文学データベースのための文書の構造化と意味管理

Structured Documents and Semantic Objects for Literature Databases

横田 一正　三宅 忠明　國島 丈生　劉 渤江*　田槙 明子†

Kazumasa Yokota, Tadaaki Miyake, Takeo Kunishima, Bojiang Liu, Akiko Tamaki

岡山県立大学 情報工学部

〒719-1197 岡山県総社市窪木 111

Faculty of Computer Science and System Engineering

Okayama Prefectural University

Soja, Okayama, 719-1197

キーワード: 文学データベース、構造化文書、XML、意味オブジェクト

Keywords: literature database, structured document, XML, semantic object.

あらまし: 電子化された文学作品をデータベースに蓄積し、情報検索や統計処理のツールを使って文学的考察を行うことは多く行なわれており、文学研究に大きく貢献してきている。しかし、さらに深く文学研究を進めるためには、構造的あるいは内容的関連を表すことが重要で、作品内の特定の部分を抽出したり、それらの間、さらには複数の作品間にリンク付けを行う必要がある。そのために、文書管理技術によって文学作品を構造化し、さまざまな意味情報を扱い、さらに文学研究の支援ツールを整備したシステムの研究開発を行っている。本稿では、ケルト文学の『デァドラ伝説』を対象に、XML による構造化文書の生成、総索引表(コンコーダンス)等の作成、構造化文書間のリンク付けなどをサポートしたプロトタイプシステムについて報告し、XML で記述できない利用者のオブジェクトの扱い、それらを含む管理機構について議論する。

Summary: There are many works for literature databases, in which elecroronic documents are stored and informaton retrieval and statistics tools are supported. However, for advanced literature research, it is important that structural or semantical similarities are represented. That is, any part in a document is specified as an object and linked to another one in the same or aother document. For the objectives, we have been developing a system based on some document management techniques, in which literature documents are structured, user-specified objects are defined, and various support tools for literature research are provided. In this paper, we report a prototype system for a literature database which includes varieties of a Celtic story, *The Dierdire Legend*. The system structures documents based on XML, generates various concordances, and supports link tools among multiple documents, including personal memos. Further, in this paper, we discuss representaton and management of user-specified objects, which cannot be described in XML, and their management.

1 はじめに

『デァドラ伝説』[15]は8世紀頃から伝わるケルト(アイルランド)文学の代表的伝承物語であるが、口承によって伝えられたため数多くの変種が存在している。岡山県立大学はその収集において世界的に見て最大規模のひとつであり、現在100編近くを所蔵している。岡山県立大学では上記のような分析・研究のために、1997年からそれらの電子化を行なっているが、当初は、それは出版を目的としたもの

*岡山理科大学総合情報学部

†(株) リョービシステムサービス

で、文学研究のためのデータベース化や支援システムの構築を目的としてこなかった。従来の文学データベースの研究では、電子化された文書を対象に、情報検索や統計処理技術によって、文学研究のための有用な情報を引き出すことがほとんどであり、データベースシステムの観点からのデータベース化は考慮されなかった。

比較文学の分野では複数の作品における類似性や語彙の変化に着目し、作品の系統樹を作成することが大きな課題となっている。とくに『デァドラ伝説』のように変種の多いものにとっては重要である。さらに、変種間やそれらの語彙間の差異を比較文学的に分析するすることによって、文化的あるいは政治的な背景を考察することができ、より深い文学的考察を行うことができる。このため、作品間の構造的類似性や内容的類似性を可能にするためには、構造化された文学データベースの構築と、それに基づく支援システムがより進んだ文学研究のために必要不可欠となってくる。そこで、これらをサポートした『デァドラ伝説』のデータベース化を行っている。

電子化文書を構造化するためには XML (eXtended Markup Language)[2][24] が一般的になってきており、本研究でも XML に基づくアプローチを採用している。従来構造化は文書の再利用を目的としたものが多いが、本研究で XML を採用した理由は、それがデファクトスタンダードになりつつあること以外に、以下のようなものである。

- 文書スキーマの定義
  文書に対する検索機能が必要になると考えられ、文書の意味的構造が明確になっていることが重要である。言い換えれば、文書スキーマが明確に定義できることが必要である。XML では、DTD (Document Type Definition) によって独自の文書型定義を行うことができる。

- リンク機能
  XML では、XLink[12], XPointer[13] という 2 つの規格によって、強力なリンク機能を提供している。具体的には、文書中の任意の粒度のオブジェクトをアンカーとできる、3 つ以上のアンカーに対してリンクを張ることができる、文脈外リンク (out-of-line link)[1]によって元の文書に手を加えずにリンクを張ることができる、などの機能を有用と考えている。

- インターネットでの技術的優位性
  HTML と同様、WWW での利用を考慮して作られた規格であり、Netscape Navigator や Internet Explorer などの主要な WWW ブラウザでの対応が期待できる。また、オフィスアプリケーションでも将来の XML 対応を公表しているものもあり、文書フォーマットとしてデファクト・スタンダードとなりうることが期待できる。

- レイアウト情報の定義
  見栄えにこだわった WWW ページが氾濫している現状から分かるように、電子文書といえども、それを扱うのが人間である以上、レイアウト情報を保持する機能は必要になる。XML では、CSS (Cascading Style Sheet)[23, 1], XSL[3] などのレイアウト定義言語を持ち、レイアウト情報を定義することができる。

