

CAI教材データベースにおける HTML文書の統合管理

山北隆典† 富士 隆† 三枝武男†

†学習情報通信システム研究所 ‡北海道情報大学

インターネット上に大量に存在するHTML文書の中には学習教材として有用な文書が少なくない。CAIによる学習にこれらの資源を取り入れるためには、CAI教材データベースでHTML文書を統合的に管理することが望ましい。本稿では、HTML文書統合の要件として、埋め込みリンクの制御（提示の際の隠蔽やリンク先の変更）、学習で利用するすべての文書の属性管理、柔軟な再利用を可能にするためのコンテンツのカスタマイズ、および異種文書（ハイパーフレーム）との提示方式の差異の吸収を取り上げる。さらに、それらの要件を解決するアプローチとして、Dexterモデルに基づいたハイパーテキスト・データベースによるデータ管理方式について述べる。

A CAI Material Database for Integration of HTML Documents

Takanori Yamakita† Takashi Fujii† Takeo Saegusa‡

†Software Research Laboratory ‡Hokkaido Information University

There are many HTML documents, which are available as learning materials, on the Internet. In order to utilize these documents for learning in CAI systems, it is necessary to store them in the database together with all original instruction components of the CAI system uniformly. In this paper, we point out four important issues which we need to tackle the integration of HTML documents. These issues are the control of their embedded links(hiding or changing), the property management of them, the tailoring their contents, and the encapsulation of each presentation styles. Moreover, we describe our basic framework based on the Dexter model to solve above issues.

1 はじめに

CAI(Computer Assisted Instruction)システムでは教材を部品化し、適切に属性付けしたうえでデータベース上に管理しておくことで、学習者の状況に応じたきめ細かな教材の提供が実現できると共に、教材部品の再利用による教材作成作業の軽減が期待できる。我々は、高度個別型CAIフレームワークの開発過程で、教材を部品化するモデルとしてハイパーフレーム [3] を提案してきた。さらにハイパーフレームの管理を目的としたCAI教材データベースを教材の相互利用の観点から検討してきた [1]。

一方、WWW(the World-Wide Web)の爆発的な普及に伴って、大量のHTML(Hyper Text Markup Language)文書がインターネット上に作成されている。その中には(学習や教育を意図するか否かにかかわらず)学習教材として有効に利用できる内容を持った文書が少なくなく、このような文書の活用が望まれている。

我々は学習教材として有効と思われるインターネット上の既存のHTML文書を高度個別型CAIの教材部品として積極的に取り入れ利用するためのアプローチとして、HTML文書をハイパーフレームと共に統合的に管理できるCAI教材データベースの構築に取り組んでいる。本稿では統合化の要件を考察し、HTML文書統合という観点からCAI教材データベースでのデータ管理方式について述べる。

2 関連研究の問題点と統合の要件

2.1 関連研究

既存のHTML文書を学習教材として積極的に利用した例としてK-12を対象とした研究 [4] がある。このシステムでは教材作成者が既存のHTML文書に paths¹ を定義することで、カリキュラムに基づいた学習を支援している。既存のHTML文書をカリキュラムでのコンテキストで利用できるように、学習者に提示する際にコンテンツの一部を隠蔽したり、導入文を追加したりでき、部品としての利用を可能にしている点も特徴である。ただし、HTML文書以外は扱っていない。

また、WWWのデリバリ機能を利用したCAIシステムが多数提案されている。その多くは、WWW

¹連続した複数のリンクを順序付けしたナビゲーションの道筋。

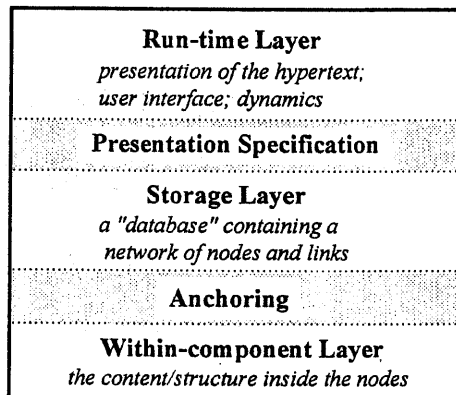


図 1: Dexter モデル

サーバのバックエンドに ICAI(Intelligent CAI) もしくは ITS(Intelligent Tutoring System) と呼ばれるシステムを位置付け、その出力結果を HTML 化したのち WWW クライアントへ転送するといったアーキテクチャを採用している [1][7]。こうした WWW ベースの CAI システムに HTML 文書を利用している例として、CALAT [8] がある。このシステムでも教材作成者によって、既存の HTML 文書とバックエンドの ICAI システムの文書との間にリンクを設定でき、それらの間のナビゲーション環境を提供している。HTML 文書としては独立性の高い辞書的な内容のものを想定している。

