マルチメディア通信と分散処理　109－18
電子化知的財産・社会基盤　17－18
（２００２．９．１３）

Content-based networkにおけるオントロジー自動変換システムの提案

酒井 隆道† 太田 昌克 寺元 光生
NTT未来ねっと研究所

近年，内容指向のメッセージ配送を行うContent-based network（CBN）がアプリケーション層ネットワークの一形態として着目を浴びている．CBNをグローバル環境において適用する場合，異なるオントロジーを有するエンドユーザ間におけるメッセージ配送を実現するための，オントロジー変換技術の確立が課題となる．本稿では，このようなオントロジー変換を実現するためのアーキテクチャを示し，フレーム形式のオントロジーモデルを対象としたオントロジー自動変換方法を提案する．

An ontology transformation system in content-based network

Takamichi Sakai† Masakatsu Ohta Mitsuo Teramoto
NTT Network Innovation Laboratories

A content-based network(CBN), that performs a content-based routing, is paid attention as a kind of an application-layer network in recent years. When we lay out the CBN at a global environment, we must establish an ontology transformation technique in order to realize the message routing between end-users of different ontologies. In this paper, we propose an ontology transformation architecture and an automatic ontology transformation method for the frame-based ontology model.

1. 序論

近年，Gnutella，JXTA等のP2Pネットワークが着目を浴びているが，それらの特徴のひとつとして，それがアプリケーション層（overlay）のネットワークであることが挙げられる．この特徴によって，下位層であるIPの束縛から解放され，自由で柔軟な名前空間やネットワーク構成を実現可能としている．

このようなアプリケーション層のネットワークの一形態として「Content-based network（以後CBNと略す）」が提案されている[1][2]．これは，メッセージ内に記述されたメタ情報に基づいた内容指向の配送（routing）を行うネットワークのことである．従来型ネットワークにおいては，各ノードに対して一意にIP等のネットワークアドレスが付与され，メッセージはそのヘッダにおいて指定されたネットワークアドレスを持つノードに対して配送される．これに対してCBNでは，各ノードは自身が受信したいメッセージに関する条件を記述したフィルタをCBNに登録し，メッセージはそのヘッダに記述されたメタ情報と合致するフィルタを登録しているノードに対して配送される．

†sakai@ma.onlab.ntt.co.jp

CBNでは広域（global）な環境に適用できることを主目標の一つに掲げており，その実現に際しては二つの大きな課題がある．一つ目はルーティングである．広域環境においてスケーラブルなメッセージ配送を実現できるルーティングアルゴリズムの考案が課題となる．そして二つ目が本稿で扱うオントロジーである．ここで，オントロジーとはメタ情報とフィルタの構造（データモデル）や用いる語彙を規定する仕様のことであり，CBNにおけるアドレス体系となる．CBNでは，まさにその内容指向な特性によって「内容の理論（Content-based AI）[3]」であるとされるオントロジー工学の手法が活用できる．

CBNにおける具体的なオントロジーの課題として，特に広域環境を想定した場合，異なるオントロジーを有するエンドユーザ間でのメッセージ配送を実現するための，メッセージのオントロジー変換を実現する課題がある．なぜならば，従来型ネットワークにおいては階層的で一意な標準アドレス体系をグローバルに規定することができたが，オントロジーに関してグローバルな標準体系を確立することは現実的でない．

以上の状況を踏まえ，本稿ではまず第2.節においてCBNにおけるオントロジー変換を実現するための

アーキテクチャを示す．続く第 3. 節において具体的なオントロジーモデルを定義し，そのオントロジーモデルを対象とするオントロジー自動変換方法の提案を第 4. 節において行う．そして第 5. 節，第 6. 節において実装と関連研究の説明を行い，第 7. 節でまとめを行う．

2. オントロジー変換アーキテクチャ

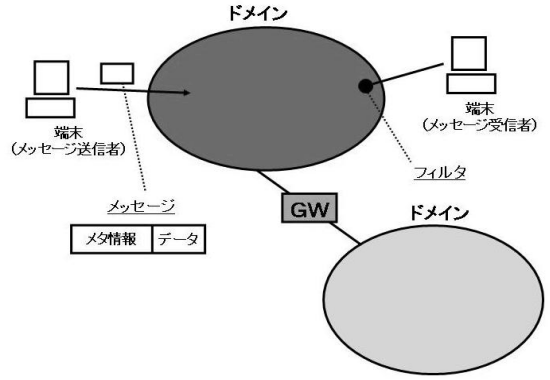


図 1: アーキテクチャの概念図

CBN においてオントロジー変換を実現するためのアーキテクチャの概念図を図 1 に示した‡．まず，メッセージ受信者は，自身が受信したいメッセージに関する条件を記述したフィルタ (filter) を CBN に登録しておく．そして，メッセージ送信者は，データとそのメタ情報（メタデータ）から構成されたメッセージを CBN に対して送信する．CBN は，ヘッダにおいて指定されたメタ情報と合致するフィルタを登録しているメッセージ受信者に対して，メッセージを配送する役目を担う．

