

情報組織化によるコンテンツ流通システム：Net-X

日高 哲雄 斉藤 典明 阿部 貴博 関 良明

NTT 東日本会社移行本部 研究開発センタ

あらまし：

インターネットの普及により、ネットワークを介してデジタルコンテンツを流通させることが注目されつつある。しかし、膨大なコンテンツの中からユーザが望む情報を探し出すことが難しくなるという問題も生じている。この問題を解決するために、Dublin Core Metadata を用いて、キーワードを利用したハイパーリンクを自動生成し、コンテンツを連携させる情報組織化技術を取り入れたシステムを開発した。本稿では、情報組織化技術の仕組みとその技術に基づき構築したコンテンツ流通システム Net-X について述べる。

Net-X: A Contents Distribution System based on Information Organizing Technology

Tetsuo Hidaka Noriaki Saito Takahiro Abe Yoshiaki Seki

NTT-East Provisional Headquarters Research and Development Center

Abstract :

The distribution of digital contents on networks has become more active by the spread of the Internet. However, it has become difficult for customers to search out the information that they expect from among huge contents. In this paper, with the solution to this problem in mind, we proposed the information organizing method of generating hyperlinks based on a keyword automatically in the metadata of contents, and relating contents. In addition, we describe a contents distribution system based on information organizing technology, Net-X.

1.はじめに

インターネットの普及により、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。しかし、膨大なコンテンツの中からユーザが望む情報を探し出すことが難しくなるという問題も生じている。

この問題を解決するために、筆者らは、ユーザが参照しているコンテンツから関連のある次のコンテンツを提供する情報組織化技術を導入する。情報組織化は、情報の関連付け、情報の分類、データマイニング等の要素技術により実現されるが、特に、本稿では、情報の関連付けに着目した。

情報の関連付けとしては、WWWのハイパーリンク機構がある。しかし、この機構はWebページに埋め込まれたリンクを辿るものであり、ユーザ（以下、読者）が疑問や興味を持った単語にリンクが張られているとは限らない。これは、Webページの作者と読者の知識レベルや興味の相違に主な要因がある。

このような問題に対して、本稿では、作者や読者の視点を超越し、コンテンツ全体を管理しているシステム側から情報を関連付けるシステムNet-Xを提案する。

一方、膨大なデジタルコンテンツを効率良く管理するために、データに関するデータであるメタデータの付与に関する議論も活発に行われている[1]。メタデータには、コンテンツ自身には含まれない著者名や発行年月日等の重要な情報も含まれている。そこで、では、関連付けの際に、コンテンツ自身の情報（一次情報）だけでなく、メタデータも有効活用する。

以下、2章では、情報組織化技術とメタデータについて説明し、3章では、実際に構築したシステムについて述べる。

2.情報組織化技術とメタデータ

2.1.情報組織化技術

情報組織化技術とは、ネットワーク上に分散して存在する情報と情報を自動的に結び付け、情報

の利用を活発にする技術である。情報の関連付けはその一要素技術であり、情報の関連付けを実現するリンク機構は、

- ① 作者が指定したリンク（WWWのHTML）
- ② 読者が指定したリンク（検索エンジンによるキーワード検索）
- ③ システムが全コンテンツの構造を判断し、自動生成されるリンク

に分類できる。①②は、現在のWWWですでに実現されている。③の基本構造に関しては、ノウハウ蓄積システムFISHで提案されている[2]。この自動リンク機能は、コンテンツの作成時点におけるリンクではなく、参照時に、キーワードを用いてハイパーリンクを自動的に生成するため、

- 新たなコンテンツとの動的なリンクが可能
- 作者や読者の気づくことのできなかったコンテンツ間の関連に基づいたリンクを表示するKnowledge Awareness機能が実現可能

という利点がある[3]。

本稿で提案するシステムでは、ノウハウ蓄積システムFISHによる自動リンク機能を一般のコンテンツに拡張した。

2.2.メタデータ

ネットワーク上の様々なデジタルコンテンツを扱うために、コンテンツの属性を明示的に示す

表 1 : Dublin Core Metadata

Element	Label
1. Title	Title
2. Author or Creator	Creator
3. Subject and Keywords	Subject
4. Description	Description
5. Publisher	Publisher
6. Other Contributor	Contributor
7. Date	Date
8. Resource Type	Type
9. Format	Format
10. Resource Identifier	Identifier
11. Source	Source
12. Language	Language
13. Relation	Relation
14. Coverage	Coverage
15. Rights Management	Rights

