

## 電子論文における内容的階層構造に基づく 再生および編集機能に関する検討

高屋 和幸<sup>†</sup> 笠井 裕之<sup>†</sup> 児玉 明<sup>‡</sup> 富永 英義<sup>†</sup>

<sup>†</sup> 早稲田大学 理工学部 電子・情報通信学科

〒169-8555 東京都新宿区大久保 3-4-1

TEL : (03) 5286-3385

FAX : (03) 3200-6735

<sup>‡</sup> 広島大学 地域共同研究センター

〒739-0046 東広島市鏡山 3-10-31

TEL : (0824) 21-3646

FAX : (0824) 21-3639

E-mail : takaya@tom.comm.waseda.ac.jp

あらまし 本稿では、既存の電子論文に対する要求条件・機能を挙げ、それらを実現するための内容的階層構造を持つ電子論文について検討を行った。特に、簡易再生機能および簡易電子論文編集機能に着目し、両機能を内容的階層構造に基づいて構成されたストリーム情報から実現する際に必要な情報および処理について述べた。さらに、内容的階層構造を持つ電子論文作成処理についても併せて述べた。内容的階層構造を持つ電子論文を実際に作成し、簡易再生機能および簡易電子論文編集機能の実装を行った。

キーワード 電子論文, 内容的階層構造, 簡易再生機能, 簡易電子論文編集機能

## A Study on the Process of Playing and Editing for Electric Thesis with Hierarchical Structure based on Contents

Kazuyuki TAKAYA<sup>†</sup> Hiroyuki KASAI<sup>†</sup> Mei KODAMA<sup>‡</sup> Hideyoshi TOMINAGA<sup>†</sup>

Dept. of Electronics, Information and  
Communication Engineering, WASEDA University

3-4-1 Ohkubo Shinjuku-ku, Tokyo,  
169-8555 JAPAN

TEL : +81-3-5286-3385

FAX : +81-3-3200-6735

Center for Technology Research and Development,  
HIROSHIMA University

3-10-31, Kagamiyama, Higashi-Hiroshima,  
739-0046 JAPAN

TEL : +81-824-21-3646

FAX : +81-824-21-3639

E-mail : takaya@tom.comm.waseda.ac.jp

**Abstract** In this paper, we explain the electric thesis which has hierarchical data structure based on contents. At first, we state requests for existing electric thesis, and the functions to play and edit for the content based hierarchical electric thesis. In addition, we explain the processing algorithm to play the electric thesis simply, and to edit simple electric thesis. At last, we made the hierarchical electric thesis with making algorithm we stated.

**key words** Electric Thesis, Hierarchical Structure, Simple Playing, Elementary Electric Thesis Editing

## 1. はじめに

本稿では、内容的階層構造を持つ電子論文について検討する。特に、簡易再生機能および簡易電子論文編集機能に着目し、両機能を内容的階層構造に基づいて構成されたストリーム情報から実現する際に必要な情報および処理について述べる。また、内容的階層構造を持つ電子論文を実際に作成し、簡易再生機能および簡易電子論文編集機能の実装を行う。なお、紙媒体による論文は文章領域と図表領域で構成されているが、本稿で扱う電子論文はテキスト情報と静止画像情報により構成されているものとする。

## 2. 要求条件

本節では、電子論文への要求条件・機能を挙げ、機能実現に必要な情報構造について説明する。

### 2.1 要求機能

従来の電子論文は、紙媒体の論文をスキャナから入力して作成した文書画像で構成されているものが大半を占めている。この場合、利用者が論文を閲読する際の全ての処理は文書画像単位で行なわれる。従って、論文中から必要な情報部分のみを見たり取り出したりする場合に多くの処理を必要とする。

一方、フルテキスト型の電子論文も存在するが、テキスト情報が構造化されていないために、必要部分だけを見たり取り出したりする場合には利用者の手作業が必要である。

このような状況を回避するためには、利用者が必要な情報をより容易に取得・再生および編集・再利用できる機能が必要である。これらの機能を実現するためには、機能実現性を考慮して情報を予め構造化しておく必要がある。本稿では、前者の機能を簡易再生機能、後者の機能を簡易電子論文編集機能と呼び、以降の検討を進める。ここで、簡易再生とは、元の構造化された電子論文の部分的な再生を行い、詳細度の低い内容だけを表示することを意味する。

