

XML を利用した知識流通システムのアーキテクチャ

佐藤宏之 大友健治 増尾剛

{sato.hiroyuki, otomo.kenji, masuo.tsuyoshi}@lab.ntt.co.jp

NTT 情報流通プラットフォーム研究所

コンテンツに関わる背景情報（コンテキスト）をグループやコミュニティ間で流通することによって、知的創造活動を支援する知識流通システムを提案している。コンテキストとは、ユーザの複数のコンテンツに関わる行為から、それらのコンテンツ間の関連を知識として抽出したものである。本論文では、XML を用いたコンテキストの表現、および流通のアーキテクチャ、本方式の利点について述べる。

The Architecture of a Knowledge Sharing System using XML

Hiroyuki Sato Kenji Otomo Tsuyoshi Masuo

{sato.hiroyuki, otomo.kenji, masuo.tsuyoshi}@lab.ntt.co.jp

NTT Information Sharing Platform Laboratories

A knowledge sharing system which enables users to share background information about contents among groups or communities is proposed for intellectual creation works. The background information composed of relationships among the contents is called "context". We extract the context as knowledge from user's activities when they handle several contents. This paper explains representation of the context using XML, architecture for knowledge sharing, and merits of this method.

1. はじめに

World Wide Web (WWW)に代表されるように、現在ネットワーク上にはさまざまなコンテンツが溢れている。電子メディアの発達に伴い、オフィスなどにおいて、電子的なドキュメントの参照、作成、レビューなどの、「ドキュメントベース」で作業が進むこともコンテンツ数の増加に拍車をかけている。そのため、溢れるコンテンツを有効に活用して、どのように知的創造活動を行うかが大きな課題となっている。本研究では、この課題を解決するため、ユーザのコンテンツの参照、利用といった行為などから、コンテキストと呼ぶコンテンツ間の関係情報を、コンテンツの活用に関わる知識として抽出し、流通する知識流通

システムを提案する。本論文では、XML を用いたコンテキストの表現、および流通のアーキテクチャ、本方式の利点について述べる。

2. 知的創造活動におけるコンテンツの活用

コンピュータを利用した知的創造活動では、まずネットワーク内に埋もれている情報、すなわちコンテンツをユーザに認知させることが重要である。何かを創造したり、新たな作業を行うときには参考とする情報が必要となる。現在、多くの検索エンジンがキーワードに基づいてネットワーク上の情報を探索し、ユーザへ大量のコンテンツを提示することに成功している。

このときに、コンテンツに加えて、「過去にどのような状況でどう利用されたのか」といった背景情報があると、ユーザが自身の創造物や作業にそのコンテンツの内容を反映させる方法がわかったり、役に立つかの判断ができたりすることがある。このため、従来からコンテンツにさまざまな属性を表すメタタグを付与することによって、情報探索時に付加的なコンテンツの特徴をユーザに提示することが考えられている。

しかし、従来の方法では、背景情報を詳しく付与しようとするほどタグの種類が多くなり、登録に時間や人手のコストがかかるようになる。記述内容の一貫性にも多大な注意を払わなければならない。また、コンテンツの背景にある情報は、さまざまなユーザの独自のノウハウなどに基づく行動などから形成されると考えられ、1つの静的な特徴を記述するのが難しいと我々は考えている。

1つのコンテンツに対していろいろな観点から利用するユーザがいるため、次々と新しい背景情報が追加されることもある。

そこで、本研究では、コンテンツの背景を示す情報として、コンテンツ間の関係（リンク）情報に着目した。複数のユーザによる背景情報の動的な追加を簡易に記述できる柔軟性をもち、さらに複数のユーザの作業プロセスをナビゲーションする構造をもったリンク情報を、ユーザ間で流通することによって、知的創造活動におけるコンテンツの再活用が促進されると考える。我々はユーザの行動などに基づいたコンテンツ間のリンク情報から生成された流通可能なコンテンツの背景情報を「コンテキスト」と呼んでいる。

3. コンテキスト

3.1 リンクによるコンテキストの表現

コンテンツの背景情報をリンクとして表わすのは、知的創造活動において、ユーザが他のコンテンツを参照するときには、コンテンツが利用されたプロセスなどをひとかたまりにしたコンテンツ群が、再活用のための参考になると考えるからである。コンテンツ群はコンテンツ間のリンク情報