XML で特定するオブジェクトが (表題、著者、章、段落など) が他の文書とのリンクの単位となると共に、検索等のデータ操作の単位ともなる。

本稿では、まず 2節でこれまでに実装したプロトタイプシステムの概要を説明し、3節から7節まではその中の個々のサブシステムについて報告する。8節でプロトタイプシステムの経験から今後の拡張点について、とくに XML の構文的制約で記述できないオブジェクトについて議論し、まとめる。

2 プロトタイプシステム

電子化されたテキスト文書は構造を持たないフラットファイルであり、そのままでは表層的な分析しか行なうことができない。そこで、深層的な分析が行なえるよう、個々の文献に対して基本的に、以下に示す 4 種類の構成要素を持たせている。

- テキスト文書
  個々の物語を分析や加工処理を施さず、単に電子化したフラットファイル。

- 構造化文書
  テキスト文書に構文的な概念を持たすために章、段落、文などに構造化したもの。構造化は XML(eXtensible Markup Language) に用いられるようなタグを挿入することで行う。この構造化情報により文章の位置を特定することができる。

[1] XML に関する技術用語の日本語訳は [24] によった。

- 内容記述
  物語のストーリを個々の場面に分割し時系列で結び、トップダウンで記述したもの。これは抽象化のレベルにより複数の記述がありうる。

- メモ文書
  原テキストに対応させてユーザ独自のメモを記述したもの。

内容記述もメモもXMLで記述されるので、共通のリンク機能によって、関連付けることができる。つまり、リンク機能をサポートした問合せ機能は、元の文書、内容記述文書、メモ文書（個人文書）を関連付けながら検索することが可能になる。

上記の4種類の文書をもつ文学データベースのためのプロトタイプシステムは、図1に示すように、構造化システム、索引抽出システム、検索システム、内容記述システム、メモ記述システムの5つのサブシステムから構成される。ここでは内容記述システム以外の4つについて説明する。

3 構造化システム

プロトタイプシステムの構造化システムでは、最小構成要素を文書の文単位として構造化を行なうこととした。文書を構造化するための構造識別子は、表1に示す7つである。

表 1: 各構成要素の構造識別子

構造識別子	構成要素
⟨title⟩	表題（著者等も含む）
⟨chapter⟩	章名
⟨section⟩	節
⟨paragraph⟩	段落
⟨sentence⟩	文（韻文も含む）
⟨stanza⟩	韻文
⟨source⟩	出典

一般の構造化文書は再利用を主たる目的としたものが多く、構造化の粒度はもっと大きなものがとられるが、文書の構文的条件を考慮した検索やリンクの粒度等のために、文単位としている。また『デァドラ伝説』は散文だけではなく、韻文や戯曲、それらの混在したものなど、さまざまな形で書かれているが、構造化するためには、構文の文法記述が必要で、プロトタイプシステムではその単純化のために、散文と韻文を含む散文にとどめることにした。

このシステムにより生成される構造化文書は、図2のようにXMLに基づいたタグを構造識別子としてテキスト文書の各構成要素の前後に挿入することで構文的概念を付加したものである。

```
⟨section number="1"⟩
⟨paragraph number="1"⟩
⟨sentence number="1"⟩
What caused the exile of the sons of Uisliu?
⟨/sentence⟩
⟨sentence number="2"⟩
It is soon told.
⟨/sentence⟩
⟨/paragraph⟩
⟨paragraph number="2"⟩
⟨sentence number="3"⟩
The men of Ulster were drinking in the house of Conchobor's
storyteller, Fedlimid mac Daill.
⟨/sentence⟩
⟨sentence number="4"⟩
Fedlimid's wife was overseeing everything and looking after them all.
⟨/sentence⟩
.....
```

図 2: 構造化文書

文書からの各構成要素の切り出しは、lexにより字句解析ルーチンを生成することにより行なっている。切り出し規則は、文書内の空白や改行、ピリオドなどのインデント情報を元にしている。ただし、対象データとして扱っているデァドラ伝説には、文献によって韻文が含まれるなど文書ごとに論理構造が異なっているため、構造化のための文法記述はある程度文書に依存した形となっている。このシステムで使用しているlexは、入力文字列が2つ以上の規則に適合したり、規則に適合する文字列がいくとおりも存在するような仕様を含むファイルにも字句解析ルーチンを生成する。このとき、字句解析ルーチンは

- 適合する文字数が最大になる規則を優先する
- 適合する文字数が同じ場合には、仕様書ファイルの先に書かれている規則を優先する

という2つの暗黙の選択基準のいずれか1つを採用することにより、トークンの切り出しを行なっている。

このような性質を持つ字句解析ルーチンでは、入れ子状態にある節、段落、文、韻文の規則を一度に定義した場合、字句解析ルーチンは前者の選択基準を採用してしまい、結果として文のみをトークンとして切り出すことになるた