### 2.2 内容的階層構造

本稿では、詳細度が低い情報の配下に詳細度が高い情報を並べて構造化したものを内容的階層構造として定義する(図1参照)。本稿で扱う内容的階層構造は、MPEG-2における各種スケラビリティプロファイルで規定されている階層符号化とは異なるが、実際に扱う情報はストリーム状に構造化されている。

従来の電子論文では、文書画像毎に処理が行われたり、テキスト情報全体に対して処理が行われる。そのため、論文の内容に基づいた処理を行うためには多くの処理を必要とするが、内容的階層構造を持つ電子論文は、論文の内容度を考慮して構成されたデータ構造に応じた再生機能および編集機能の実現が可能となる。すなわち、論文の内容の深さに応じた再生や編集、章や節など

の形式的単位毎の再生や編集などが可能になる。

## 3. 実現機能と情報構成

本節では、内容的階層構造に基づいて前述の簡易再生機能および簡易電子論文編集機能を実現するために必要な情報について述べ、それらがどのように表現されているかについて説明する。なお、本稿では、入力情報から領域分割処理によって生成された、電子論文のコンテンツ情報となるテキスト情報および静止画像情報をソース情報と呼ぶ。

### 3.1 簡易再生機能

最初に、簡易再生機能の概要と、機能実現のための情報構成について述べる。

#### 3.1.1 機能概要

簡易再生機能とは、内容に沿って構造化された電子論文データを再生する際に、利用者からの要求を受けた部分だけを選択して再生する機能である。簡易再生例を図1に示す。

簡易再生の種別は、内容度別および章節別に分けられる。内容度に基づいた簡易再生とは、利用者が要求する電子論文再生時の詳細度を受け、予め内容の詳細度に基づいて構造化されているソース情報を、その内容度を示す情報をもとに選択しながら表示するものである。例えば図1において、利用者から最も簡潔な内容の再生要求が与えられた場合、Layer1のソース情報だけが1-2-3-4の順に再生される。

論文の章節構造に基づく簡易再生とは、利用者の要求に応じて論文の章や節毎に再生処理を行うことである。例えば、利用者から第2節だけの再生要求があった場合、2-2.1-2.1.1-2.2の順にソース情報が再生される。

これらの2つの指標を組み合わせることで、利用者の要求する部分のみを復号・再生することが可能となる。

この他にも、電子論文を構成する各ソース情報のメディア種別に基づいて、利用者が指定したメディア情報のみを再生する場合が考えられるが、本稿では、このメディア別簡易再生は検討対象としない。

#### 3.1.2 構成情報

前述の簡易再生機能を実現するためには、まず各ソース情報の内容度識別情報が必要である。これにより、利用者は任意の内容度を電子論文に要求することが可能となる。また、電子論文の中から再生する章や節を指定するために、各ソース情報がどの章あるいは節に属していたかを示す情報、すなわち章節識別情報が必要である。これら2種類の情報は、各ソース情報領域の直前に付加されて構造化されるものとする。

電子論文再生時には、各ソース情報の再生処理順序を表す情報が必要である。これをシナリオ情報と呼ぶ。シナリオ情報は論文の内容の展開を示す情報であり、全てのソース情報の識別番号を再生順序に従って記述したものである。例えば、10個のソース情報からなる電子論

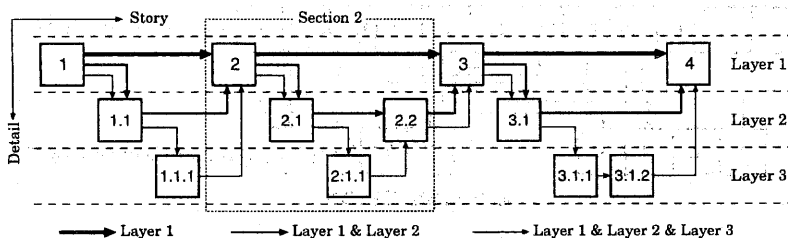


図 1: 内容的階層構造を有する電子論文の簡易再生例

文があり、ストリームの先頭に近い順に1から10までの識別番号を与える。この電子論文を最初から復号・再生する場合、シナリオ情報には1から10までの識別番号が順に記述される。シナリオ情報は、構造化データの先頭に位置し、各ソース情報よりも先に処理される。シナリオ情報により、以降の電子論文再生処理では再生対象情報を利用者が逐一指示する必要がなくなり、機械的に実現可能となる。このシナリオ情報に記述されている内容と、各ソース情報を対応付けるために、各ソース情報には一意に定まる識別番号を付加する。この識別番号は、内容度識別子、章節識別子と併せて構造化される。