情報として 15 項目からなる Dublin Core Metadata が RFC2413 で提案されている (表 1)[4][5]。

この 15 項目は、タイトルや概要など情報の内容に関するもの、著者や権利管理など情報の知的所有権に関するもの、日付や言語など情報の属性に関するものの 3 つの種類から成り立っている (表 2)。

表 2 : Dublin Core Metadata の分類

Content	Intellectual Property	Instantiation
Title Subject Description Type Source Relation Coverage	Creator Publisher Contributor Rights	Date Format Identifier Language

様々なデジタルコンテンツにこれらのメタデータを付与してネットワーク上で蓄積することによって、情報資源の活用が活発になることが期待されている。

Dublin Core のメタデータに対する研究事例がいくつかある [1]。CIMI(Consortium for the Computer Interchange of Museum Information)では、メタデータを機械的に付与する研究が行われており、AGLS(Australian Government Information Locator Service)では、サービスや機能に関するエレメントを追加する研究が行われている。また、ISO 10162, 10163 で規定されている情報検索プロトコル Z39.50[6]に Dublin Core のメタデータを導入する研究も行われている。

本稿で提案する方式では、Dublin Core に準拠したメタデータを利用して、コンテンツ自身が持たない情報によるリンクを生成する。

3. コンテンツ流通システム Net-X

筆者らは、Dublin Core に準拠したメタデータの中に、キーワードを利用してハイパーリンクを自動生成し、コンテンツを連携させる情報組織化技術を取り入れたシステム Net-X を開発した。

Net-X により、新しいコンテンツ間の連携が生まれるので、コンテンツ流通の促進が期待できる。

3.1. Net-X の概要

Net-X のシステム構成を図 1 に示す。

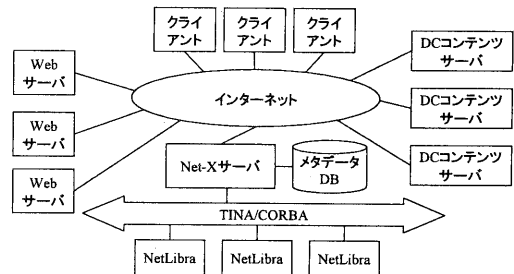


図 1 : システム構成図

Net-X では、次の 3 種類のサーバに格納されたコンテンツを対象とする。

- **Web サーバ :**
既に存在する Internet の Web サーバ
- **DC コンテンツサーバ :**
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コンテンツを扱える専用の Web サーバである。今回は、有料コンテンツを扱う。
- **NetLibra :**
TINA/CORBA の分散オブジェクトを利用して構築された電子図書館システム [7][8]
Web サーバのメタデータは、Net-X サーバのロボット機能により定期的に収集され、メタデータ DB に保存される。DC コンテンツサーバ (Dublin Core 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。NetLibra は、分散オブジェクトである TINA/CORBA を利用して構築されており横断検索が可能である。そこで、Net-X サーバでは、この検索機構を活用し、逐次検索する手法を採用した。
Net-X は、コンテンツ自身の本文に対してリンクを作成する方法と、メタデータ中の要旨にリンクを作成する方法を併用している。これにより、動画や静止画等のデジタルコンテンツへの対応、購入前の有料コンテンツのプレビューへの対応を

実現できる。

なお、自動リンクの基となるキーワードテーブルは、メタデータ中の特定の属性値と、要旨から形態要素解析によって抽出された名詞の集合で構成する。

3.2.Net-X の動作例

Net-X を利用するにあたっては、2 つの立場がある。一つはデジタルコンテンツを提供する立場であり、もう一つがデジタルコンテンツを利用する立場である。ここでは、デジタルコンテンツの利用方法について説明する。利用手順を図 2 に示す。

Net-X 内の有料コンテンツを利用する場合は、会員登録を行う (利用者登録)。登録が完了すると、コンテンツの探索画面へアクセスできるようになる。コンテンツの探索画面例を図 3 に示す。画面の左側は、登録コンテンツの探索用のエリアであり、

- 「分野」「地域」「年代」の階層構造のカテ

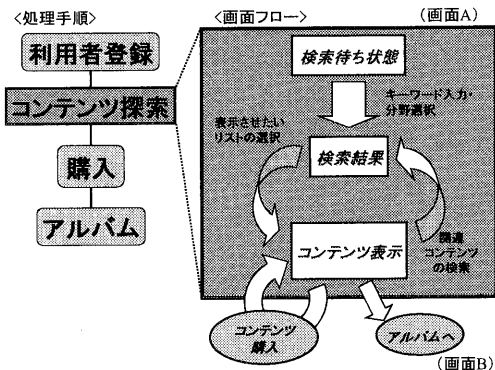


図 2: Net-X の利用手順

ゴリから探索する方法

- 「キーワード」を入力し検索する方法
が用意されている (検索メニュー)。これらの検索および探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 Web コンテンツと DC 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。