既存の HTML 文書と他のアプリケーション文書との統合の研究として、open Hypermedia system (OHS) の研究がある。Microcosm DLS [2] は、Microsoft Word、Microsoft Access、Lotus 1-2-3 等のアプリケーションの文書と HTML 文書をリンクによって結び付けるシステムである。しかし、文書の提示はそれぞれに対応するアプリケーションが責任を持つこと、文書全体として統一的な属性管理を行わないなど、異種文書を統合的に扱おうとするシステムではなく、あくまでリンクサービスを提供するシステムである。一方、DHM [5] はハイパーテキストの参照モデルである Dexter Hypertext Reference Model [6] (以降、Dexter モデルと略す) に基づき、異種メディアの文書を HTML 文書も含めて統合的に扱えるハイパーメディアシステムのオブジェクト指向フレームワークである。図 1 に Dexter モデルの概略を示す。

関連研究	HTML文書へのリンク定義機能	HTML文書の内容変更機能	Embedded links管理	HTML文書管理	異種文書との統合利用
Walden's Paths ^[4]	Page to Page	追加・変更・部分削除	なし	一部文書のURL (Embedded linksによるリンク先文書は除く)	なし
CALAT ^[8]	HTMLの範囲内	なし	なし	一部文書のURL (Embedded linksによるリンク先文書は除く)	ナビゲーション (HTML文書は辞書的利用)
Microcosm DLS ^[2]	m対nの双方向 (span to span)	なし	なし	一部文書のURL (Embedded linksによるリンク先文書は除く)	ナビゲーション (文書提示に責任は持たない)
DHM ^[5]	m対nの双方向 (span to span)	なし	なし	一部文書のURLと各種属性 (Embedded linksによるリンク先文書は除く)	ナビゲーション、属性値による検索 (管理外の文書は検索対象外)
Ours	m対nの双方向 (span to span)	追加・変更・部分削除、複数文書のマージ	Link DBで管理 (隠蔽・リンク先変更可能)	利用可能な全文書のURLと各種属性	ナビゲーション、属性値による検索 (利用可能な全文書)

表 1: 類似研究との比較

2.2 学習への適用における問題点

ところで、いずれのシステムにおいても既存のHTML文書にもともと埋め込まれていたハイパーリンク (以降、embedded linksと呼ぶ) に関しては管理の対象外となっている (表1参照)。学習システムにとって、学習者の興味に応じた自由な探索の余地を残しておくことも1つのポリシーであるが、他方、学習者にとって不必要な情報、あるいは学習に悪影響を与えるような情報であると教材作成者が判断した文書へのハイパーリンクは、教材作成者の裁量で隠蔽できる機能も必要である。

また一方で、embedded linksの先の文書が学習にとって有効な内容の文書であるならば、ハイパーリンクによるナビゲーションだけでなく、ハイパーフレームに付与しているような学習支援に必要な属性によって教材データベースを検索した結果としても学習者に提供できなければ、教材部品として統合されたとはいえない。そのためにはembedded linksで辿ることのできる文書をすべてデータベースが把握し、属性管理をしなければならない²。

2.3 HTML文書統合の要件

以上述べてきた関連研究での経験や問題点の考察から、学習教材として有効なHTML文書を教材部品として高度個別型CAIの教材データベースに統合するためには以下の要件を実現できるような情報