利用者が簡易再生機能を利用する場合に、内容度と章節番号を用いて必要箇所の指定を行なうための判断基準が必要となる。そこで、電子論文の目次情報を用意する。目次情報は、各ソース情報のタイトルをシナリオ情報順に並べて記述したものである。目次情報は、電子論文を生成する際に同時に作成しておき、各ソース情報よりも先に構造化データ中に記述され、復号・表示される。目次情報も一つのソース情報として扱う。

1. で述べたように、電子論文はテキスト情報および静止画像情報という異なるメディア情報で構成される。よって、再生時には、メディアの種類に基づいて異なる処理を行う必要がある。そのためには各ソース情報がテキスト情報なのか、静止画像なのかを示すメディア識別子が必要となる。このメディア識別子も、ソース情報識別番号、内容度識別子、章節識別子と同様に、各ソース情報の前に付加されて構造化される。

### 3.2 簡易電子論文編集機能

次に、簡易電子論文編集機能の概要と、機能実現に必要な情報について説明する。

#### 3.2.1 機能概要

簡易電子論文編集機能とは、元の電子論文の一部を内容度や章節毎に取り出して新たな電子論文を作成する機能である。簡易電子論文編集機能例を図2に示す。

3.1.1で説明したように、簡易電子論文編集機能も内容度別、章節別に分類される。

内容度別の簡易電子論文編集とは、利用者から要求された電子論文の詳細度に基づき、元の構造化された電子論文中の各ソース情報を、その内容度を示す情報をもとにして選択して編集し、新たに簡易電子論文として生成

するものである。その際、元の電子論文の各ソース情報の内容度を示す情報が用いられる。図2の例では、最も詳細度の低い4つのソース情報(1, 2, 3, 4)のみを選択して編集することで、最も簡潔な内容をもつ電子論文が構成される。

章節毎の簡易電子論文再生とは、元の電子論文の中から利用者の指定した章や節などの部分のみを取り出して編集し、電子論文を作成するものである。その際、元の電子論文において各ソース情報がどの章あるいは節に属していたかを示す情報が用いられる。図2の例では、第2節のソース情報(2, 2.1, 2.1.1, 2.2)のみを選択して編集することで、元の論文の部分的な内容をもつ電子論文が構成される。

以上の2種類の指標を組み合わせて用いることで、利用者の要求により忠実に、新しい簡易電子論文を作成することが可能となる。

#### 3.2.2 構成情報

以上で述べた機能を実現するためには、各ソース情報の内容度識別情報が必要である。また、各ソース情報が出現した章または節の識別番号も必要である。さらに、新たに作成した電子論文を再生する時のために、3.1.2で述べたソース情報の識別番号およびメディア識別番号も必要である。以上で述べた4種類の情報は、いずれも文章領域および図領域の前に付加されて構造化される。

### 4. 機能実現手法

本節では、前節で述べた内容的階層構造を持つ電子論文に必要な情報を用いた簡易再生機能および簡易電子論文編集機能の実現手法について述べる。両機能を実現するには、前節で述べた各情報を以下で説明する手順に従い、電子論文作成時に予め構造化しておく必要がある。

両機能を利用する際は、「章または節Aの、内容度Bまで」という形式で指定された部分を処理の対象とする。Aは章節識別番号であり、図1および図2の場合では章節番号1, 2, 3, 4の組合せを指定する。Bは内容度識別番号であり、同様に内容度1, 2, 3の組合せで指定する。