次に，図 1 に示す通り，オントロジーの一意性が保証された空間としてドメインを規定する．ここで，オントロジーとはメタ情報とフィルタの構造（データモデル）や用いる語彙を規定する仕様のことであり，CBN におけるアドレス体系となる．すなわち，ドメイン内においてはオントロジーを共有することで，メタ情報とフィルタ間の照合の正当性を確保する．

そして，ドメイン間にはゲートウェイ（以後 GW と略す）を設置する．本 GW がメッセージを転送する際に，そのオントロジー変換を行うことによってドメイン間のメッセージ共有（流通）を実現する．具体的には，GW はメッセージ内のメタ情報に関して，転送元ドメインから転送先ドメインへのオントロジー変換を行う

‡ここに示したアーキテクチャは，CBN を実現する一解にすぎない．

2.1 オントロジー変換方式

上記アーキテクチャにおいて，複数のドメインが存在する場合のオントロジー変換方式として，3 つの方式が考えられる．以下それぞれの方式について図 2〜図 4 を用いて説明した後に，それらの方式の特性に関する比較検討を行う．

図 2〜図 4 は，いずれもドメイン 1，2，3 が GW を介して直列に連結された状況下で，ドメイン 1 において送出されたメッセージがドメイン 2，そして 3 に転送される様子を示している．図中のメッセージの括弧（）内の数字は，このメッセージがどのドメインのオントロジーに基づいて記述されたものであるのかを示している．なお，ここでは直線のトポロジを用いて議論を行っているが，以下の議論は他のトポロジにおいても同様に成立する．

2.1.1 オントロジー変換方式 1

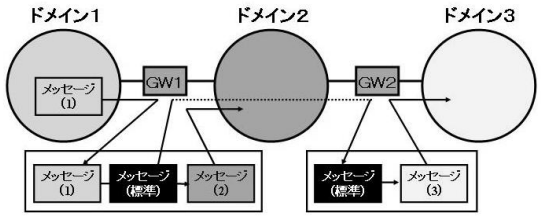


図 2: オントロジー変換方式 1

本方式は，メッセージを常に標準オントロジー（共有オントロジー）に変換してから転送する方式である．図 2 において，GW1 はドメイン 1 内で送出されたメッセージを標準オントロジーへと変換し，メッセージのコピーをドメイン 2 をトンネリングして GW2 に引き渡すと共に，本メッセージをドメイン 2 のオントロジーへと変換してドメイン 2 へ送出する．GW2 は，GW1 から受け取ったメッセージをドメイン 3 のオントロジーへと変換し，ドメイン 3 へ送出する．

2.1.2 オントロジー変換方式 2

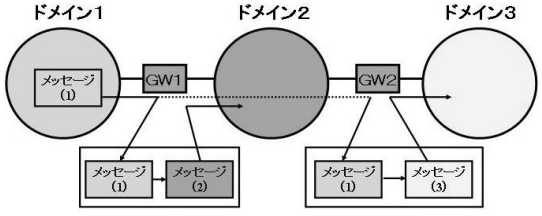


図 3: オントロジー変換方式 2

本方式は，常にメッセージの送信元ドメインと送信先ドメイン間での一対一の変換を行う方式である．

図 3 において，GW1 はドメイン 1 内で送出されたメッセージを受信し，メッセージのコピーをドメイン 2 をトンネリングして GW2 に引き渡すと共に，本メッセージをドメイン 2 のオントロジーへと変換してドメイン 2 へ送出する．GW2 は，GW1 から受け取ったメッセージをドメイン 3 のオントロジーへと変換し，ドメイン 3 へ送出する．

