

電子図書館におけるコンテンツ流通の提案

日高 哲雄、斉藤 典明、阿部 貴博、関 良明

NTT 東日本 研究開発センター

〒180-8585 東京都武蔵野市緑町 3-9-11

{t.hidaka, saito.n, takahiro.abe, yoshiaki.seki}@rdc.east.ntt.co.jp

概要

インターネットの普及に伴い、ネットワークを介したデジタルコンテンツの流通が盛んになってきた。その原動力となった Web システムは、現在膨大なコンテンツが公開されており、重要な情報収集手段となっている。一方で、体系化された知識と情報の蓄積、共有を担う電子図書館も情報収集において重要な役割を果たす。そこで、筆者らは、電子図書館システムのコンテンツと Web システムのコンテンツをシームレスに検索できるようにすることを目的とし、自動リンク機能によりコンテンツ間の連携を実現するコンテンツ流通システム Net-X を開発した。

キーワード

電子図書館、メタデータ、コンテンツ流通、Dublin Core Metadata Element Set、インターネット

A Contents Distribution System for Digital Library

Tetsuo Hidaka, Noriaki Saito, Takahiro Abe, Yoshiaki Seki

NTT-East Research and Development Center

3-9-11 Midori-cho Musashino-shi, Tokyo 180-8585 Japan

{t.hidaka, saito.n, takahiro.abe, yoshiaki.seki}@rdc.east.ntt.co.jp

Abstract

The distribution of digital contents on networks has become popular because of the widely recognized Internet and the WWW(World Wide Web). The WWW is an important way of acquiring various kinds of information. On the other hand, Digital Library is another important way. We propose a contents distribution system, Net-X, which provides seamless retrieval services over the WWW and Digital Library.

Keywords

Digital Library, Metadata, Contents Distribution, Dublin Core Metadata Element Set, Internet

1. はじめに

電子図書館プロジェクトは、様々な国で盛んに進められている [1]。一方、膨大な量のコンテンツが公開されている Web システムのコンテンツ（以下、Web コンテンツ）も電子図書館と同じく重要な情報収集手段として多くの人に利用されている。これら 2 つのシステムのシームレスな検索を実現するためには、従来から使われてきた図書館システムの書誌情報による検索だけでなく、コンテンツの特性を考慮した新たな検索方法が必要となる。そこで、本稿では、電子図書館システムと Web システムの統合に求められる機能を検討し、その結果をもとにシステムの提案を行う。

以下、2 章では、電子図書館におけるコンテンツ流通について検討し、3 章では、電子図書館を拡張したコンテンツ流通システム Net-X の実装について述べる。

2. 電子図書館におけるコンテンツ流通

2.1. 電子図書館の役割と機能

図書館には、体系化された知識と情報の蓄積、共有という重要な役割がある。そのために、図書館は、資料を収集・組織化することによりコレクションを作成し、それを利用者に提供するとともに利用者の情報アクセスを支援する役割が期待されている [1]。この役割は、図書館を踏襲する電子図書館においても同様に期待される。具体的には、大学図書館等で提供されている電子図書館のサービスは、以下の 3 つに分類可能である [2]。

- (1) 自館によるデジタルコレクションの蓄積と提供
- (2) 既にデジタル化された資料の提供
- (3) サブジェクトゲートウェイ機能の提供

(1) は、歴史的資料や貴重資料を図書館自身が個別にデジタル化し提供するサービスである。(2) は、デジタル化された資料を利用するための許諾を得て、あるいはライセンスを購入してデジタル資料を提供するサービスである。この 2 つの方法は、図書館自身が自らコンテンツを提供するサービスである。一方、(3) は、ネットワーク上の他の電子図書館と相互に連携することによってメタデータを作成し、利用者を他の電子図書館の資料にナビゲートするサービスである。

2.2. 電子図書館システム NetLibra の概要

筆者らは、2.1 節で述べた (1) と (2) のサービスを提供する電子図書館システムとして、NetLibra を開発してきた [3]。NetLibra は、広域に分散した電子図書館をネットワークで相互に接続し、シームレスな電子図書館の利用環境を提供するシステムである。