#### 4.1 電子論文作成処理手順

最初に、内容的階層構造を有する電子論文の作成手法について述べる。なお今回は、入力として $\text{\TeX}$ ソース

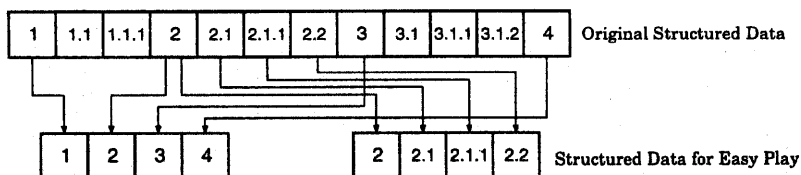


図 2: 簡易電子論文編集例

情報が与えられた場合の電子論文作成処理について述べる。作成処理は、 $\text{T}_{\text{E}}\text{X}$  ソース情報解析、ソース情報生成、中間フォーマット生成、中間フォーマット解析および符号化(構造化データ作成)に大きく分けられる。中間フォーマットは SGML を基本として記述されている<sup>(1)</sup>。具体的な作成手順を以下に示す。また、図 3 に作成処理器の構成を示す。図 3 中の各数字は、各処理手順と対応している。

1. 入力  $\text{T}_{\text{E}}\text{X}$  ソース情報から、論文のタイトル、著者などの情報を抽出し、目次情報生成器に渡す。
2.  $\text{T}_{\text{E}}\text{X}$  ソース情報を解析し、本文領域を領域分割処理器に渡す。領域分割処理器では、 $\text{T}_{\text{E}}\text{X}$  ソース情報中に記述されている `\section` などの章や節を示すコマンド(以下、章節コマンドと呼ぶ)を検出する度に、そこまでのテキスト情報を分割し、1つのソース情報として生成する。表領域は、本文領域と同様に扱う。図領域からは、図題を抽出し、内容度識別子、章節番号、メディア識別子を生成する。図本体はそのままソース情報とする。
3. 領域分割処理により分割されたテキスト領域情報を  $\text{T}_{\text{E}}\text{X}$  情報に整形してソース情報を生成する。
4. 章節コマンドと共に記述されている各章または節のタイトルを、目次情報生成器に渡す。
5. 目次情報生成器は、論文全体のタイトルや著者、および各ソース情報のタイトルを順に並べて目次情報を生成する。目次情報も  $\text{T}_{\text{E}}\text{X}$  形式で記述される。
6. 入力情報の解析が終了したら、生成したソース情報の数を中間フォーマットに記述する。
7. 各ソース情報の内容度識別子、章節番号、メディア識別子、ファイル名を、入力情報中の出現順序に従って中間フォーマットに記述する。以上で中間フォーマットの生成・記述を終了する。
8. 中間フォーマット解析器がソース情報数を中間フォーマットから抽出し、ソース情報識別番号生成器およびシナリオ情報生成器に渡す。
9. シナリオ情報生成器がソース情報数に基づいてシナリオ情報を生成した後、符号化器で符号化する。ソース情報数が 10 の場合、1 から 10 までを順に記述したシナリオ情報が生成される。
10. 符号化器で目次情報を符号化する。

11. 各ソース情報を、ソース情報識別番号、内容度識別子、章節番号、メディア識別子と一緒に符号化する。ソース情報識別番号は、ソース情報符号化時にソース情報識別番号生成器によって生成され、符号化器に渡される。
12. 全てのソース情報を符号化したら作成処理を終了する。

#### 4.2 簡易再生処理手順

次に、簡易再生処理手順について説明する。簡易再生処理器の構成を図 4 に示す。図 4 中の各数字は、各処理手順と対応している。

1. ストリームデータを先頭から読み込んで行き、シナリオ情報を復号してシナリオ情報解析器に渡す。
2. 目次情報を復号し、フォーマットで表示形式に変換した後にブラウザで表示する。
3. 利用者から、再生を要求する内容度 ( $c_0$ ) が内容度識別子解析器に入力される。また、再生を要求する章や節の番号 ( $n_0$ ) が章節番号解析器に入力される。
4. シナリオ情報に従い、ソース情報識別番号を復号しながら各ソース情報を走査して行き、各ソース情報の内容度識別子 ( $c$ ) を復号して内容度識別子解析器に渡す。さらに、各ソース情報の章節番号 ( $n$ ) を復号して章節番号解析器に渡す。
5. ソース情報の内容度 ( $c$ ) が、利用者の指定した内容度 ( $c_0$ ) よりも低い ( $c < c_0$ ) の場合、かつソース情報の章節識別番号 ( $n$ ) が、利用者の指定した再生要求章節番号 ( $n_0$ ) に等しい ( $n = n_0$ ) 場合、直後のソース情報を復号して再生する。両方の条件を満たさない場合はソース情報を復号しない。
6. ソース情報を復号・再生する場合は、メディア識別情報とソース情報をフォーマットに渡し、表示用データに整形された後に表示される。
7. 電子論文データの最後まで到達する、あるいは利用者からの再生停止命令が与えられるまで、4 から 6 までの処理を繰り返す。