²不必要な文書へのハイパーリンクは極力隠蔽しないと際限のない状態になってしまうことに注意が必要である。

管理が必要であることがわかる。

1. HTML文書上のembedded linksを文書の提示の際に適切に制御 (すなわち、隠蔽やリンク先の変更) できるような情報をデータベースで管理できること。
2. 学習教材の部品として利用するすべてのHTML文書について、ハイパーフレームと同様に学習支援に必要な属性付けができ、それを利用した検索が可能であること。
3. HTML文書を可能な限り任意のコンテキストで利用できるように、もともとのコンテンツ内のコンテキスト依存部分の変更や隠蔽、別のコンテキストでの利用をスムーズにするような記述の追加、および教材部品間の粒度をそろえるための分割やマージといった機能が文書の提示の際に実現できること。
4. DHMが異種メディア文書を扱ったのと同様に、HTML文書とハイパーフレームとの提示方式の差異を吸収できること。

3 基本構想

3.1 試作システムの範囲

現在、今回の検討事項を確認するための試作システムを構築している。本システムは次のような機能を実現する。

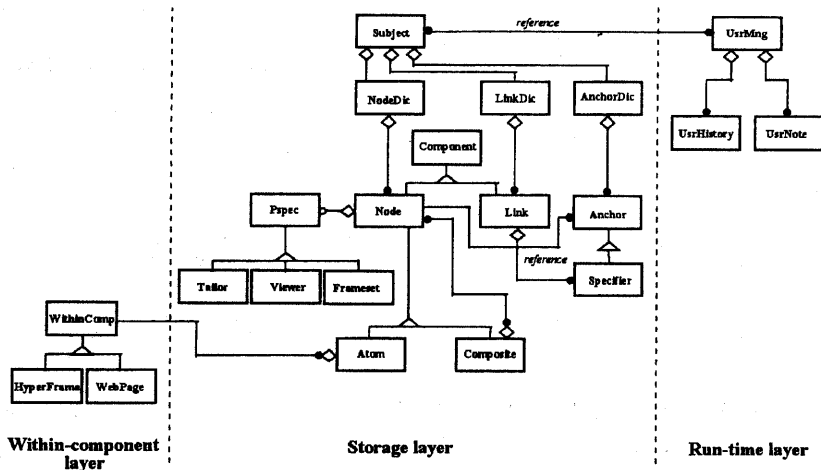


図 2: CAI教材データベースのスキーマ

1. 教材部品 (ハイパーフレーム、HTML 文書) に付与した属性によるキーワード検索。
2. 教材部品間に設定したリンクに基づくナビゲーション検索。

なお、教授方略、個別化アルゴリズムに基づく、個別化された教材部品提供機能の実装は範囲外とする³。

3.2 基本モデル

教材部品に対して属性を用いたキーワード検索とリンク情報を用いたナビゲーション検索の双方を実現するために、いわゆるハイパーテキスト・データベースとしてシステム化する。そのために、ハイパーテキストの参照モデルである Dexter モデルを基本モデルとして採用する⁴。

上で述べた CAI 教材データベースへの要件をこの Dexter モデルに基づき解決する。

3.2.1 Embedded links の制御情報の管理

HTML 文書上の embedded links 情報は Dexter モデルに基づくリンクオブジェクトとして管理する。

³教授方略、および個別対応のアルゴリズムの検討はすでに文献[3]でなされていること、今回の試作範囲内でも個別化アルゴリズムが利用する属性を網羅していることによる。

⁴我々は分散したハイパーテキスト間での相互利用を表現するためには Dexter モデルの拡張が必要であると提案している[11]が、本稿の内容は拡張を必要としない範囲なので、簡単のため拡張モデルは示さない。

HTML 文書が提示されるときには、もともとの embedded links の記述をそれに対応するリンクオブジェクトを指すように変更して提示する。そのための変更情報を Dexter モデルの presentation specification (以降、pspec と略す) で管理する。Dexter モデルではノードやリンクに対する提示の仕様を pspec に記述する。

