

コンテンツ流通とコンテンツ流通システムの提案

NTT東日本 研究開発センタ

〒180-8585 東京都武蔵野市緑町3-9-11

斉藤典明、日高哲雄、阿部貴博、関良明

あらまし：

インターネットの普及により、ネットワーク上には様々なデジタルコンテンツが流通するようになった。これらのデジタルコンテンツを誰もが効果的に流通できることを可能にするためには、現在のWebページのキーワード検索とは異なるデジタルコンテンツに適したコンテンツを容易に探し当てる方法が必要となる。そこで、今後のコンテンツ流通の特徴を検討すると共に、情報組織化手法を用いたコンテンツ流通システムNet-Xを開発した。

Contents Distribution and Contents-Mediation System

Noriaki SAITO, Tetsuo HIDAKA, Takahiro ABE, Yoshiaki SEKI

NTT-East Research and Development Center

3-9-11 Midori-cho Musashino-shi, TOKYO 180-8585 JAPAN

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

Abstract:

Various digital-contents are distributed on the network after the Internet became popular. To enable any people to distribute digital contents on the network, new digital contents finding method which is different from current searching method of Web pages will be needed. We considered the characteristics of the digital-contents distribution and developed the contents mediation system, Net-X.

1. はじめに

インターネットの普及により、ネットワーク上には様々なデジタル情報が流通するようになった。このデジタル情報には、非常に短い断片的な情報から、書籍をデジタル化したような非常にボリュームがあり作品性の強いデジタルコンテンツまで様々なものがあ

る。これらのデジタル情報を誰もが効果的に流通できることを可能にするためには、現在のWebページのキーワード検索とは異なるデジタルコンテンツに適したコンテンツを容易に探し当てる方法が必要となる。そこで、ネットワーク上のデジタルコンテンツの流通についてモデル化し、そこでの要求条件に

について検討を行った。

2. デジタルコンテンツの流通

ネットワーク上で流通するデジタル情報には、非常に短い断片的な情報から、書籍をデジタル化したような非常にボリュームがあり作品性の強いデジタルコンテンツまで様々なものがある。これらのデジタル情報を次の3つの視点で整理した。

(1) 情報共有

ネットワーク上には様々な有用な情報を持った人達がいる。この有用な情報をもった人達の様々な知識を電子化し、必要なときに必要な人がその情報を取り出せるようにする試みが情報共有である。そして、その手段としてハイパーテキストが用いられた。このハイパーテキストの利用だけでは、現在知ってる既知の情報源からのハイパーリンクを辿るしか必要な情報が探し得なかった。そこで、インターネット上の様々な情報を網羅的に収集し、その収集した情報を検索することにより必要な情報を取り出すサーチエンジンサイトが活用されるようになった。ここでは、必要な情報は必要なときに検索によって探し出すことが可能となった。一方、自分の要求が漠然としている場合や、周辺の情報も得るためには、その時々を検索だけでなく、普段からその情報に接している必要がある。このような場合には、検索ではなく、ニュースグループやコミュニティサイト、メーリングリストのようにあるテーマを持ったネットワーク上のコミュニティを利用する事が有益である。

サーチエンジンサイトやネットワークコミュニティの利用では、検索の場合はWebページごと、ネットワークコミュニティの場合は単体のメッセージごとに、情報が断片的に扱われる。また、そのような断片的な情報はテキ

ストによるメッセージが中心なため、ネットワーク上でキーワード検索により、その情報そのものを検索することが可能となる。そのためサーチエンジンなどの利用ではキーワードにより必要な情報そのものを探し出すことが可能となる。特に情報そのものに対するキーワード検索は常に変化する情報を迅速に捕らえるには適している。

しかしながら、直接的な情報の内容からだけでは、どのような人物の情報なのか、情報の出所はどこなのかなど、情報を取り巻く情報により必要な情報の選別を行うことが困難となる。そのため、必要な情報のあたりをつけることは可能になるが、厳密に絞り込むには困難さを伴う。