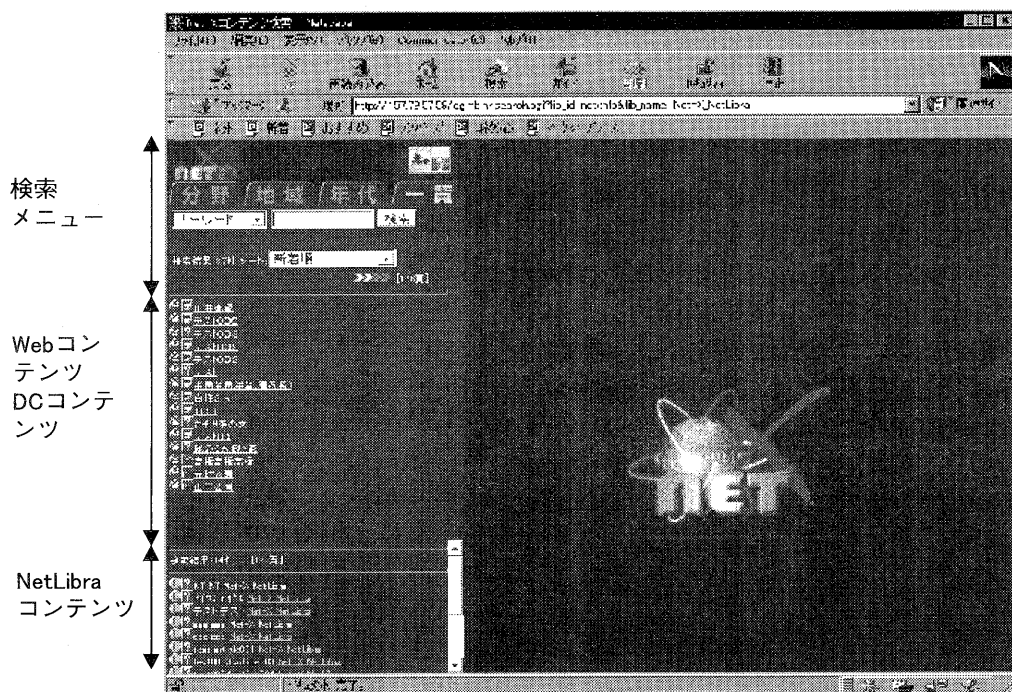


図 3: コンテンツの探索画面 (画面 A)

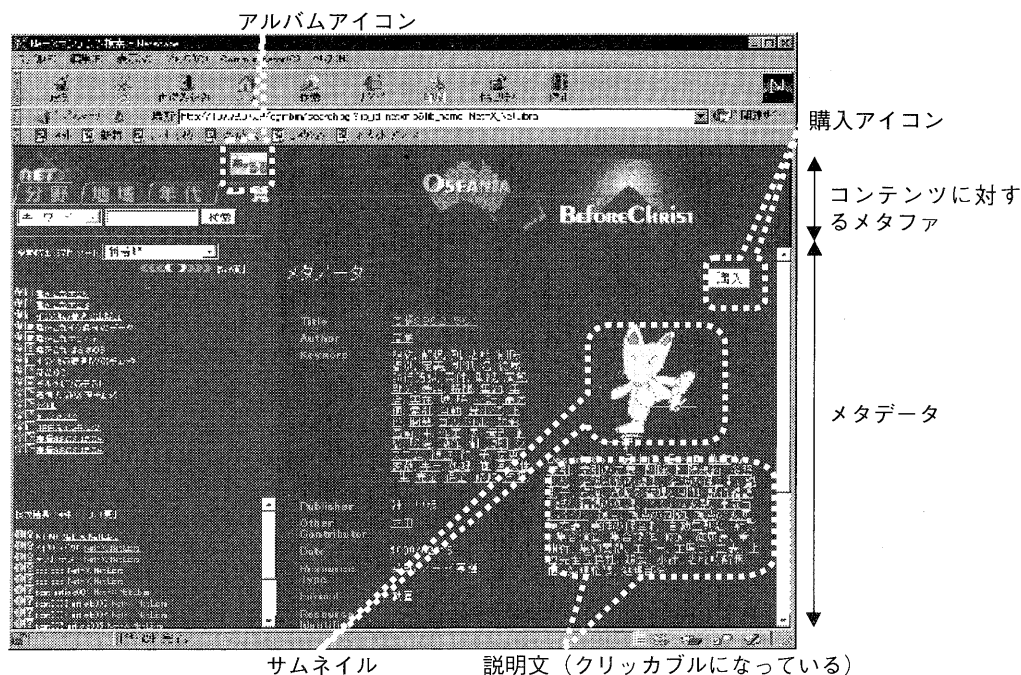


図 4：メタデータ表示画面 (画面 A)