また、NetLibra は、CORBA(Common Object Request Broker Architecture) による分散オブジェクト技術を利用して、ネットワーク上に分散したデジタルコンテンツの所在場所（格納サーバの位置）を意識することなく必要な情報にアクセスする横断検索機能やサーバの自律運用等の特徴を持つ。

NetLibra では、分散した電子図書館にコンテンツを配置することが可能なため、大量のマルチメディアコンテンツを扱うことが可能である。しかし、インターネットが普及し、Web システムを使って膨大な量のコンテンツが公開されている現在、NetLibra に閉じたコンテンツの検索だけでは十分とはいえ、2.1 節で述べた (3) サブジェクトゲートウェイ機能への対応も必要である。そこで、次節では、この観点から現在の NetLibra に付加すべき機能について述べる。

2.3. NetLibra に付加すべき機能

現在の NetLibra に付加すべき機能として以下の項目がある。

(1) Web コンテンツへの対応

Web システムで公開されているコンテンツは、雑多でありかつ膨大な量であるので、有用なコンテンツを取捨選択する仕組みを持つことが必要である。

(2) 検索方法の高度化

NetLibra に閉じたコンテンツの検索の場合は、コンテンツが体系化されており、従来から利用されている書誌情報による検索(図1)でも問題は少ない。しかし、Web コンテンツへ対応した場合、体系化されていない情報が大量に含まれることとなり、検索結果が不要なもので埋め尽くされる、検索結果が膨大な量になるなどの問題が生じる。そこで、書誌情報による検索だけでなく、関連のあるコンテンツへナビゲートする機能を持つことが必要である。

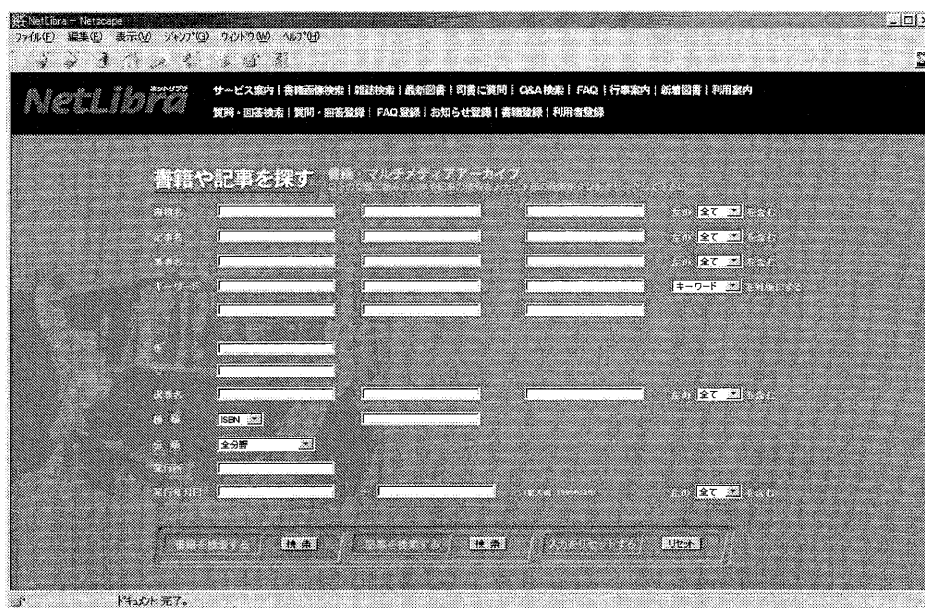


図1：NetLibra の検索条件入力画面

(3) 有料コンテンツへの対応

インターネット上では、有料コンテンツを利用者に提供するシステムも増加している。そのため、コンテンツの種別としても、無料コンテンツだけでなく、印刷用画像やソフトウェアなどの有料コンテンツも扱える仕組みを持つ必要がある。

2.4. NetLibra の機能拡張の実現方法

2.3 節で述べた NetLibra に追加すべき機能の具体的な実現方法を以下に示す。

(1) Web コンテンツへの対応

Web コンテンツと NetLibra のコンテンツをシームレスに検索できるようにするために、Web コンテンツのメタデータを収集し、検索用の DB に保存する必要がある。そこで、検索ロボットにより、有用なコンテンツを取捨選択し、Web コンテンツのメタデータを収集する。