(2) コンテンツ配信

ネットワーク上には、未整理で断片的な情報ではなく、編集過程を経るなど体系化された一つの情報がデジタル化されネットワーク上で提供される場合がある。このような情報はネットワーク上の特定の場から広告などを通じて利用者に知らされ、利用者はその発信源へアクセスすることによってコンテンツの配信を受ける。ここでは、テキスト文書化されていないバイナリーデータ化されたコンテンツが多数ある。特に、過去に出版された図書をイメージ情報としてデジタル化し蓄積した電子図書館の試みや、非文書コンテンツを多数所有する電子博物館などの試みである。このようなバイナリーデータに対しては、キーワードにより直接検索することができない。あるいはバイナリーデータの説明のために通常のテキスト情報を混在させる構成がとられている。そのため、正確にコンテンツを探し出すには単純なキーワード検索以外にコンテンツを探し出す方法が必要となる。このため電子図書館では二次情報への検索であ

表1. コンテンツ流通の特徴

	情報共有	コンテンツ流通	コンテンツ配信
情報の形態	テキストが中心	バイナリが中心	バイナリが中心
情報の特性	断片的	作品	作品
探索手法	キーワード検索場(コミュニティ)の活用	今回の課題	場(図書館等)の活用 二次情報への検索
情報取得要求	要求が不明確でも可	要求が不明確でも可	知名度や評判により探索 書名や著者が明確
情報の変化	頻繁	頻繁	緩やか
特徴	あたりをつけるのに適 絞り込みは苦手	無名コンテンツの存在 ネット上不特定多数存在	あたりをつけるのは苦手 絞り込みは得意

る書誌検索が用いられている。

また、図書などは、該当するキーワードを持った書籍を本文に対する全文検索により探すことは妥当ではない。なぜならば、本文には大量の文章を含んでおり、様々な部分がキーワードに対してマッチしてしまい、その図書の内容を言い当てている確実性は低くなる。そしてそのような検索結果は、検索リストのノイズになるからである。

このようなことから、様々な二次情報への検索は、探索中のコンテンツを様々な項目で絞り込んで行く場合に適した方法である。一方で、多数のコンテンツの中から漠然と必要なコンテンツを見つけ出すためにあたりをつけるような場合には適さない。

(3) コンテンツ流通

一方で、インターネットの普及により、様々な人達がデジタルコンテンツを自由に流通させようとしている。これは、これまででは一つのコンテンツを流通させるために、様々な仲介段階を経て、コンテンツの消費者の目に触れていた。一方、インターネット上では、仲介なしにコンテンツ作成者が直接的にコンテンツをコンテンツ消費者に配信することが可能になったからである。

そのため、ネットワーク上では様々な人達がコンテンツを提供することを試みている。し

かしながら、現在のネットワーク上では、コンテンツを伝達する仕組みはあるものの、コンテンツが円滑に流通するための仲介となる仕組みが存在しない。そのため、コンテンツ消費者は目的のコンテンツの存在を知ることなしに、知名度によってコンテンツを探すことになる。

そこで、情報共有の特徴とコンテンツ配信の特徴の間に位置する考え方としてコンテンツ流通の概念を検討した。ここでは、あらゆるデジタルコンテンツがネットワーク上で効果的に流通するための仕組みの実現を目標とする。そのための要求条件を以下に挙げる。

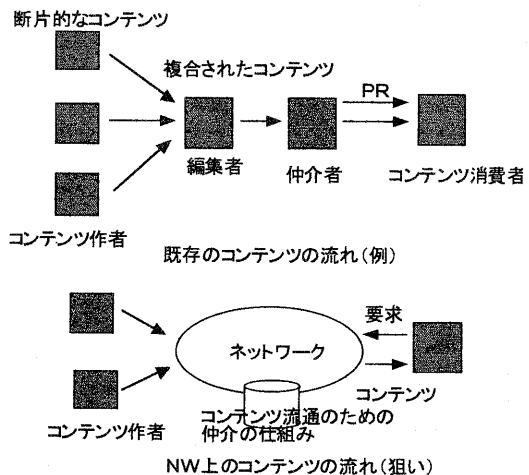


図1. コンテンツ流通の狙い

コンテンツ流通の要求条件：

- ・知名度によらないコンテンツ発見の可能性
- ・内容に基づくコンテンツ発見の可能性
- ・二次情報によるコンテンツ発見の可能性
- ・社会的評価によるコンテンツ発見の可能性