#### 4.3 簡易電子論文編集処理手順

最後に、簡易電子論文編集処理手順を以下に示す。また、簡易電子論文編集処理器の構成を図 5 に示す。図 5 中の各数字は、各処理手順に対応している。

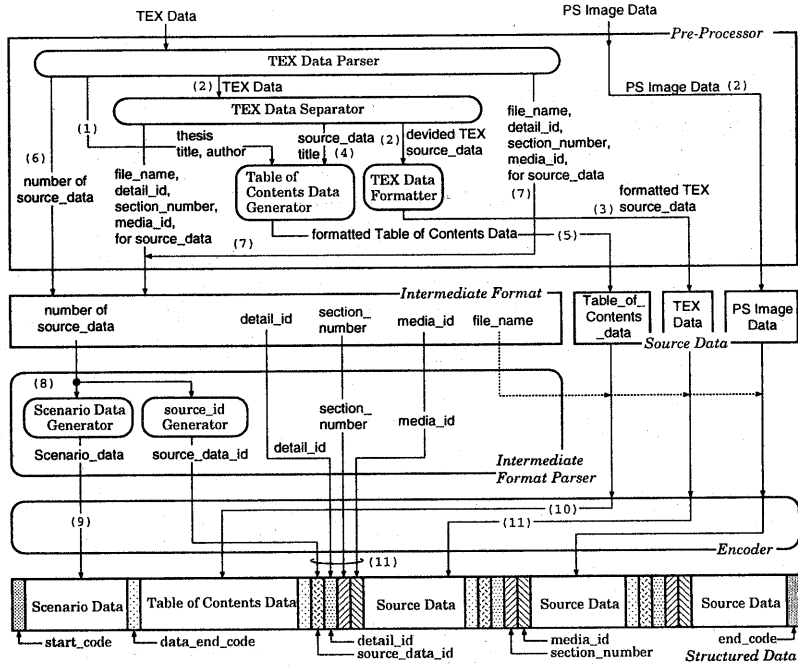


図 3: 作成処理器

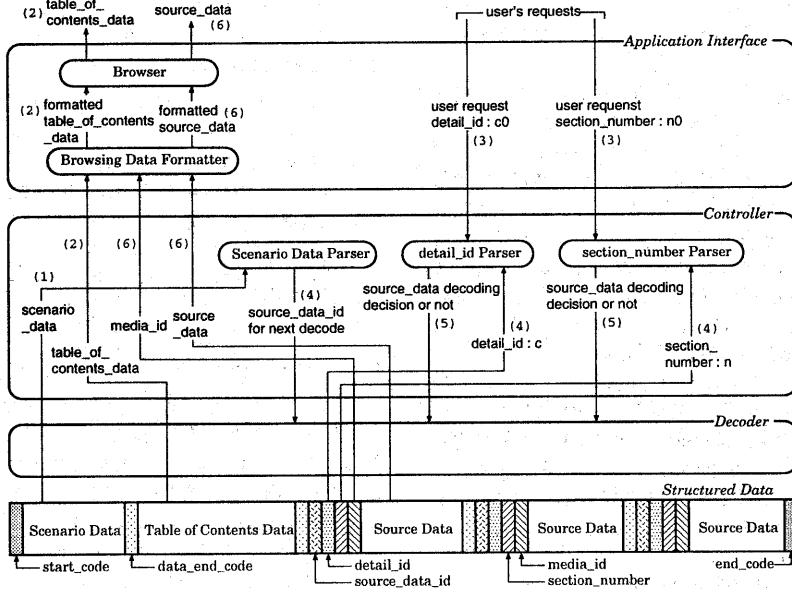


図 4: 簡易再生処理器

1. ストリームデータを先頭から読み込んで行き、シナリオ情報を復号してシナリオ情報解析器に渡す。
2. 利用者から、編集の対象とする内容度(c0)が内容度識別子解析器に入力される。また、編集の対象とする章あるいは節の番号(n0)が章節番号解析器に入力される。