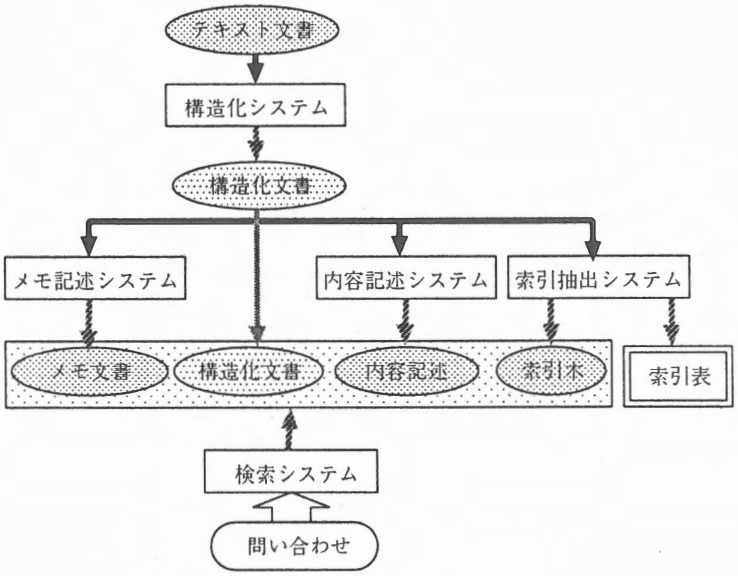

図 1: 文学データベースシステム

め、トークンの切り出しは以下に示す4段階に分け、テキスト文書の構造化を行なう。

1. 電子化されたテキスト文書を入力文字列として、表題、章名、韻文、出典とその他の5つのトークンに切り出し、その他以外のトークンの前後にタグを挿入する。

2. 第1段階で生成されたデータを入力文字列として、節と第1段階でタグの付けられた表題、章名、出典をトークンとして切り出し、切り出された節の前後にタグを挿入する。このとき、節には第1段階でタグの付けられた韻文が含まれていることもあるということを規則として書いておく。

3. 第2段階で生成されたデータを入力文字列として、表題、章名、節のタグ、出典と段落をトークンとして切り出し、切り出された段落の前後にタグを挿入する。ここで、段落に韻文が含まれる可能性があることを規則として書いておく。

4. 第3段階で生成されたデータを入力文字列として、表題、章名、出典、節と段落のタグと文に切り出し、切り出された文の前後にタグを挿入する。ここで、文に韻文が含まれる可能性があることを規則として書いておく。

図3に構造化システムのインタフェースを示す。

この構造化システムで生成される構造化文書には一文毎にタグが付けられているので、必要な情報の存在する位置の論理的な特定を行うには良いのだが、フラットファイルとしてみた時に構造識別子という付加情報が含まれており、図2のように構造化文書だけを見て文章を理解するのはなかなか難しい。そこで、構造化文書から構造識別子を抜きとり成形し、図4のようにテキスト文書として表示するブラウザの構築も作成している。これにより、構造化文書一つでテキスト文書の役割も果たすようにもなる。

4 メモ記述システム

文書が紙に書かれている場合、文書中にユーザが独自のメモを書き込むということがあるが、これに近い機能を電子文書上でも行なえるようにするシステムがメモ記述システムである。

メモ記述システムでは、先に述べた構造化文書ブラウザを利用して、構造化文書からでもタグを取り除いたテキスト文書からでも、メモを付加したい任意の文字列を指定することが可能である。、指定した文字列に対してメモ文書を作成し元文書とリンクを張ることでメモを付加していく。