3. Net-Xシステム概要

上記の要求条件のもと、効果的なコンテンツ流通を実現するための機能の実現を検討する。検討にあたってベースとなるシステムとして、コンテンツに付与される二次情報として Dublin Core Metadata を用い、コンテンツの流通を活性化することを狙ったコンテンツ流通システム Net-X を開発した[1,2]。Net-X は、図2のような構成になっており、ネットワーク上の様々なデジタルコンテンツを検索し、さらに検索したコンテンツと関連するコンテンツを情報組織化技術により探し出す仕組みをもつ。Net-X で扱うデジタルコンテンツは Dublin Core Metadata に基づいて構成された DC コンテンツ、DL サーバに蓄積されている DL コンテンツ、インターネット上の通常の Web ページである I コンテンツの3種類のデジタルコンテンツを扱うことが可能である。

・DCコンテンツは、コンテンツの登録時にそのメタデータが Net-X サーバに登録、蓄積

表2. Dublin Core Metadata (RFC2413)

	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

される。

・DLコンテンツはコンテンツの登録時にその書誌情報がメタデータとして DL サーバに登録されている。DL 内のコンテンツおよび書誌情報は、Net-X サーバからは DL の検索インターフェースを介して操作される。

・Iコンテンツは、インターネット上の通常の Web ページであり、Net-X サーバのロボットにより検索される。検索された結果は Dublin Core Metadata へマッピングされ Net-X サーバで蓄積、管理される。

Net-X では、これらの3種類のコンテンツを Dublin Core Metadata の形式で扱い、検索だけでなくコンテンツ相互を情報組織化機能により結びつける。本論文では、この情報

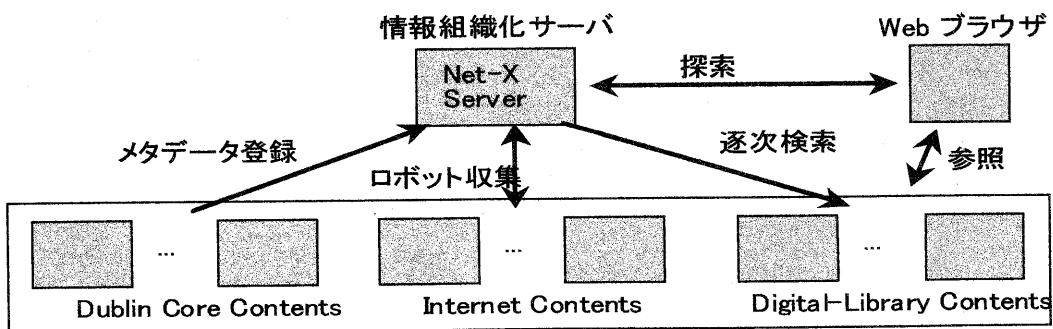


図2. Net-Xシステム構成

組織化機能を用いてデジタルコンテンツがより効果的に流通できる方式の実現について述べる。

4. コンテンツ探索手法

Net-Xでは、ネットワーク上のコンテンツをコンテンツ消費者が探し出すための機能として、コンテンツの検索機能と、関連コンテンツへのハイパーリンク機能を有しており、これらについて以下に説明する。

(1) コンテンツ検索

コンテンツをはじめに探し出すための機能として、Net-Xではキーワード検索とカテゴリ検索を実現した。

キーワード検索は、オプションとしてメタデータのディスクリプションに対する検索、コンテンツの著者に対する検索、コンテンツの題目に対する検索の3種類が可能である。これはDCコンテンツに対しては、メタデータのうち概要と、著者、題目の3種類の情報を検索テーブルで管理しており、それぞれに対して検索を行う。DLコンテンツに対してはDLのそれぞれの検索インターフェースに対して検索を行う。ここではDLシステムは、NetLibraを用いた。Iコンテンツに関しては、ロボットで収集されたデータは全てDCメタデータの概要の情報として扱っている。

カテゴリ検索に関しては、分野、地域、年代による3種類が可能である。各項目について、Net-Xサーバ内でそれぞれの要素間の関係を階層構造によって管理している。カ

テゴリ検索の際に指定された階層の属性を持つコンテンツがリストアップされる仕組みになっている。

DCコンテンツについては、分野はResource Type エlement内のデータを持ち、地域と年代についてはCoverage Element に地域と年代のサブElementを追加することによって実現した。

DLコンテンツについては、書誌情報の中に特定様式で分野、地域、年代のデータを記録した場合において、Net-Xサーバ内の階層情報を参照してカテゴリ検索が可能となる。

