

技術解説記事データベースのためのXML文書作成ツールの開発

4V-6

徳永秀和 青江順一

高松工業高等専門学校 徳島大学工学部

1. はじめに

ワードプロセッサの普及さらにインターネットの普及によって、電子化文書が急増している。しかし、この多量の文書を有効利用するためのシステムは、文書単位の検索として文献検索やホームページの検索がある程度である。一方、SGMLなど文書構造を利用しようとするものがあるが、印刷への応用が中心であり、文書構造を知識として積極的に応用するシステムはまだない。そこで、我々は、技術解説記事的を絞り、文書構造を利用し複数文書を統合して個人的知識として活用するための技術解説記事データベースの開発を行っている[1]。本論文では、基本的なアイデアと構造化文書を作成するためのツールについて述べる。

2. 技術解説記事の文書構造

文書の最も基本的構造は木構造である。各ノードはタイトルなどの属性を持つ。そして、文1つが葉となる。また、文書には様々なリンクが存在する。キーワード索引、原因結果や事例などのノ

ード間の関係などである。〔図1〕

3. 検索、個人知識化

検索する記事は、数十個に絞りこまれているものとする。これらの記事に対して、文書構造を利用した検索をする[2][3]。検索結果を利用して、新たな文書を作成する。この時、新たな文書にはもとの文書へアクセスできる情報を入れておく。この、新たな文書が個人的知識となる。

4. XMLによる文書の記述

もとの文書と検索により作成した文書を記述する方法としてXMLを利用する。XMLは構造化文書記述のSGMLとインターネット用文書記述のHTMLを統合した機能を持っており、木構造とリンクを簡単に記述する能力を持っている[4]。また、近い将来ホームページ記述言語として広まる可能性が高い。このような理由よりXMLを採用した。