任意の文字列をメモを付加するオブジェクト(メモオブジェクト)として指定する方法として以下の2つが考えられる。

- メモオブジェクトとしたい任意の文字列にタグを付けることで構造化文書を変更する
- 構造化文書から指定したメモオブジェクトの位置情報を抽出する

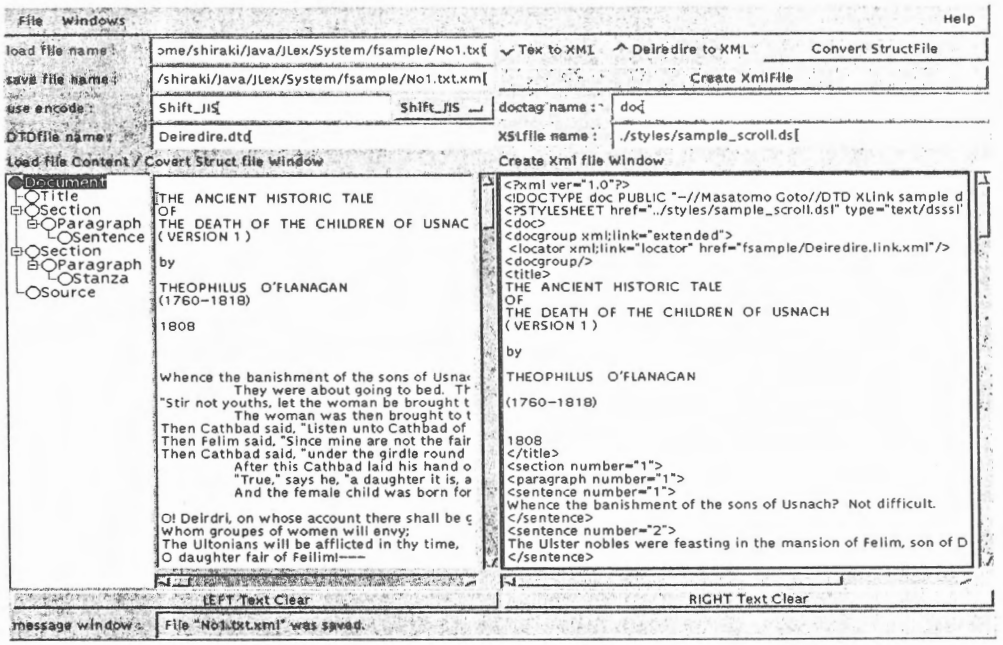

図 3: 構造化システムの実行例

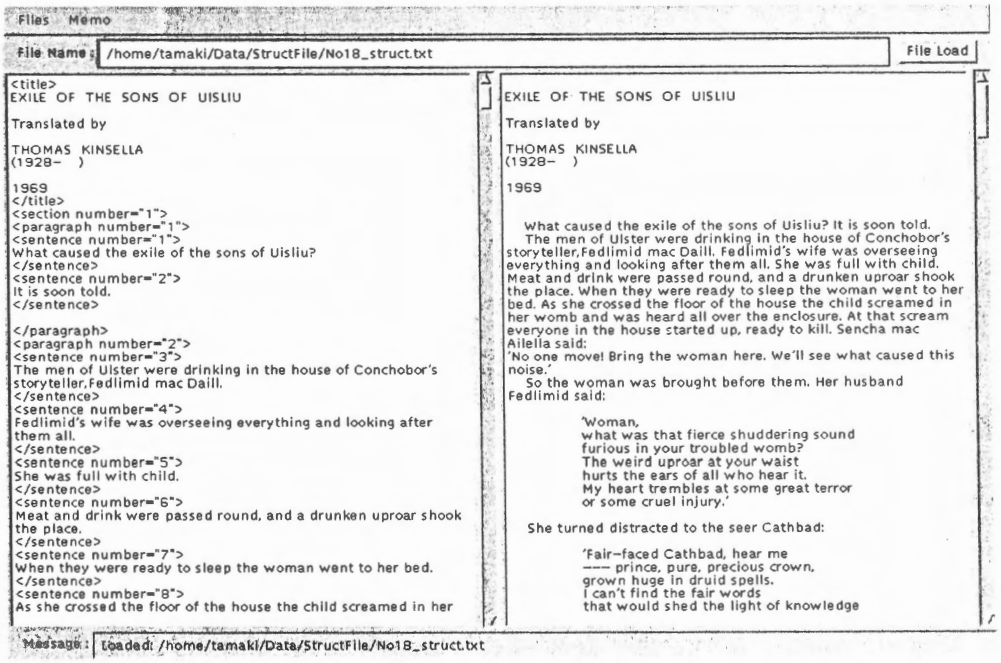

図 4: 構造化文書ブラウザ

前者のタグ挿入形式でメモオブジェクトを指定した場合、構造化文書にタグが直接書き込まれるため構造化文書が変更される。また、メモオブジェクトとして指定した任意の文字列が既存のメモオブジェクトと重なるとき、メモオブジェクトが入れ子状態になっていれば良いのだが、少しずれた形で指定されている場合、メモオブジェクトを図5のように2つに分割してタグを付けるなどの処理が必要となる。ただし、

```
× <para>...<kw>...</para>
     <para>...</kw>...</para>
○ <para>...<kw>...</kw></para>
     <para><kw>...</kw>...</para>
```

図 5: アンカーの分割

これはシステム側の処理上の問題であり、ユーザには関係のないことであるから、実際に表示させる時には分割されたメモオブジェクトを、あたかも一つのものであるかのように表示しなければならない。メモオブジェクトの前側に挿入される開始タグには、各メモオブジェクトに対するメモIDとユーザIDが情報として付加される。

後者の位置情報によるメモオブジェクト指定方式では、構造化文書からメモオブジェクトの存在する位置情報を見るだけなので、構造化文書が変更されることはないが、ユーザが文書に直接変更を加えた場合、メモオブジェクトの存在位置を指定しているので、位置情報にも変更を加えなければならなくなる可能性がある。本研究で処理対象データとして扱っているデァドラ文書は、読み込み専用データであるので、構造化文書を参考文献として位置情報を取り出す形式を採用することとした。

実際に文書にメモを付加するには、構造化文書ブラウザ上で任意の文字列を選択するか、もしくはメモ記述エディタに直接メモオブジェクトを指定する。ただし、構造化文書のタグは構成要素の判別のために付加された構造識別子なので、現在の実装では、それをメモオブジェクトとすることはできない。また、メモ記述エディタからメモオブジェクトを指定した場合、指定文字列が構造化文書内に複数箇所ある場合はユーザに確認を取る。指定がなければ最初に現れた文字列の位置がメモオブジェクトの位置情報として記録される。また、メモは基本的に私的なものであるから、各ユーザが付加したメモ文書のみを表示させるようにする必要がある。既存のメモ文書は特定のウィンドウより開いたり、必要でなくなったメモの削除を行なうことができる。

5 リンクエディタ

複数の構造化文書を関連付けるため、構造化システムの一部としてリンクエディタがある。これは、文脈外リンクを用いることにより、元文書に手を加えることなく管理することを可能にした。文脈外リンクを用いた理由は、必ずしも文書が書き込み可能ではないからである。リンクの定義がどの異種文書からも参照できる必要があるので、図6のような拡張リンクの定義のみを集めたXML文書(リンク定義文書)を文書群ごとに一つ設け、文書群中の各文書から参照するという方法をとった。この方法を用いると、すでに存在する文書群内のリンクの定義の変更はリンク定義文書に対して行えばよく、元文書に手を加えなくてもよい。

内容的に関連のある文書をひとつのまとまりとして定義することができる。つまり一般的に、XMLの構造化文書を含むハイパーテキスト文書の場合、たとえ内容が同一であっても設計方法は多様で、図7のように結果の文書集合の同定は構文的には困難である。

そこで同一内容の文書集合を文書群として定義している。つまり文書群とは、内容に対応した論理的な単位であって、図8のようにひとつの文書が複数の文書群から共有されても、入れ子構造になっていても構わない。この文書群は、文書単位のリンク付けで十分で、操作の基本的な単位となる。

6 索引抽出システム

索引抽出システムでは、文献中で使用されている語彙の使用例や使用頻度を知ることで比較文学研究の促進を図ったり、検索の高速化や高度化を図るためのデータを作成する。このシステムは、文書より使用されているすべての単語を切り出すキーワード抽出システムと総索引表(concordance)を生成する索引表作成システムの2つのサブシステムから構成される。