Iコンテンツについては実装は行っていない。

(2) コンテンツ間のハイパーリンク

Net-Xでは、参照しているコンテンツからさらに関連するコンテンツを探し出す情報組織化の機能を備えている。これはハイパーリンクにより様々な関連するコンテンツがつながる仕組みである。我々の研究ではこれまで情報共有の分野でこの手法を用いてきたが、ここでは様々なハイパーリンクが混在してしまう問題点があった。そこで、ここでは3種類のハイパーリンクを色分けにより明示する仕組みを実現した。

DCコンテンツに関しては、メタデータ参照の際に、著者に関するリンク、題目に関するリンク、任意のキーワードに関するリンクがDCメタデータの概要の部分に埋めこまれる仕組みである。

DLコンテンツに関しては、コンテンツ本体のテキスト部分に関して、著者、題目、任意のキーワードのハイパーリンクが埋めこまれる。

Iコンテンツに関しては、参照しているWebページのテキスト部分をNet-Xサーバに送

表 3. Net-Xにおける検索方式

	DCコンテンツ	DLコンテンツ	Iコンテンツ
キーワード検索			
概要	○	○	○
著者	○	○	×
題目	○	○	×
カテゴリ検索			
分野	○	△	×
地域	○	△	×
年代	○	△	×

表4. Net-Xにおけるハイパーリンク方式

リンク元	リンク先		
DCコンテンツ→	DCコンテンツ	DLコンテンツ	Iコンテンツ
著者	○	○	×
題目	○	○	×
キーワード	○	○	○
DLコンテンツ→	DCコンテンツ	DLコンテンツ	Iコンテンツ
著者	○	○	×
題目	○	○	×
キーワード	○	○	○
Iコンテンツ→	DCコンテンツ	DLコンテンツ	Iコンテンツ
著者	△	△	×
題目	△	△	×
キーワード	△	△	△

信し (HTMLのPOSTメソッドを用いたページを作成する)、関連コンテンツへのハイパーリンクを埋め込み返送する機能を実現した。

5. コンテンツ流通への適用と課題

誰もがネットワーク上でデジタルコンテンツを提供可能とするには、現在のWebページのキーワード検索とは異なるデジタルコンテンツに適したコンテンツを容易に探し当てる方法が必要となる。そこで、今後のコンテンツ流通の特徴を検討した。Net-Xをこの特徴に基づいて検討し、今後の課題を明らかにする。

(1) 知名度によらないコンテンツ発見の可能性 : コンテンツの概要に関する任意のキーワードによる検索の他、情報組織化機能により参照しているコンテンツのテキスト情報の中に他の関連コンテンツへのハイパーリンクが埋めこまれる。このため、知名度に依存しないでコンテンツを探し出すことが可能となる。

(2) 内容に基づくコンテンツ発見の可能性 : DCコンテンツの場合、コンテンツ本文への検索は行なわないが、メタデータの概要部分に、コンテンツをよりの確に説明する文章があれば可能となる。このメタデータの概要部分を本文よりも効果的に扱える手法の検討については、今後の課題として必要である

う。

(3) 二次情報によるコンテンツ発見の可能性 : 今回は著者、題目、分野、地域、年代という項目に基づいてコンテンツを探し出す仕組みを実現した。ただし、コンテンツの探索において「あたりをつける→絞り込む」という過程を意識した手法の検討が必要である。

(4) 社会的評価によるコンテンツ発見の可能性 : 現在の方法では、一旦付与したメタデータの値の変更は容易ではない。そのため、コンテンツの参照動向に基づいた、ネットワーク上の評価となるような動的な情報を記録できるような仕組みが必要であろう。

6. まとめ

誰もがネットワーク上で容易にコンテンツを流通可能とするための環境を検討し、これを実現するためのシステム化の要求条件の検討を行った。この検討の下、コンテンツ流通システムを実現した。今後は本要求条件の下、機能拡張を行う予定である。

表5. Net-Xの特徴

要求条件	対応
(1) 知名度によらない検索	○
(2) 内容の検索	○
(3) 二次情報の検索	○
(4) 社会的評価による検索	×

参考文献

- [1] 日高他, "情報組織化によるコンテンツ流通システム: Net-X", 情報処理学会研究会報告 GW32-2, p. 5-10, 1999.
 [2] N. Saito 他, "Net-X: Contents-Mediation System with Information-Organizing Technology", APSITT' 99 Proceedings, p. 308-312, 1999.