で構成できる。1つのコンテンツが複数のコンテンツ群の構成要素となることによって、そのコンテンツがもつさまざまな背景や特徴を動的に提示できると考える。

さらに、リンクを辿るという行為は、人間の発想や記憶と親和性が高いと考えられる。連想記憶モデルでは、関係のある事柄の結びつきを記憶しその事柄の一部から全体を思い出すとされ、リンク情報が記憶において重要な役割を果たしている。また、興味、関心、必要性に応じて発見的に情報を辿るという行為、刺激が新たな発想を想起させるということが期待できる。

3.2 コンテキストの分類

我々は、目的とするコンテンツに効率的にアクセスしたり、効果的に利用したりするために必要なコンテキストの種類を考察した[1]。

(1) 分類／カテゴリに関する情報

ネットワーク上のコンテンツは、内容が同じコンテンツでも、個々のユーザによって別の分類やカテゴリに従って整理されていることが多い。ディレクトリ構造（ファイルシステム、Web サイト、ブックマークなど）内でのコンテンツの位置付けは重要な背景情報と成り得る。

(2) 関連するコンテンツ

コンテンツ作成者の意図に基づき関連付けられたコンテンツと、コンテンツ作成者以外の第三者によって関連付けられたコンテンツがある。前者にはコンテンツ内にリンク情報として埋め込まれたコンテンツや、修正の前後といったバージョンの異なるコンテンツなどが考えられる。後者には第三者による、引用、参照などの作業プロセスや、注釈、コメント、メッセージなども含まれる。

(3) 関連する人や組織

コンテンツの作成者や管理者だけでなく利用者からもコンテンツの重要度を計れることがある。

4. コンテキストの記述方法

前述のように、コンテキストは後から動的に生成されることがあるため、既にネットワーク上に

存在するコンテンツとは別に第3者的に記述されて管理されることが必要である。また、更新・引用などの一般的なコンテンツを扱う行為に基づいて簡易に個々のリンクが生成でき、複数人で自由に付加できる柔軟性を備える必要もある。本研究ではこれらを考慮し、XMLにおいてハイパーリンクを表現する XLink[2]の仕様を拡張した表現形式を、コンテキストの記述に利用した。

4.1 XLink

XLink (XML Linking Language)はHTMLのアンカータグ<A>に相当する機能をXMLにおいて利用するための仕様である。しかし、HTMLにおけるハイパーリンク機能に加え、以下に示す機能がある。

- 双方向のリンクが記述できる
- 1つのリンク元から複数のリンク先を記述できる
- リンクの記述を、リンク元データとリンク先データの両方から独立して記述できる
- リンクに対して意味を記述できる
- リンクを辿るタイミングやブラウザなどによる表示時の処理を記述できる

XLinkはW3C (World Wide Web Consortium)において標準化が進められている。2000年12月に勧告案となっており、諮問委員会のレビューに基づく投票結果とディレクタの裁定次第では、近くに正式なW3C勧告となる可能性がある。

4.2 コンテキストのXMLによる記述

XLinkのリンク表現を利用して、複数のコンテンツ間の関係を記述するイメージを図1に示す。コンテンツAを中心に、コンテンツAがコンテン

ツBを引用して作成され、その後、コンテンツAをコピーしてコンテンツCが作成された様子を表わしている。これらのコンテンツ間の関係をコンテンツとは独立して記述しているのが、contentA_context.xmlと表示されたコンテキスト(XMLファイル)である。

図2にこのXMLファイルの内容を示す。また図3にコンテキストを記述するための文法規則をDTD (context.dtd)によって定義している様子を示す。ここでは、コンテキストの構成や、リンクの意味付けなどに関する情報の記述に、新たに我々が <http://www.ntt.co.jp/2000/context> の名前空間を用いて宣言した context の接頭辞が付いたタグや属性を利用する。もちろん、W3CのXLink仕様で既に規定されているリンクに関するタグや属性はそれに従うことを宣言している。