実際に比較文学研究において索引を利用しようとする場合、抽出語や不要語はユーザや使用目的によって異なる。また、文献によって、使われている語が同じとは限らず、すべての不要語をキーワード抽出システム内で自動的に削除してしまうというのは難しい。そこで、単語の切り出しは文書の題名、章名、出典を除く構成要素を対象に、前置詞や冠詞等も区別なくすべての単語を切り出し、切り出された単語から外

```
⟨docgroup xml:link="extended"⟩
⟨location xml:link="locator" href="main.xml"/⟩
⟨location xml:link="locator" href="www.xml"/⟩
⟨location xml:link="locator" href="ohp.xml"/⟩
⟨docgroup/⟩
⟨relation xml:link="extended"⟩
⟨point xml:link="locator" href="main.xml#root().child(1, section)"/⟩
⟨point xml:link="locator" href="www.xml#ID(section_2)"/⟩
⟨point xml:link="locator" href="ohp.xml#span(ID(p3),ID(p5)"/⟩
⟨relation/⟩
```

図 6: 拡張リンクによる関連付けの例

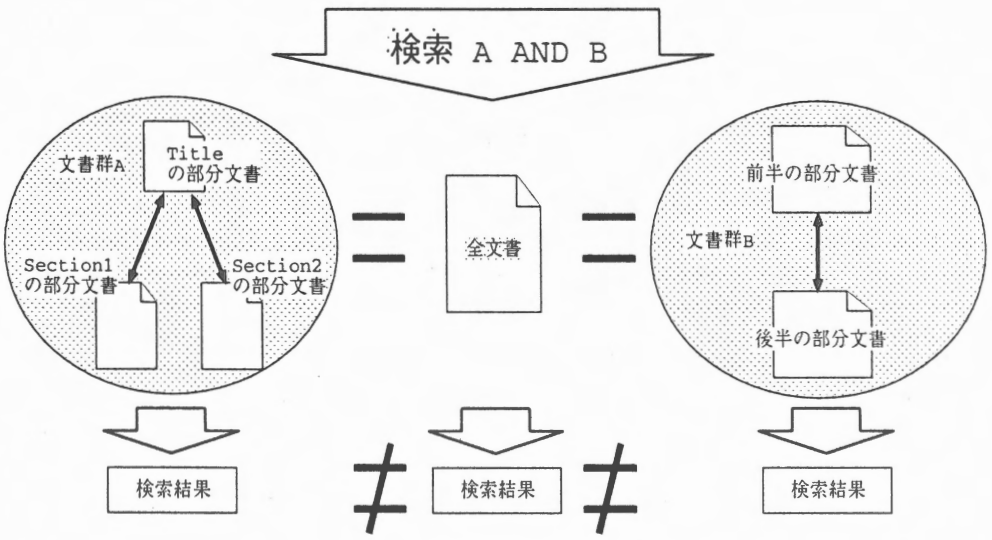

図 7: 同一文書の検索結果の相違

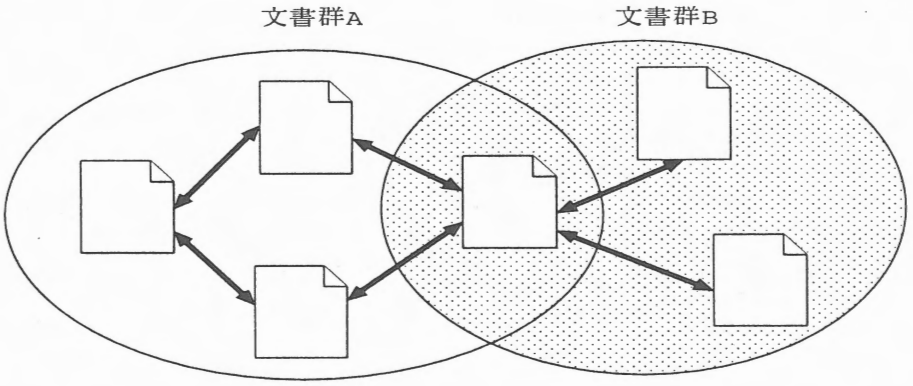

図 8: 文書関連・文書群の概要

部プログラム（索引表作成システムの前処理）によって目的に応じた語を抽出したり、不要語をユーザが削除することで修正キーワードファイルを作成することとした。内容記述のためには固有名詞の切り出しも行う [5]。なお、単語の切り出しは先の構造化システムでも使用した lex を使用し行なっている。

索引表作成システムでは、基本的にキーワード抽出システムにより作成された修正キーワードファイルを用いて総索引表を作成する。生成される索引表には、キーワードとその前後の文脈を表示させる索引表と検索・内容記述システムにおける処理対象データとして使用する索引表の 2 種類がある。前者の索引表は、一般に使用される KWIC 索引と KWOC 索引の両者の特徴をいかしたたもので、出現頻度表の生成も行なわれる。これは図 9に示すように、出現頻度とキーワードを使用頻度が高い順に並べて変えたのもで、これにより、どのキーワードがもっとも多く使用されているかが視覚的にわかるようになる。頻度表のキーワードをクリックすることにより、それに対応する KWIC 索引を表示できる。ここで、ロケーション情報とは一番右端にある”(2,30,132,20)” のような情報のことで、表示されている数字はキーワードが存在する節、段落、文、韻文中にキーワードが存在する場合は韻文の番号で構成される。このロケーションは最初は文単位に作成されているので、それを操作することで、パラグラフ単位や節単位などで表示することも可能である。