収集したメタデータの格納方法として、

- Web コンテンツのメタデータを NetLibra と同じ DB に保存する方法
- Web コンテンツのメタデータを NetLibra と別の DB に保存する方法

の2通りがある。Web コンテンツは、電子図書館のコンテンツとは異なり、改編や削除の周期が短いので、検索ロボットを起動するたびにデータの修正が数多く発生すると考えられる。データのバックアップ等、システムの保守・運用という観点からは、更新の多いデータは、更新の少ないデータと別に扱うほうが良い。そこで、Web コンテンツのメタデータを NetLibra と別の DB に保存する後者の方法を採用した。

(2) 検索方法の高度化

関連のあるコンテンツへ利用者をナビゲートする機能を実現するために、筆者らが開発したノウハウ蓄積システム FISH[5] で提案した自動リンク機能を用いてコンテンツ間のリンクを生成する。

FISH では、内容が多様で、従来のデータベースのような属性と値による定型化が困難な情報を、仮想的なカードに細分化し、各カードにキーワードを付与して表現する。各カード間は、キーワードによりリンクで結ばれ、関連情報の参照を容易にする。

キーワードによるリンクは、HTML のようにコンテンツ作成時に埋めこまれるのではなく、カード参照時に自動的に生成されるため、

- 新たなカードとの動的なリンク機能
- カード作成者の気づかなかったカード間の関連に基づいたリンクを表示する Knowledge Awareness 機能 [6]

を実現できる。

この自動リンク機能を、電子図書館のコンテンツと Web コンテンツに拡張することにより、コンテンツ間の連携が可能となり、関連のあるコンテンツへナビゲートする機能を実現した。なお、自動リンクの基となるキーワードは、コンテンツおよびメタデータから形態要素解析によって抽出された名詞を用いた。

(3) 有料コンテンツへの対応

有料コンテンツ対応としては、市販されている電子決済モジュールを Web システムに組み込むことにより実現した。有料のコンテンツは、購入を希望する利用者にはできるだけ紹介したいという販売側の意思があるので、そのメタデータは、コンテンツ登録時に詳細に登録されることが多い。そこで、有料コンテンツのメタデータは、コンテンツ入力時に整備することとした。その際のメタデータの形式は、インターネット上での情報資源の発見を目的として提案され、標準化が進んでいる Dublin Core Metadata[4] に基づき保存する。同様に、ロボット収集した Web コンテンツのメタデータに関しても、可能な項目に対して、Dublin Core Metadata に準拠する形式で保存する。

3. コンテンツ流通システム Net-X の実装

2章の検討結果をもとに、電子図書館と Web コンテンツを統合して扱い、コンテンツ間の連携を実現するコンテンツ流通システム Net-X を開発した。概念図を図 2 に示す。