<context:context>要素を構成するものとして、コンテキストの内容が記述されている。コンテキストは大まかに、中心となるコンテンツをリソースとして指定する部分(コンテンツ情報: <context:content>要素)、関係のあるコンテンツをターゲット・リソースとして指定する部分(それと中心となるコンテンツとの意味付けをする部分(リンク情報等: <context:locator>要素、<context:action>要素)から構成されている。後者は複数記述することが可能である。

図2の例では、<context:content>要素に、コンテンツAに関するURLや作成情報が記述されている。また、図2最後の<context:locator>要素と<context:action>要素のペアによって、コンテンツAとコンテンツCとの関係が記述されている。<context:locator>要素では、コンテンツCに関するURL、生成に使われたアプリケーション名など

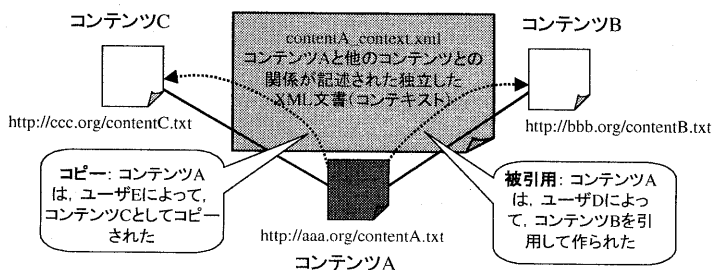


図1: 複数のコンテンツ間の関係を記述するXML文書(コンテキスト)のイメージ

<pre> <?xml version="1.0" encoding="UTF-8"?> <!DOCTYPE context:context SYSTEM "context.dtd"> <context:context> <!-- コンテンツ情報 --> <context:content xlink:role="document" stamp="2001/03/05 10:23:00 +0900"> <context:appID>FileManager</context:appID> <context:url>http://aaa.org/contentA.org</context:url> <context:creation-date>2000/11/03 00:30:32 +0900 </context:creation-date> </context:content> <!-- 登録者情報 --> <context:locator xlink:title="userD" xlink:label="action1" xlink:role="http://www.ntt.co.jp/2000/context/user" xlink:href="userD@sample.org" appID="FileManager"/> <context:action xlink:arcrole="http://www.ntt.co.jp/2000/context/creator" xlink:to="action1" time="2000/11/03 00:30:32 +0900"/> </pre>	<pre> <!-- 以下リンク情報等 --> <context:locator xlink:label="action2" xlink:role="http://www.ntt.co.jp/2000/context/document" xlink:href="http://bbb.org/contentB.txt" appID="FileManager" time="2000/11/03 00:30:32 +0900"/> <context:action xlink:arcrole="http://www.ntt.co.jp/2000/context/cited" xlink:to="action2" user="userD"/> <context:locator xlink:label="action3" xlink:role="http://www.ntt.co.jp/2000/context/document" xlink:href="http://ccc.org/contentC.txt" appID="FileManager" time="2001/03/05 10:22:51 +0900"/> <context:action xlink:arcrole="http://www.ntt.co.jp/2000/context/copy" xlink:to="action3" user="userE"/> </context:context> </pre>
--	--

図 2: コンテキストの記述例 (contentA_context.xml)

<pre> <?xml version="1.0" encoding="UTF-8"?> <!ENTITY % link " (context:locator , context:action) "> <!ELEMENT context:context (context:content, %link:*)> <!ATTLIST context:context xmlns:xlink CDATA #FIXED "http://www.w3.org/1999/xlink" xmlns:context CDATA #FIXED "http://www.ntt.co.jp/2000/context" xlink:type (extended) #FIXED "extended" xlink:title CDATA #IMPLIED stamp CDATA #REQUIRED > <!ELEMENT context:content (context:appID+, context:url, context:creation-date)> <!ATTLIST context:content xlink:type (resource) #FIXED "resource" xlink:title CDATA #IMPLIED xlink:label (content) #FIXED "content" xlink:role CDATA #REQUIRED security CDATA #IMPLIED stamp CDATA #REQUIRED > </pre>	<pre> <!ELEMENT context:locator (context:comment*)> <!ATTLIST context:locator xlink:type (locator) #FIXED "locator" xlink:title CDATA #IMPLIED xlink:label NMTOKEN #REQUIRED xlink:role CDATA #REQUIRED xlink:href CDATA #REQUIRED appID CDATA #REQUIRED security CDATA #IMPLIED stamp CDATA #REQUIRED > <!ELEMENT context:action EMPTY> <!ATTLIST context:action xlink:type (arc) #FIXED "arc" xlink:title CDATA #IMPLIED xlink:arcrole CDATA #REQUIRED xlink:from (content) #FIXED "content" xlink:to NMTOKEN #REQUIRED appID CDATA #IMPLIED time CDATA #IMPLIED stamp CDATA #REQUIRED user CDATA #IMPLIED > <!ELEMENT context:appID (#PCDATA)> <!ELEMENT context:url (#PCDATA)> <!ELEMENT context:creation-date (#PCDATA)> <!ELEMENT context:comment (#PCDATA)> </pre>
---	---