さらに、異なる文書から抽出された出現頻度表ををマージすることによって文書間の比較を容易に行なうことができるようになる。実際にマージを行なった例を図 10に示す。

これによって『デァドラ伝説の』のさまざまな変種を比較することにより、文学的考察を深

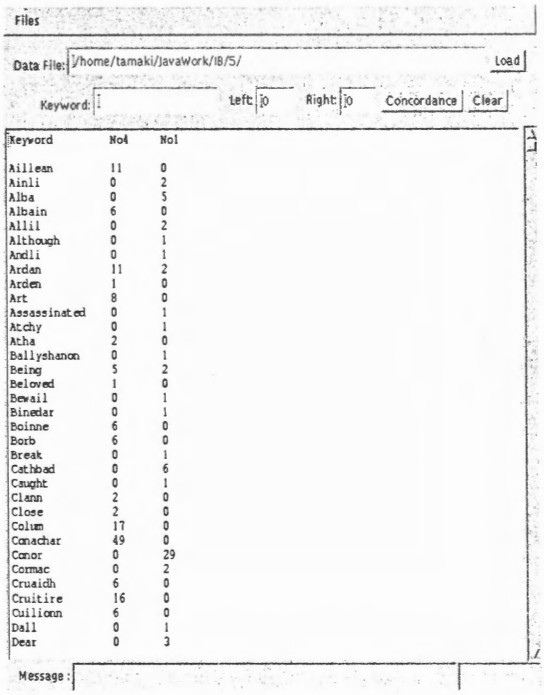

図 10: マージした出現頻度の表示

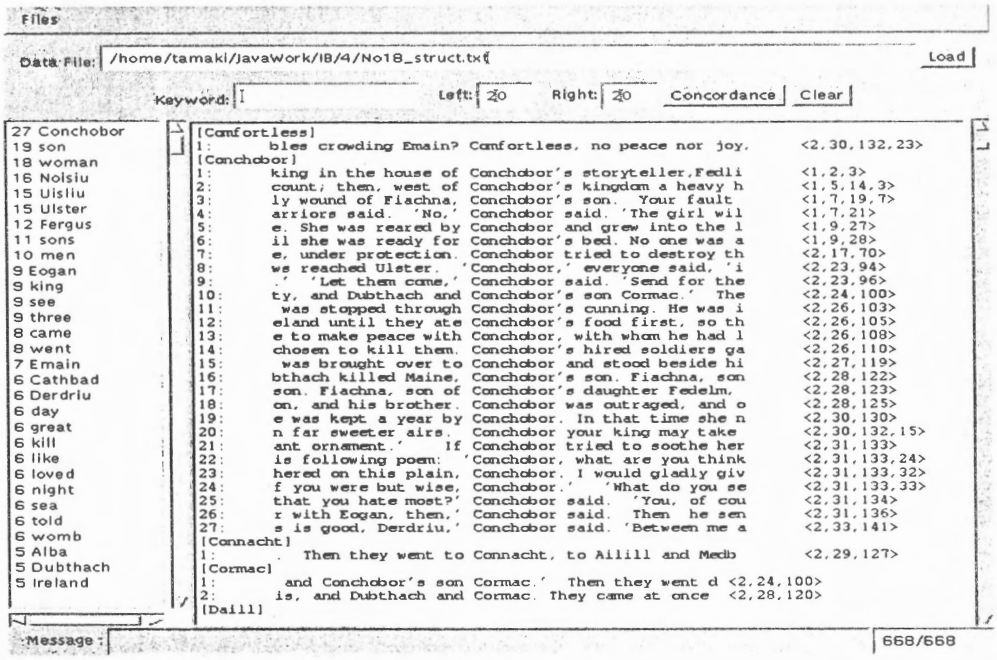

図 9: 出現頻度の表示

めることもできる [16, 17]。

7 検索システム

検索処理はまず、索引抽出システムから出力されたキーワード表から、必要となるキーワードを抽出し、索引木 (B+ 木) を生成している。従来の情報検索システムと異なるのは検索結果である。

単独の文書を対象にした結果得られるのは文書の部分構造の集合である。XML で構造化された文書は木構造になっており、ロケーションは木の葉 (文) を指している。B+ 木はこのノードアドレスを指しており、ロケーションの階層を操作することによって、同一文章中、同一パラグラフ中などの構文的条件を検索条件に付加できる。したがって検索結果の構造のレベルは検索条件によって異なっている。基本的な操作は以下のようになっている。

- 縮退: **Reduce** Loc **by** i

  $Loc = [l_1, l_2, \cdots, l_n] \Rightarrow [l_1, l_2, \cdots, l_{n-i}]$

- ロケーション検索:
  Select Loc **where** ⟨tag⟩ が C を満足する

  $\{Loc \mid D$のLoc'の$\langle tag \rangle$が
  (Cから得られた)C'を満足し、
  Locは縮退操作によって得られ、
  Cを満足する$\}$