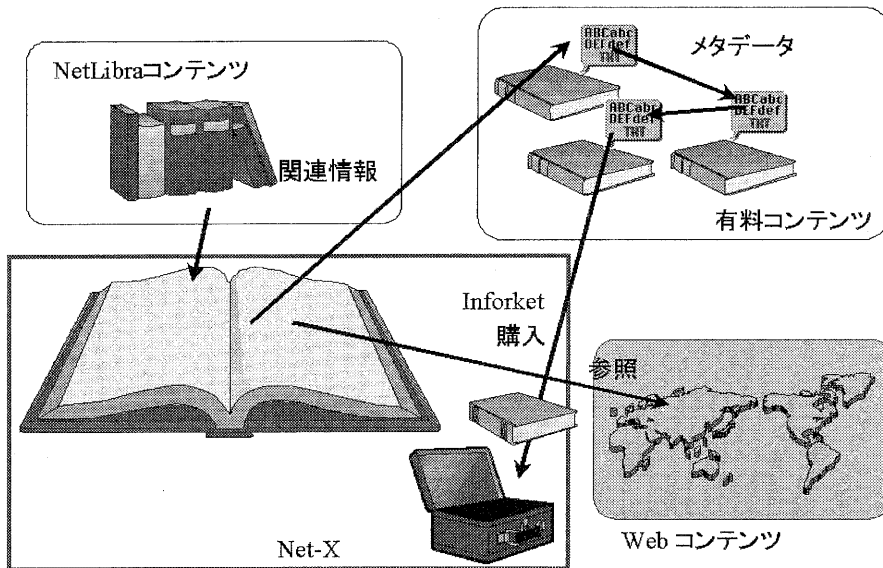


図 2：Net-X の概念図

3.1. Net-X のシステム構成

Net-X のハードウェア構成を図 3 に、ソフトウェア構成を図 4 に示す。「既存 Web サーバ」は既に存在する Internet の Web サーバ、「有料コンテンツサーバ」は全コンテンツに対して Dublin Core に準拠したメタデータが整備された有料コンテンツを扱える専用の Web サーバ、NetLibra サーバは、電子図書館のコンテンツを保存したサーバである。Net-X サーバと「既存 Web サーバ」、「有料コンテンツサーバ」は、インターネットで接続されている。一方、Net-X サーバと「NetLibra サーバ」は、CORBA を利用してネットワークで接続されている。

既存 Web サーバのメタデータは、検索ロボットにより定期的に収集され、メタデータ DB に保存される。有料コンテンツサーバのメタデータは、コンテンツ登録時に Net-X サーバに送られる。利用者が Web クライアントから検索を実行すると、入出力制御部から検索モジュールに検索指示が出され、Web コンテンツ検索部はメタデータ DB を検索し、NetLibra 検索部は、CORBA を利用して NetLibra サーバを検索する。自動リンク生成部は、コンテンツ自身の本文に対してと、メタデータ中の要旨に自動リンクを作成することができる。

3.2. シームレスな検索の実現

従来の検索システムでは、「検索条件入力」－「検索結果一覧」－「コンテンツ表示」という流れの繰り返しにより目的のものを発見する。したがって、「コンテンツ表示」で表示されるコンテンツ内のキーワードに興味を持った場合には、「検索条件入力」でそのキーワードを手動で入力する必要があった。

一方、Net-X では、「コンテンツ表示」で表示されるコンテンツ内にキーワードによる自動リンクが張られているので、そのリンクをクリックすることにより、「検索結果一覧」が表示される。すなわち、従来の

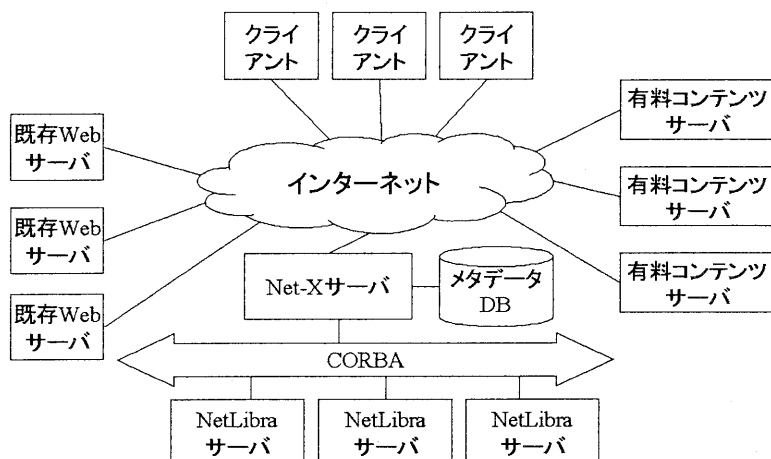


図 3：ハードウェア構成図

検索システムに必要な「検索条件入力」を省略できるのである。

また、従来のシステムでは、利用者が「コンテンツ表示」に興味を持ったキーワードで検索を行った場合に、検索結果 0 件という結果もあり得る。その場合、他のキーワードを考え、再入力する必要があった。しかし、Net-X では、自動リンクのキーワードとコンテンツは必ず対応しているので、検索結果 0 件ということはない。