図 3: context.dtd (部分的に抜粋して編集したもの)

が記述されている。<context:action>要素には、action3 とラベル付けされたコンテンツ C は、コンテンツ A を userE がコピーして生成したことが示されている。この他、図 2 では、コンテンツ A

とその登録者 (userD) や、コンテンツ B との関係についての記述例も示している。

図 3 の DTD は、我々が定義した context.dtd の中から図 2 で示した例を定義する部分を抜粋し、

それらだけで完結するように編集し直したものである。実際の context.dtd では、ディレクトリにおけるコンテンツをノードとしたツリー構造の親子関係などについて記述する仕様も定義している。

5. 知識流通システムのアーキテクチャ

前述のコンテキストを個人やグループ間で流通することが、知的創造活動に必要な知識の共有につながることを考え、知識流通システムのアーキテクチャを提案する。図4にシステムの概要を示す。

5.1 コンテキストの入力

コンテキストはコンテンツを利用するユーザの挙動などから取得され、ユーザが大きなコストをかけなくても、流通可能な情報として蓄積されていくことが重要と考える。そこで、ユーザが利用するさまざまなアプリケーション（例えば、文書作成ソフトウェア、ファイルマネージャ、Webブラウザとそのブックマークエディタなど）において生じた3.2で示したコンテンツ間の関係を、4.2で示したXML表現でマッピングしてコンテキストを生成する機能を提供することが考えられる。

このような機能が付加されたアプリケーションは知識流通クライアントと呼ばれ、イントラネットなどのドメイン毎に存在するコンテキストビューロ（知識流通サーバ）と通信を行う。この通信の中で、コンテキストビューロは各クライアントで生成されたコンテキストであるXMLファイルを受信して、蓄積・管理する。

ユーザには、蓄積されたコンテキストを参照し、GUIによる操作で内容を修正したり、追加し

たりすることを可能にするコンテキストエディタも提供される。

5.2 コンテキスト流通機構

Webブラウザのようなアプリケーションからも各ユーザのコンテキストを取得した場合、その量は膨大となる。1つのコンテンツに関して個々のユーザの行為や意味付けといった重複や矛盾を含み得る、動的に更新される膨大な量のコンテキストを、1つのコンテンツに関して一貫性を維持して管理することは困難である。

そこで、コンテキストの管理と流通は集中管理型ではなく分散管理型にしている。我々はネットワーク上のファイル交換サービスであるGnutella[3]のようなPeer to peer型のプロトコルに着目した。Gnutellaでは、流通するファイルはGnutella servantsと呼ばれるサーバとクライアントを兼ねた機能単位でのみ管理され、ホップ数で指定された範囲内のGnutella servantsが管理するファイルが交換、流通の対象となる。コンテキストもある機能単位で管理し、到達可能な範囲をホップ数により制御できるPeer to peer型のプロトコルがその流通に適していると考えられる。

コンテキストを管理するコンテキストビューロでは、他のドメインのコンテキストビューロとインターネットなどのネットワークを介してPeer to peer型のプロトコルによる通信を行い、関連するコンテキストの有無を問い合わせ、存在する場合はそれを取り込むことを可能とする。ホップ数を指定してコンテキストビューロ間の通信が行われることによって、異なるドメイン間のコンテキストの流通がXMLファイルの受け渡しによって行われる。

