

#### 4. アプローチ

上記の課題を達成するため、われわれは次のアプローチをとることとした:

- ・ 表組みや箇条書きなど、伝統的なページや文書の組み方を利用する。これによって、利用者にとって直感的で分かり易いプログラムとなるだろう。
- ・ 試行錯誤の場面を観察して、そこで行われる操作を抽象化してみる。

#### 5. 実装例: STORYWRITER

最後に、そのようにして開発した文書編集プログラムの例を紹介する。

##### STORYWRITER

STORYWRITER<sup>[2] [9]</sup>は、多色マーカーツールである。文書の読者は、自由にマーカーを追加することができる。また、マークを手掛かりにして文書の部分を並べ替えることができ、文書を読み解く助けにすることができる。さらに、マークを手掛かりにして、文書を編集することができる。

読者がマーカーを追加すると、それに応じて、STORYWRITER は、class 属性の値を新規に用意する。読者が文書の部分をマークすると、STORYWRITER は、その部分を em 要素で囲い、そのマーカーに対応する class 属性値を設定する。

図 5 では、読者が“殺人”とマークした部分に、“cc88”という class 属性値が設定されている。em 要素とは、HTML で強調を表す要素である。

STORYWRITER の Cross ビューでは、同じ class の em 要素が、表の同じ行に抜き書きされる。

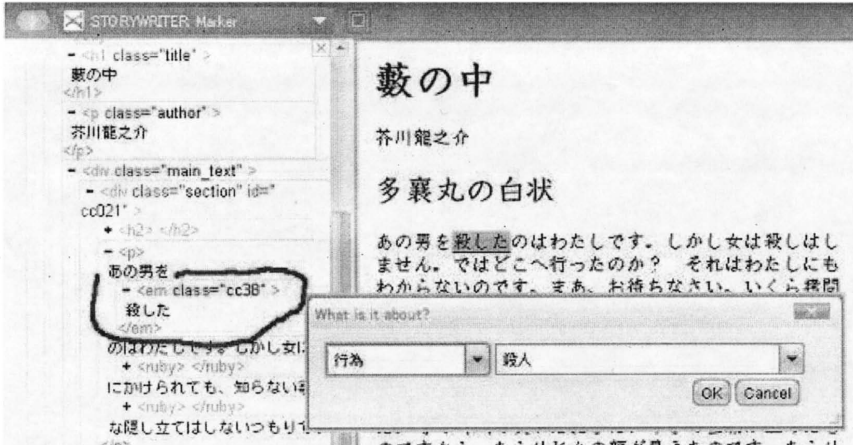


図5 マーカーはクラス属性で区別される

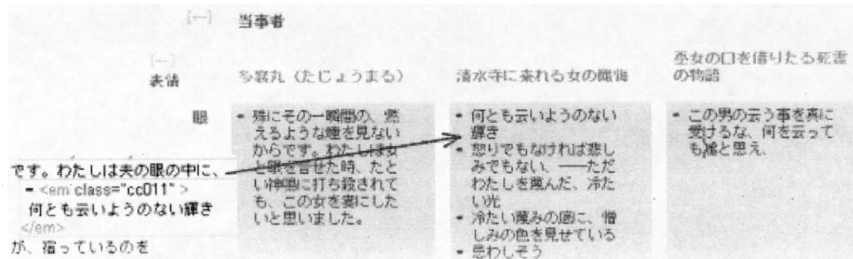


図6 STORYWRITER のクロスビュー

図6では、class 属性値が“cc011”である em 要素を抜き出して、表の同じ行に配置している。ここで、class 属性値が“cc011”である em 要素は、UI 上は“視線”というマーカーに対応している。

この例では、“視線”というタグがシステムに導入されたわけではない。STORYWRITER システムは“視線”の意味を知らない、そのような意味に基づく処理が実装されているわけではない。読者が“cc011”という class 属性値にどのような意味を重ねようとも、STORYWRITER は同じ動作をする。

すなわち、ここここは同じ意味だということだけを前提に、役に立つ文書操作機能を提供できた。それぞれの括りの意味は、ユーザーだけが知っている。それでも、利用者に便宜を提供することができている。

#### STORYWRITER 開発経緯

##### STORYWRITER の開発当初

は、“When”や“Where”といった特別な値が class 属性値として設定されていた。すなわち、時間や場所は普遍的で客観的だからビルトインマーカーとして用意して良いだろうと考えていた。

しかし、素材とした「藪の中」では、“午少し過ぎ”や“その内にやっど気がついて見ると”など、幅のある時刻や主観的な時間経過の記述がある。また、この物語そのものが、登場人物の供述の食い違いをポイントとする平行物語であり、それぞれの主観的な主張にこそ物語の本質があり、《読み》の面白さがある。客観的な時間を特別扱いしたところで、この《読み》に対して、さして寄与するところはないのではないかと。スケジューラーのような便利さを実現できないであろう。むしろ、システムが複雑になるデメリットが大きいと考えた。

#### 6. まとめ

この報告では、マークアップそのものに注目するという、われわれの取り組みの特徴を論じ、実装例を示した。

#### 考察

簡単な考察をもってまとめたい。

《マークアップそのもの》の効果は、自然言語における指示語、日本語の「こそあど」ことばの効果に似ている。すなわち、指示語は次のような特徴をもち:

- ・ 指示語そのものは、話し手や聞き手が了解しているよりも、かなり抽象的な意味しか持っていない。
- ・ 指示する対象は、話し手と聞き手の双方にとって明確であることを建前としている。

そして、次のような効果を持つ:

- ・ 短い・簡単な表現で膨大な内容や複雑な内容を話したり、検討したりできる。
- ・ ああでもない、こうでもないと何度も考え直すことができる。

われわれがやろうとしていることは、指示語の働きを、電子的・定性的な情報交換、情報利用に持ち込み、その品質を高めて効率を向上し、記録に残すものなのかもしれない。

文献

- [1] 山口琢: Web 文書の usability を高める"Slide Show for XHTML", 情報処理学会研究報告 Vol. 2006, No.83(20060728) pp. 55-58 (2006.7)
- [2] 小林龍生、山口琢: Parallel Narratology 試論: ハイパーテキストにおける相互参照の観点から、情報処理学会研究報告 Vol.2007, No.77(20070726) pp. 25-30 (2007.7)
- [3] 山口琢、新ワードプロセッシング: CrossConcept、Template It!の試作、情報処理学会研究報告 Vol.2007, No.77(20070726) pp. 31-35 (2007.7)
- [4] Taku Yamaguchi, Tatsuo Kobayashi: Implementing CrossConcept: A Computer Software Supporting Creative Thinking, IJCKS 2007 (KSS2007/KICSS2007) (2007.11)
- [5] 山口琢、小林龍生、野口尚孝: CrossConcept における概念操作モデルと知性・感性の工学的支援、第2回横幹連合コンファレンス (2007.11)
- [6] 大場 みち子、山口琢: プロセスとしてのドキュメンテーション、情報処理学会研究報告 Vol.2008, No. 10 (2008/1・2/31・1) pp. 131~137 (2008.1)
- [7] World Wide Web Consortium: Extensible Markup Language (XML) 1.1 (Second Edition) (2006.9)
- [8] 株式会社ジャストシステム: xiv Community

[9] 山口琢: STORYWRITER、<http://www.yamahige.jp/storywriter/>