3.2.1.Web コンテンツの表示

検索結果リストの中から Web コンテンツを選択した場合、新しいウィンドウにそのページが表示される。

3.2.2.DC コンテンツの表示

検索結果リストの中から DC コンテンツを選択した場合、右側にそのコンテンツのメタデータが表示される (図 4)。

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 15 項目のメタデータとコンテンツのサムネイルが表示されるが、どの項目を表示させるかはオプションである。

メタデータの概要に関しては、情報組織化機能の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イルを参照しながらの情報探索中に購入する必要のある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する。

3.2.3.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情報組織化機能の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。

3.2.4.購入したコンテンツの表示

情報探索中に電子決済により購入したコンテンツは、自端末に蓄積されアルバム機能により参照できるようになっている (図 5)。

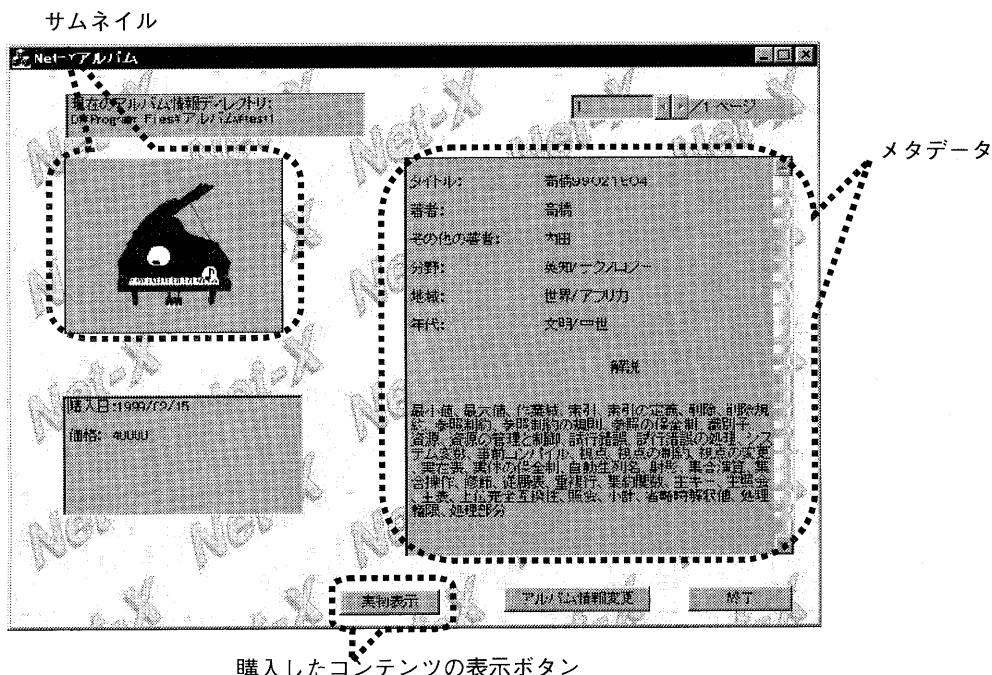


図 5：購入したコンテンツの表示（画面 B）

4. おわりに

本稿では、膨大なコンテンツの中からユーザが望む情報を探し出すことを容易にすることを目的に、情報組織化技術である情報の関連付けと Dublin Core のメタデータを利用して、コンテンツ間に自動リンクを生成し、コンテンツを連携させるシステム 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情報を関連付けるので、増えつづける膨大なコンテンツにも対応でき、また、コンテンツの作者や読者が気づかなかったリンクも生成できる。

今後の課題として、自動リンクの妥当性の評価がある。これに関しては、各種コンテンツを投入し、その性能評価を行う予定である。

参考文献

- [1] 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

イジタル図書館ワークショップ, pp.3-18, 1999.3.

- [2] 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.
- [3] T. Yamakami, Y. Seki, : Proc. of the IFIP TC8/WG8.4 Working Conference, pp.215-225, 1993.
- [4] <http://purl.org/dc/>
- [5] S. Weibel, J. Kunze, C. Lagoze, M. Wolf : Dublin Core Metadata for Resource Discovery, RFC2413, 1998.
- [6] <http://lcweb.loc.gov/z3950/agency/>
- [7] 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.
- [8] <http://www.tinac.com/>