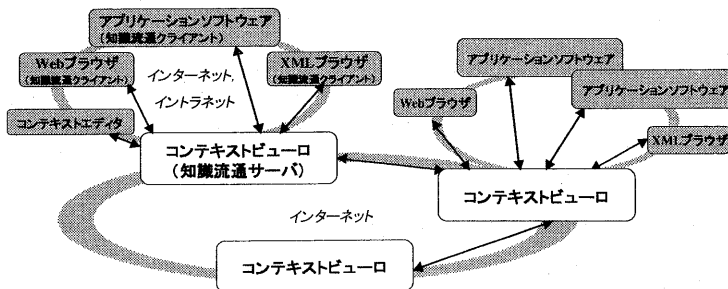


図4: 知識流通システムの構成

5.3 コンテキストの活用

アプリケーションによって生成され、コンテキストビューロによって蓄積・管理され、さらに流通されるコンテキストを、必要に応じて、ユーザにフィードバックすることが重要である。

単純に Web ブラウザや XML ブラウザを用いて、コンテキストビューロにアクセスして、あるコンテンツに関するコンテキストを参照することは可能であるが、フィードバック方法を工夫することにより、知識流通システムをプラットフォーム技術とした次のような応用例が考えられる。

検索エンジン

キーワード検索を行う検索エンジンに加えて、コンテキストに含まれるコンテンツ間の意味付けされたリンク情報を活用して、検索精度を向上させることが考えられる。

ナレッジマネジメント

近年、企業内で分散していた情報やノウハウを共有・活用する仕組みを提供し、業務アウトプットの質を向上させたり、効率化を図ったりする試みとしてナレッジマネジメント[4]が注目されている。知識流通システムの運用により、1つのコンテンツからその生成に関わる背景やその利用のされ方などを、リンクを辿って参照することが可能になり、マニュアルがなかったり、ワークフローが定型化していない業務における個人が有する知識の共有を支援することができる。

6. 考察

4.の仕様と5.の方式のメリットを考察した。

知識を流通するためのフォーマットとしてXMLを利用したメリットには、既存のアプリケーションにそれを利用するユーザの行為からコンテキストをマッピングする機能が付加し易く、フォーマットが統一された大量のコンテキストを取得し易くなることが挙げられる。また、XML自体がデータ交換に適しているという特性をそのまま享受し知識流通のサーバを分散型で構築し易いということもある。さらに XLink を用いることにより、コンテンツ間の意味付けと、ナビゲーションの両方を簡単に記述できるというメリットがあ

る。将来的に多くの Web ブラウザなどのアプリケーションが XLink に対応することにより、簡単にコンテキストを用いたコンテンツ探索を行うことができるようになると考えられる。

知識流通システムのアーキテクチャは、ユーザに対して、ネットワーク上に散在するコンテンツ間に存在する関係を把握しながら、興味のある関係を辿って情報探索する手段を提供する。Peer to peer 型の通信を用いたため、まずは自分が所属するドメインや自分が知っているドメインからコンテキストを収集し、次にその過程で興味をもったコンテンツをキーにして、別のドメインからさらにコンテキストを集めることにより、知的創造活動における情報探索の効率を向上させることができると考えられる。

7. おわりに

コンテンツに関わる背景情報を XML を用いて表現し、グループやコミュニティ間で流通することによって知的創造活動を支援する知識流通システムのアーキテクチャを提案した。

今後は、実運用を通して、さまざまなアプリケーションから必要なコンテキストを取得し、システムのパフォーマンスを検証する。また、この知識流通システムをプラットフォーム技術とした応用システムを構築し、利用シーンに合わせたコンテキストの提示手法の評価を行う予定である。

参考文献

- [1] 大友健治, 佐藤宏之, 増尾剛: 知識流通システムによる知的創造活動支援, 情報処理学会第 62 回全国大会, 8B-4 (2001).
- [2] W3C XML Pointer, XML Base and XML Linking, <http://www.w3.org/XML/Linking>.
- [3] Gnutella Protocol Specification v0.4, <http://dss.clip2.com/GnutellaProtocol04.pdf>.
- [4] T.H.Davenport, ナレッジマネジメント実践法, Diamond Harvard Business, August-September 1999, pp.26-36 (1999).
- [5] 佐藤宏之, 神戸雅一, 小村誠一: コミュニティにおける知的創造活動を支援する知識流通方式, 情報処理学会第 61 回全国大会, 2k-1, 4-13 (2000).