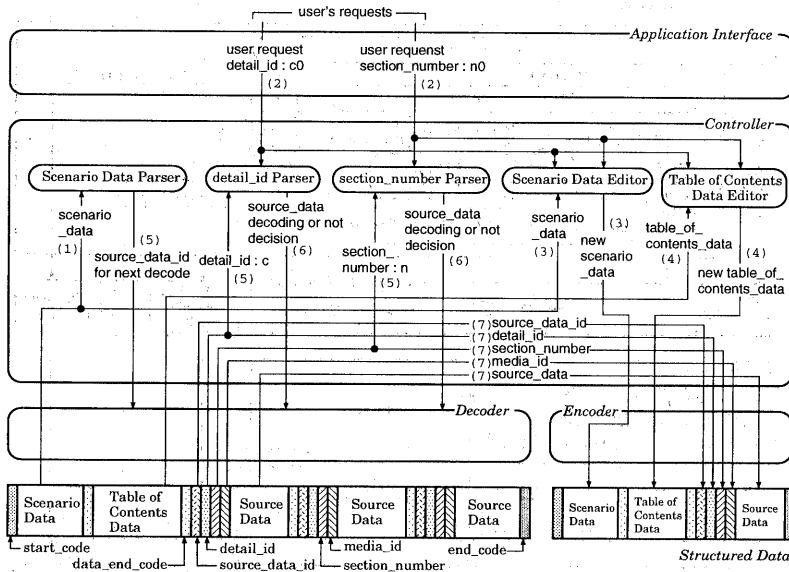


図 5: 簡易電子論文編集処理器

3. 復号されたシナリオ情報をシナリオ情報変換器に入力し、利用者から与えられた処理条件を基にシナリオ情報を再編集・変換し、新しい簡易電子論文の先頭部分に記述する。新しいシナリオ情報は、元のシナリオ情報から利用者の要求に従い必要部分のみを抽出して作成する。
4. 目次情報を復号して目次情報変換器に入力し、利用者からの処理条件を基に目次情報を再編集・変換し、シナリオ情報に続けて新しい簡易電子論文に記述する。新しい目次情報は、元の目次情報から利用者の要求に従い必要部分のみを抽出して作成する。
5. 元のシナリオ情報に従い、ソース情報識別番号を復号しながら各ソース情報を走査して行き、各ソース情報の内容度識別子(c)を復号して内容度識別子解析器に渡す。さらに、各ソース情報の章節番号(n)を復号して章節番号解析器に渡す。
6. 復号したソース情報の内容度(c)が、利用者の指定した内容度(c0)よりも低い(c<c0)の場合、かつソース情報の章節識別番号(n)が、利用者の指定した再生要求章節番号(n0)に等しい(n=n0)場合、直後のソース情報を新しい簡易電子論文中に記述する。両方の条件を満たさない場合、ソース情報には何も処理を行わない。
7. 新たに作成する簡易電子論文にソース情報を記述する場合、直前に付加されているソース情報識別番号、内容度識別番号、章節番号、メディア識別番号も一緒に記述する。各情報の並び方は、元の論文の場合と同様である。
8. 電子論文データの最後まで到達する、あるいは利用者からの簡易電子論文編集停止命令が与えられるま

で、5から7までの処理を繰り返す。

## 5. 実装

3. で説明した情報を用い、4. で説明した実現手法に従って、内容的階層構造を有する電子論文を作成した。実装条件を以下に示す。

- 静止画像情報は予め PostScript 形式で用意されているものとする。
- 文章領域は復号した後に T<sub>E</sub>X ソース情報に整形し、コンパイルしてから X<sub>D</sub>VI を用いて表示する。
- 図領域は X<sub>V</sub> を用いて表示する。

今回の実装により、内容的階層構造を持つことにより、要求に応じて電子論文の一部を容易に取り出して利用することができることを確認した。

## 6. まとめ・今後の課題

本稿では、内容的階層構造を持つ電子論文について検討を行った。特に、内容的階層構造に基づく簡易再生機能および簡易電子論文編集機能の処理について検討した。また、内容的階層構造および上記機能を有する電子論文の実装を行った。

今後の課題としては、処理速度の観点からの構造化データの最適化などが考えられる。

## 参考文献

- (1) 高屋 和幸, 笠井 裕之, 児玉 明, 富永 英義: “MPP 論文の情報構成および作成手法に関する検討”, 情処学研報, AVM98-20, pp. 25-30 (1998).