- 文書検索: **Select** Doc **where** ⋯

  Reduce ロケーション検索の結果
  by 文書識別子

文書群は文書の集合として定義されるので、文書群に対する検索では、

- OR 条件では得られた文書を含む文書群の探索
- AND 条件では、AND を除去して文書群定義に含まれた文書を検索した結果、再度 AND の評価をおこなう

の 2 種類の操作が必要となる。基本操作は以下の 4 種類である。

- **Select** 文書群 G **where** C

  $\{G|G$のDがCを満足する$\}$

- **Select** 文書群 G **where** $C_1 \wedge C_2$

  $\{G|C_1$を満足するD_1とC_2を満足するD_2を含む$G\}$

- **Select** 文書群 G **where** $C_1 \vee C_2$

- **Select** 文書群 G **where** $\neg C$

文書群は一般的には検索条件を満足する極小集合の定義としている。

4節で述べたメモを含む検索も、この文書群としての検索となる。さらに 3節で述べたよう

に、『デァドラ伝説』は構造化文書やメモ文書の他に、抽象階層をなす内容記述の文書を持っている。この内容から元文書の構造を引き出すのもこの文書群検索がおこなう。

このように文書群の操作はさまざまな内容を持っている。つまり

- デァドラのようにさまざまな変種を統合すること
- 内容的に関連がある複数のデータ形式の電子化文書を論理的に統合する
- ハイパーテキストの設計の多様性に対して、内容的なまとまりを定義する
- 電子化文書の更新履歴をまとめて管理する
- メモや内容記述のような付加情報をユーザごとに統合する。

などである。

8 構造化文書の拡張

プロトタイプシステムでは多くの機能の積み残しがある。それらは以下のものである。

- 文書の再帰構造
- 構造の細分化
- 意味オブジェクトの拡張
- ラッパーとタグのフィルタリング
- データベースエンジンとの結合

構造化文書は一般的に再帰構造になる。これはロケーション情報を木にもつ索引木と、構文条件をロケーション情報で操作する縮退操作に影響する。そのためにロケーション情報に構文情報 (タグ情報) を付加するよう拡張することで解決できる。ただし問題がシステム的に解消できたとしても、「同一文章中にキーワード A と B を含む」という検索条件には多義的となるので、ユーザはそれを知らなければならない。

現在の構造化は文レベルであるが、さらに自然言語処理技術を使って詳細化することも検討している。一般的に『デァドラ伝説』のように長文の物語の内容記述は困難であるので、現在の内容記述は、場面の時系列として物語を構成し、各場面に登場人物や場所を付加するに留めている。場面は抽象化できるので抽象階層としても扱っている。因果関係など人が記述しても困難なものも多いが、意味内容を反映したオブジェクトを指定できれば、より深い内容検索を行うことができる。

構造化文書にメモを付加したり、内容記述文書等の対応をとることはプロトタイプシステムでも行っているが、それは元文書が更新されないという強い仮定に基づいていた。さらに一般的に意味内容を考えたリンク付けを行うためには、XML による構文的オブジェクトと意味オブジェクトの有機的なリンク付けが必要である。またいったん意味オブジェクトを導入すると、意味オブジェクトに対する索引木や検索が必要となる。そのために現在、意味オブジェクトの記述言語として、演繹オブジェクト指向パラダイムに基づいた知識表現言語 QUIK [?, 19, 25, 10] を予定している。現在記述実験を行っており、両オブジェクト間の意味管理やリンク管理の機能要件を明確化している。意味オブジェクトは、電子化文書の個別化や情報統合、異種文書管理で重要な役割を果たす [7, 8, 10]。これらの問題を解決するため、図 11のような統合管理システムを構想し、研究を進めている。

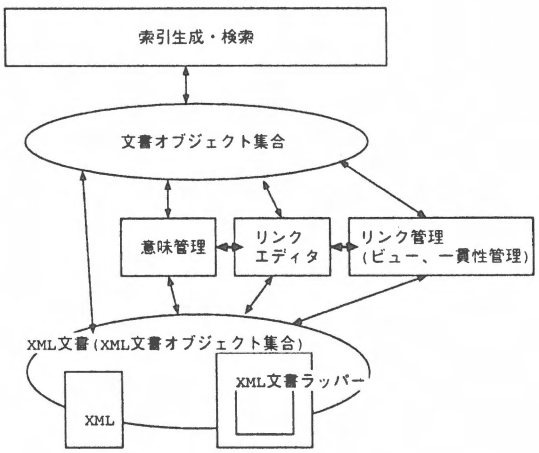

図 11: 統合管理システムの概要

現在は対象とする文書の種類が少ないが、『デァドア伝説』関係の解説や付属資料も付加するとすれば、さまざまなデータ形式を対象としなければならない。そのためには元のデータ形式を XML 形式に変換するラッパーの位置付けが重くなってくる。そのときは、変換を可逆性を保証するためのタグのフィルタリングを考えている。これはプロトタイプシステムでの構造化文書ブラウザの一般化でもある。

今回のプロトタイプシステムではデータベースエンジンとの結合まではできなかった。現在実験を開始しているのは、ObjectStore PSE Pro (for Java) である。システムはすべて

Java で実装されているので、Java オブジェクトとして蓄積し、管理する予定である。意味オブジェクトは QUIK オブジェクトとなるので、QUIK と Java の連携機能も現在実装中である。

本論文では、ケルト文学の『ディドラ伝説』を対象にした文学データベースのプロトタイプシステムについて述べ、その拡張として、さらに文学研究を進めるための課題について議論した。本研究の特徴は以下のようにまとめることができる。

- XML による構造化文書を対象にした文学データベースを指向している。
- リンク機能により、関連文書や文書群にたいする操作を実現した。
- 意味オブジェクトを導入することにより、電子化文書の管理機構を拡張している。
- 文学研究の要件を反映したツールを整備しつつある。

このプロトタイプシステムの経験から、いくつかの新しいシステムを進行中である。本稿の文学データベースの文学研究向けの強化、民間説話データベースの構築、マルチメディア情報を含む異種文書管理システム、電子化文書の効果的提示システム（ディジタルライブラリ、ディジタルミュージアム、マルチメディアブック、戯曲の視覚化）、ディジタルテーマパーク、QUIK の改良、などであり、現在平行して研究開発を進めている。