このように、Net-X では、人間の思考を妨げないシームレスな検索を実現できる。

3.3. Net-X の動作例

Net-X のコンテンツ検索画面を図 5 に示す。画面の左側は、検索用のエリアであり、

- 「分野」「地域」「年代」の階層構造のカテゴリから検索する方法
- 「キーワード」を入力し検索する方法

が用意されている。これらの検索に該当したコンテンツのリストが、その下部に表示される。リスト表示の上部分は、該当した既存 Web コンテンツと有料コンテンツのリストである。リスト表示の下部分は、該当した NetLibra コンテンツのリストである。

(1) 有料コンテンツの表示

検索結果リストの中から有料コンテンツを選択した場合は、右側にそのコンテンツのメタデータが表示される (図 6)。

メタデータの表示では、そのコンテンツの分野、地域、年代に相当する属性のシンボルが画面上部に表示される。画面中央には Dublin Core 準拠のメタデータとコンテンツのサムネイルが表示される。メタデータの概要に関しては、自動リンクにより説明文がクリック可能になって表示される。この文字をクリックすることにより、そのキーワードで検索される情報の一覧が左側に表示される。このメタデータとサムネ

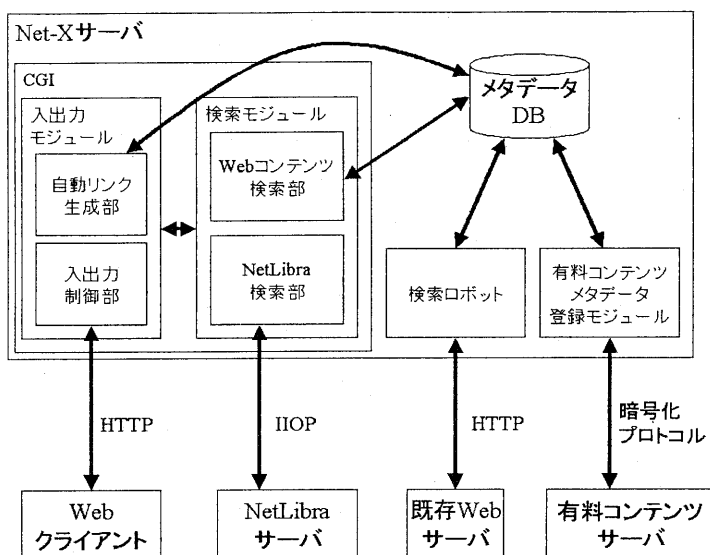


図4：ソフトウェア構成図

イルを参照しながらの購入したいコンテンツが出現した場合、画面中の「購入」ボタンをクリックすることにより、電子決済によりネットワークを介して購入できる。

(2) Web コンテンツの表示

検索結果リストの中から既存 Web コンテンツを選択した場合は、新しいウィンドウにそのページが表示される。Web コンテンツのテキスト部分に自動的にリンクを生成することも技術的には可能であるが、著作権の問題を考慮して、そのままのページを別ウィンドウで表示することとした。

(3) NetLibra コンテンツの表示

検索結果リストの中から NetLibra コンテンツを選択した場合、右側にそのコンテンツそのものが表示される。そのコンテンツの説明文は自動リンクによりクリック可能になって表示される。この文字をクリックすることにより、そのキーワードで検索される情報の一覧が左側に表示される。

4. おわりに

本稿では、電子図書館システム NetLibra と Web コンテンツのシームレスな検索の実現を目的に、コンテンツ間にキーワードによる自動リンクを生成することでコンテンツを連携させる機能をもつコンテンツ流通システム Net-X を提案した。Net-X は、コンテンツ全体を管理しているシステム側からコンテンツを関連付けるので、コンテンツ作成者が気づかなかったリンクも生成できる。

今後は、様々なコンテンツを用いたコンテンツ連携機能の評価を行う予定である。

参考文献

- [1] 杉本 重雄：“デジタル図書館実現のための要素技術と環境要素”，情報処理, pp.820-825, 1996.9.