3.2.2 HTML 文書の属性管理

Dexter モデルでは、ノード/リンクのネットワーク構造とノードの実際のコンテンツとを2つの層に分離することで異なる形式の文書もノードという概念で統一的に扱える、という特長を有するので、ハイパーフレームも HTML 文書もノードとして扱える。学習支援に必要な項目をノードの属性として定義しておけば、双方の文書すべてを対象とした統一的なキーワード検索が実現できる。付与すべき属性はすでに文献[3]で検討されている。なお、HTML 文書の URL (Uniform Resource Locator) は、ハイパーフレームの実体と同様に Dexter モデルにおける within-component layer で管理する。すなわち、ノードでは within-component layer のオブジェクトへのポインタを管理し、リンク情報においてもその終端ノードの識別子として同様のポインタを管理する。この方式によってハイパーフレームと HTML 文書が統一的に扱える。

```

struct RejectInfo {
    String start_tag; //開始位置(文字列)
    int start_order; //出現順位
    String start_offset; //オフセット
    String end_tag; //終了位置(文字列)
    int end_order; //出現順位
    String end_offset; //オフセット
};

struct ReplaceInfo {
    String old_str; //変更対象文字列
    int order; //出現順位
    String new_str; //置き換え文字列
};

struct AppendInfo {
    String tag; //開始位置(文字列)
    int order; //出現順位
    String str; //追加文字列
};

class Tailor : public Pspec {
private:
    List<RejectInfo> reject_info; //部分削除データ
    List<ReplaceInfo> replace_info; //文字列置換データ
    List<AppendInfo> append_info; //文字列追加データ

    String reject(String html); //部分削除処理
    String replace(String html); //置換処理
    String append(String html); //追加処理

public:
    String perform(String html); //カスタマイズ処理
};

```

図 3: Tailor クラスの概要

3.2.3 HTML 文書のカスタマイズ

HTML 文書を提示する際のコンテンツに対する追加・変更・削除、および分割・マージのための情報は、Dexter モデルの pspec として管理し、実際に提示されるときに解釈される。

3.2.4 提示方式の差異の吸収

ノードのコンテンツの提示方式はそのノードの pspec で指定できる。したがって、コンテンツの形式に合わせた pspec を記述しておけば、ノードに対して提示操作を行うだけで、それに対応した方式で提示できる。

4 HTML 文書の統合メカニズム

4.1 データベース・スキーマ

Dexter モデルと対応付けたデータベース・スキーマの概略を OMT 法 [9] により図 2 に示す。

また、HTML 文書の pspec は、Tailor クラスとして実装する (図 3 参照)。

なお、run-time layer のオブジェクトはセッション管理を目的としているがここでは議論しない。

```

entry
Node ID取得
Node object取得 (NodeDic)
if atom
    if hyperframe
        hyperframeデータ取得 (HyperFrame)
        提示用applet情報追加 (Viewer)
    else // HTML
        HTML文書取得 (WebPage)
        部分削除 (Tailor)
        文字列置換 (Tailor)
        文字列追加 (Tailor)
    end
else // composite
    while 子供の数だけ
        atom処理(8)
        temporal HTMLファイル作成
    end
    frameset表示用HTML文書作成 (Frameset)
end
HTML文書にheader/footer追加
HTML文書転送
exit

```

注: ()内は関連するオブジェクト

atom処理*

図 4: HTML 文書のカスタマイズ・アルゴリズム

4.2 HTML 文書のカスタマイズ・アルゴリズム

Tailor オブジェクトの情報によって HTML 文書のカスタマイズする際のアルゴリズムの概略を図 4 に示す。

4.3 システム構成

Dexter モデルに基づいた試作システムの全体構成を模式的に図 5 に示す。ハイパーフレームも HTML 文書も within-component layer に位置付けられ、storage layer では共に within-component layer のオブジェクトへのポインタ情報を管理する。したがって、文書提示の際には、ノードのポインタ情報から within-component layer のオブジェクトを特定し、ハイパーフレームの場合は Hyper-Frame DB を検索し、HTML 文書の場合には URL 情報から対応する WWW サーバにアクセスすることで、それぞれの形式のコンテンツを取得する。結果は CGI によって WWW ブラウザに送られる。このときハイパーフレームのコンテンツはその提示に適した Java applet と共にブラウザに送られる。Applet は Web サーバ上に置かれており、データベースはハイパーフレームの pspec として applet へのポインタを管理している⁵。