文書管理の最小単位をparagraphとし、sectionの再帰構造で文書を表現する。最下層のsectionは複数のparagraphを持つことになる。

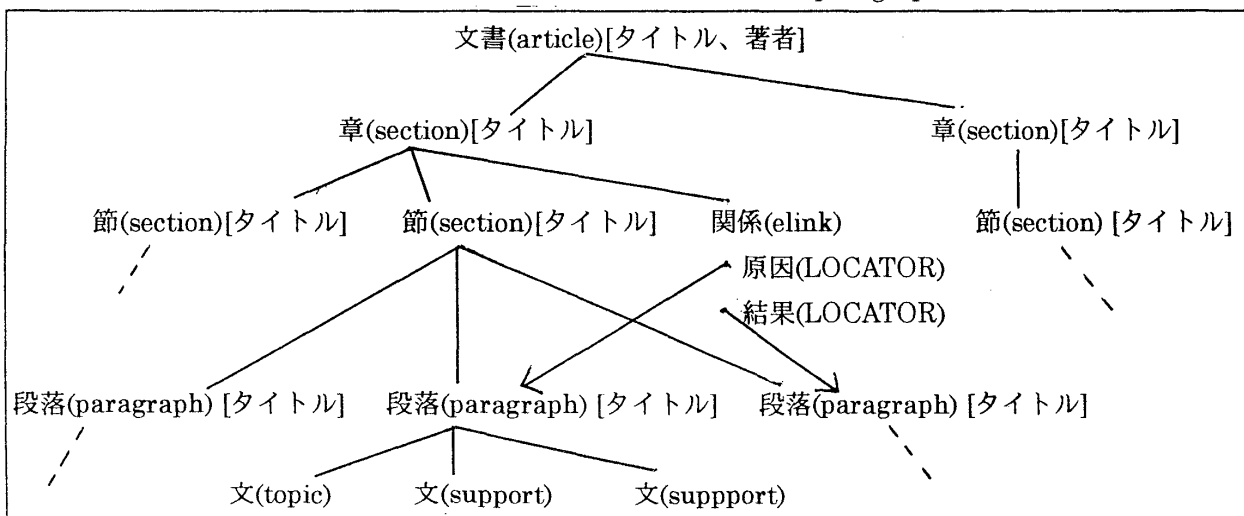


図1 文書構造

Paragraph は topic、support、result の文を持つ。また、リンクは、拡張リンクを用いてリンクを張るノードの親ノードに記述する。また、section のタイトルなどはノードの属性として持つ (図 1) (図 2)。

```
<?xml encoding="US-ASCII"?>
<!ELEMENT article ((section | elink)*)>
<!ATTLIST article
id CDATA #IMPLIED title CDATA #REQUIRED author CDATA #REQUIRED>
<!ELEMENT section ((section | paragraph | elink)*>
<!ATTLIST section
id CDATA #IMPLIED role CDATA #IMPLIED title CDATA #REQUIRED>
<!ELEMENT paragraph ((topic|support|result|elink)*>
<!ATTLIST paragraph
id CDATA #IMPLIED role CDATA #IMPLIED title CDATA #REQUIRED
mainW CDATA #IMPLIED subW01 CDATA #IMPLIED subW02 CDATA #IMPLIED>
<!ELEMENT topic (#PCDATA)>
<!ATTLIST topic id CDATA #IMPLIED>
<!ELEMENT support (#PCDATA)>
<!ATTLIST support id CDATA #IMPLIED>
<!ELEMENT result (#PCDATA)>
<!ATTLIST result id CDATA #IMPLIED>
<!ELEMENT elink (LOCATOR*)>
<!ATTLIST elink
id CDATA #IMPLIED
XML-LINK CDATA #FIXED "EXTENDED"
:
>
<!ELEMENT LOCATOR ANY>
<!ATTLIST LOCATOR
XML-LINK CDATA #FIXED "LOCATOR"
:
>
```

図2 文書用 DTD

5. 文書作成過程

文書を作成する過程は、トップダウンとボトムアップがある。トップダウンとは、文書タイトル、章タイトル、節タイトルという順に考えて行く過程である。ボトムアップとは、思いつく文を適当に書き、書き上げた文全体を眺め分類、修正して行く過程である。

6. XML 文書作成ツール

文書作成ツールは、トップダウンとボトムアッ

プによる文書作成を統合して行えるものを製作した。図 3 の上 3 つのボタンがトップダウンによる作成用であり、下 3 つのボタンがボトムアップによる作成用である。図 4 は、トップダウンにより paragraph を作成する画面である。図 5 は、ボトムアップにより paragraph を作成する画面である。なお XML パーサは IBM のものを使用している [5]。

7. おわりに

既存の文書を個人的知識として利用するための XML を利用した文書構造を示した。また、この XML 文書を作成するためのツールについて述べた。現在検索ツールを作成中であり、今後は本システムが文書の有効利用に役立つ事を示す実験を行って行く。

8. 参考文献

- [1] 有友雄祐、池田晃一、徳永秀和：XML による技術解説記事データベースの開発、平成 10 年電気関係学会四国支部連合大会 pp288, (1998)
- [2] R.Sacks-Davis, T.Arnold-Moore, and J.Zobel :Database system for structured documents, IEICE Transactions on Information and systems, E78-D(11), pp1335-1342 (1995)
- [3] W3C.Working Draft: XML Pointer Language (X Pointer)
http://www.w3c.org/TR/WD-xptr
- [4] W3C. Extensible Markup Language ;
http://www.w3c.org/XML
- [5] http://www.trl.ibm.co.jp/projects/xml

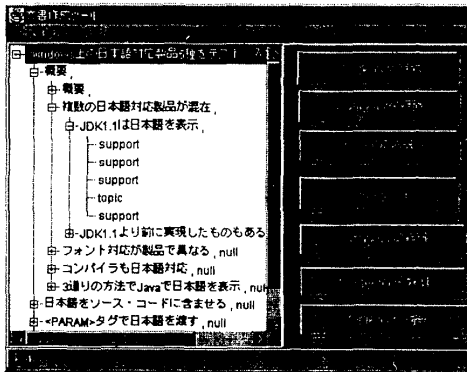


図3 文書作成ツール

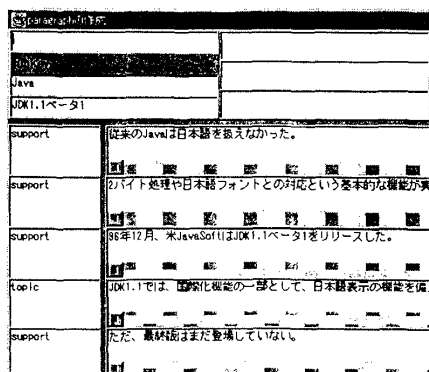


図4 パラグラフ作成(トップダウン)

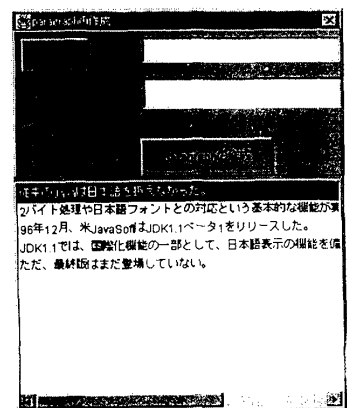


図5 パラグラフ作成(ボトムアップ)