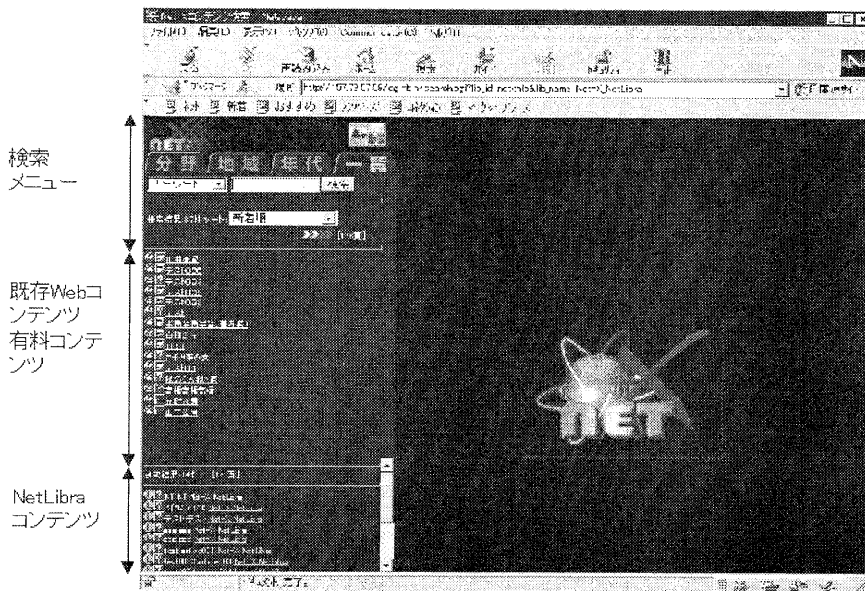


図5：コンテンツ検索画面

- [2] 杉本 重雄：図書館情報大学におけるデジタル図書館システム，第15回デジタル図書館ワークショップ，pp.17-28，1999.7.
- [3] 萱野 忠，西野 正和，関 良明，爰川 知宏：分散ネットワーク電子図書館 NetLibra の提案，第57回情報処理学会全国大会，3-1，1998.10.
- [4] 杉本 重雄：Dubline Core Metadata Element Set について-現在の状況と利用例，第14回デジタル図書館ワークショップ，pp.3-18，1999.3.
- [5] 関 良明：分散型ノウハウ蓄積システム GoldFISH における分散環境への適応，情報処理学会論文誌，Vol.36，No.6，pp.1359-13366，1995.
- [6] T. Yamakami, Y. Seki, "Knowledge awareness in asynchronous information sharing" : Proc. of the IFIP TC8/WG8.4 Working Conference, pp.215-225, 1993.

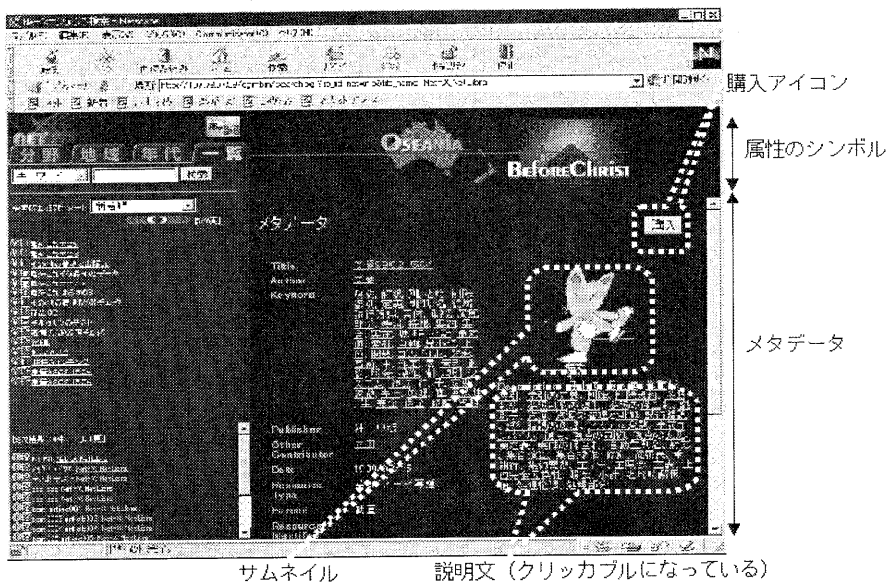


図 6：メタデータ表示画面