謝辞

さまざまな議論を頂く岡山県立大学横田研究室の皆様に感謝します。なお、本研究の一部は文部省科学研究費特定領域研究 (A)「高度データベース」および基盤研究 (C) による。

参考文献

[1] Bert Bos and Hakon Wium Lie and Chris Lilley and Ian Jacobs: "Cascading Style Sheets, level 2", *W3C Recommendation REC-CSS2-19980512* (May 1998)
available from http://www.w3.org/

[2] Tim Bray and Jean Paoli and C. M. Sperberg-McQueen: "Extensible Markup Language (XML) 1.0", *W3C Recommendation REC-xml-19980210* (Feb, 1998)
available from http://www.w3.org/

[3] James Clark and Stephen Deach: "Extensible Stylesheet Language (XSL)", *W3C Working Draft WD-xsl-19981216* (Dec, 1998)
available from http://www.w3.org/

[4] 八村広三郎: "人文科学とデータベース", 「情報処理」 Vol.38 No.5 通巻 387 号, 5 月, (1997).

[5] 本行弘明、池口仁誠、三宅忠明、横田一正: "分散環境での文学データベースの内容検索", 第 55 回情報処理学会全国大会, 九州, 9 月, (1997).

[6] 本行弘明、白木善隆、田槙明子、国島丈生、横田一正: "文学データベースのためのプロトタイプシステムの実装", 第 57 回情報処理学会全国大会, 名古屋, 10 月, (1998).

[7] 國島丈生、横田一正、白木善隆、劉渤江: "XML リンク機能による異種文書の統合方式", 情報処理学会データベースシステム研究会資料 vol.117, No.7, 1 月 (1999).

[8] Takeo Kunishima, Kazumasa Yokota, Bojiang Liu, and Tadaaki Miyake: "Towards Integrated Management of Heterogeneous Documents", *Cooperative Databases and Applications '99*, pp.39-51, Springer, Sep., 1999

[9] Bojiang Liu, Kazumass Yokota, and Nobutaka Ogata: Specific Features of the QUIK Mediator System, *IEICE Transaction on Information and System*, Vol. E82-D, No.1, pp.180-188 (1999).

[10] 劉渤江、横田一正, 國島丈生, 三宅忠明: "異種文書の統合のための意味とリンクの管理機構"、電子情報通信学会データ工学ワークショップ、5B-2, pp.1-8, 鹿児島、1999 年 3 月.

[11] W. Li, and Y. Wu: "Query Relaxation by Structure for Web Document Retrieval with Progressive Processing," *Proc. Advanced Database Systems Symp.*, pp.19-25 (1998).

[12] Eve Maler and Steve DeRose: "XML Linking Language (XLink)", *W3C Working Draft WD-xlink-19980303* (Mar, 1998), available from http://www.w3.org/

[13] Eve Maler and Steve DeRose: XML Pointer Language (XPointer), *W3C Working Draft WD-xptr-19980303* (Mar, 1998), available from http://www.w3.org/

[14] Dennis R. McCarthy and Umeshwar Dayal: "The Architecture of An Active Data Base Management System," *Proc. ACM SIGMOD International Conference on Management of Data*, pp. 215-244 (June. 1989).

[15] Tadaaki Miyake: *Select Version of DEIRDRE*, University Education Press, May, (1998).

[16] 三宅忠明、横田一正、國島丈生、田槙明子: "ケルトの悲恋ロマンス「デァドラ」における資料の言語分析", 人文科学とコンピュータ 99 シンポジウム、pp.103-104、大阪 (1999).

[17] 三宅忠明、横田一正: "情報工学的アプローチによる文学研究," 英語青年, Feb. (2000).

[18] 村田真. XML 入門. 日本経済新聞社, Jan. (1998).

[19] Toshihiro Nishioka, Kazumasa Yokota, Chie Takahashi, and Satoshi Tojo: "Constructing a Legal Knowledge-base with Partial Information", *Proc. ECAI'94 Workshop on Artificial Normative Reasoning* pp.40-55, Amsterdam (Aug., 1994)

[20] 大田友一、横田一正、西田豊明、佐藤哲司: 情報の共有と統合、岩波講座マルチメディア情報学、7 巻、岩波書店, 12 月 (1999).

[21] 白木善隆、本行弘明、田槙明子、国島丈生、横田一正: "異種文書データの統合管理システムの構想", 電気・情報関連学会中国支部連合大会, 岡山, 10 月 (1998).

[22] 田槙明子, 本行弘明, 白木善隆, 國島丈生, 横田一正: "文学データベースのための索引・検索機能の拡張", 電気・情報関連学会中国支部連合大会, 岡山, 10 月 (1998).

[23] Hakon Wium Lie and Bert Bos: "Cascading Style Sheets, level 1", *W3C Recommendation REC-CSS1-961217* (Dec, 1996)
available from http://www.w3.org/

[24] XML/SGML サロン: 標準 XML 完全解説, 技術評論社, 5 月 (1998).

[25] 横田一正: "演繹オブジェクト指向データベース Quixote の法的推論への応用", 人工知能学会誌, vol.10, no.1, pp.24-30 (1995).

[26] Kazumasa Yokota, Yutaka Banjou, Takashi Kuroda, and Takeo Kunishima: "Extensions of Query Processing Facilities in Mediator Systems," *Proc. Int. Workshop on Knowledge Representation Meets Databases (KRDB'97)*, p.17.1-8 (Aug., 1997).