⁵これは、文献 [10] によるハイパーフレームの実装を模倣したものである。

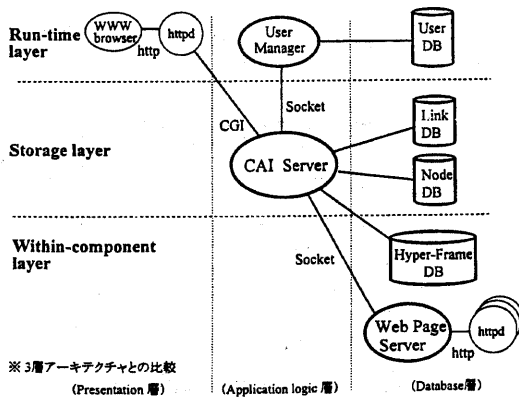


図 5: 試作システムの全体構成

5 おわりに

高度個別型CAIフレームワークに基づいたCAI教材データベース上でHTML文書を統合管理するための要件として、埋め込みリンクの制御情報の管理、HTML文書の属性管理、コンテンツのカスタマイズ、そして提示方式の差異の吸収に着目し、Dexterモデルに基づいたハイパーテキスト・データベースによるデータ管理方式を検討した。

Dexterモデルによるアプローチに基づき、within-component layerの利点を活かして異種文書を統一的に扱い、presentation specificationによってHTML文書の提示の際の埋め込みリンクの制御情報やカスタマイズ情報を管理することによって、教材作成者の意図の範囲内で、既存のHTML文書をハイパーフレームと共に教材部品として統合管理できることを述べた。

また本文中では触れていないが、既に報告している拡張Dexterモデル[11]を適用することで、本稿で述べたハイパーテキスト型教材が分散して存在する環境において、それらを相互利用することも可能であると考えている。

今後の取り組みとしてオーサリングツールの開発が考えられる。しかし、いつ変更されるかわからないというWWW上でのHTML文書の不安定さに起因する問題や、教材部品として利用できるHTML文書の発見の困難、また、HTML文書の著作権等社会制度上の問題など、この分野における課題は尽きない。

謝辞

日頃、ご指導、ご助言戴いている図書館情報大学教授 増永良文氏に感謝の意を表します。

参考文献

- [1] P. Brusilovsky, E. Schwarz and G. Weber; ELM-ART: An Intelligent Tutoring System on World Wide Web, Proceedings of the 3rd International Conference, ITS'96, pp.261-269, June 1996.
- [2] L. Carr, D.D. Roure, W. Hall and G. Hill; The Distributed Link Service: A Tool for Publishers, Authors and Readers, Proceedings of the 4th International World Wide Web Conference, Dec. 1995.
- [3] 富士 隆 他; ハイパーフレームを用いた知的マルチメディアCAIの開発, 1994年情報学シンポジウム講演論文集, pp.133-141, 1994年1月.
- [4] R. Furuta, F.M. Shipman, C.C. Marshall, D. Brenner and H. Hsieh; Hypertext Paths and the World-Wide Web: Experiences with Walden's Paths, Hypertext'97 Proceedings, pp.167-176, Apr.1997.
- [5] K. Grønbaek and N.O. Bouvin and L. Sloth; Designing Dexter-based hypermedia services for the World Wide Web, Hypertext'97 Proceedings, pp.146-156, Apr. 1997.
- [6] F. Halasz and M. Schwarz; The Dexter Hypertext Reference Model, Comm. of the ACM, Vol.37, No.2, pp.30-39, Feb. 1994.
- [7] J. Kay and R.J. Kummerfeld; An Individualised Course for the C Programming Language, Proceedings of the 2nd International World Wide Web Conference, 1994.
- [8] 丸山美奈 他; 分散教育環境CALATにおけるハイパーリンクアーキテクチャの提案 ~知的ナビゲーションの実現手法について~, 人工知能学会 SIG-J-9501-16, pp.104-111, 1995年12月.
- [9] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy and W. Lorenzen; OBJECT-ORIENTED MODELING AND DESIGN, Prentice-Hall, 1991 (羽生田栄一監訳、オブジェクト指向方法論 OMT, トッパン, 1992年)
- [10] 谷川 健 他; CORBAに基づく個人学習型CAI分散フレームワークの開発, 情報処理学会 第54回全国大会講演論文集(3), pp.97-98, 1997年3月
- [11] 山北隆典 他; Dexterモデルに基づくハイパーリンク・ビューのWWW上での実現, 情報処理学会 アドバンスト・データベース・シンポジウム'96論文集, pp.49-56, 1996年12月.