2.1.3 オントロジー変換方式 3

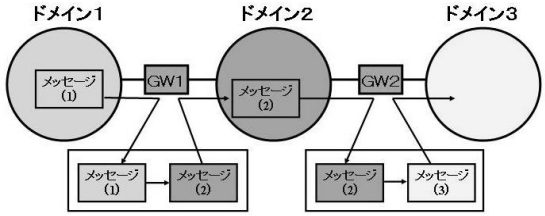


図 4: オントロジー変換方式 3

本方式は，メッセージのトンネリングを行わず常に隣接するドメインのオントロジーへと変換する方式である．図 4 において，GW1 はドメイン 1 内で送出されたメッセージを受信し，ドメイン 2 のオントロジーへと変換してドメイン 2 へ送出する．GW2 も同様にドメイン 2 内で送出されたメッセージを受信し，ドメイン 3 のオントロジーへと変換してドメイン 3 へ送出する．

2.1.4 オントロジー変換方式の比較

上記の 3 つのオントロジー変換方式に関して，3 つの項目（A〜C）から比較を行ったものを表 1 に示した．ここで，(A）の「オントロジー対応情報」とは，ある特定の 2 つのドメイン間のオントロジー変換に必要な情報（変換表）であると定義する（具体的なデータ構造に関しては後述する）．

表 1: オントロジー変換方式の比較表

	(A)各GWが保持するオントロジー対応情報の数	(B)各GWで，特定のメッセージに対して，オントロジー変換を行う回数	(C)複数のドメインを跨ってメッセージが転送される際に，当該メッセージに対してオントロジー変換が行われる回数
オントロジー変換方式1	〜2 (○)	〜2 (○)	〜2 (△)
オントロジー変換方式2	〜(ドメインの数 − 1) (×)	〜1 (○)	〜1 (○)
オントロジー変換方式3	〜1 (○)	〜1 (○)	〜(通過したGWの数) (×)

方式 1 は（A）（B）（C）のどの項目に関しても問題が無く，一番処理効率が優れている．しかし，本方式は標準オントロジーを確立しなければならないという問題点を有している．複数のオントロジーを統合するような標準オントロジーを作成することは困難であり，グローバルかつドメイン数が膨大である環境においては，標準オントロジーに関する合意を得ることは現実的には厳しいと考えられている．

方式 2 は（B）（C）の項目に関しては問題ないが，（A）の項目に問題がある．すなわち，ドメインの数が増加した場合，それに伴って各 GW が保持しなければならないオントロジー対応情報の数が比例的に増加する．よって，ドメインの数が膨大である場合は，本方式を用いることがやはり困難である．

方式 3 は（A）（B）の項目に関しては問題ないが，（C）の項目に問題がある．つまり，メッセージが複数のドメインを跨って転送されていく際に，通過するドメイン（GW）の数に比例して，本メッセージに対してオントロジー変換が行われる回数が増加する．そのため，オントロジー変換の処理効率が悪いばかりか，オントロジー変換の際に発生した誤差が蓄積されてしまうという問題点を有している．

以上の通り，各方式にはメリットとデメリットが存在し，適用領域に応じて方式を選択する必要がある．なお，上記の方式は併用することも可能であり，例えば局所的なドメイン群における標準オントロジーを確立して方式 1 を採用すると同時に，そのようなドメイン群間では方式 2 を採用する形態も可能である．

3. オントロジーモデル

CBN において，採用するオントロジーモデルの決定はその全体設計に対して多大な影響を及ぼす．CBN を利用するサービスアプリケーションの立場からは，より表現力（expressiveness）が高く柔軟性のあるオントロジーモデルが望ましい．しかし，例えばルーティングは本質的に分散アルゴリズムであり，処理効率や実装の容易性の面からは，表現力が低くともシンプルなオントロジーモデルが望ましい．一般に処理効率やスケーラビリティとオントロジーモデルの表現力はトレードオフの関係にあると言える．

具体的なオントロジーモデルの例として，channel, subject といった数値や文字列を基本とした比較的表現力の低いものから，フレーム形式のような構造化されたデータモデルを有するもの，オブジェクト指向モデルのような詳細な関係定義を可能としたもの，あるいは一階述語論理式を用いることによって公理定義を可能としたもの等が挙げられる．

CBN の基本機能は，メッセージの配送先を動的に決定するという検索機能にあり，検索という利用

側面からメタ情報が構造化されているメリットは大きい．一方で，前述の通り表現力に富んだオントロジーモデルは，どちらかと言えば集中サーバ向けであり，処理効率やスケーラビリティに問題がある．よって，CBN のオントロジーモデルとしては，構造化されつつも適度にシンプルなものが適していると考える．

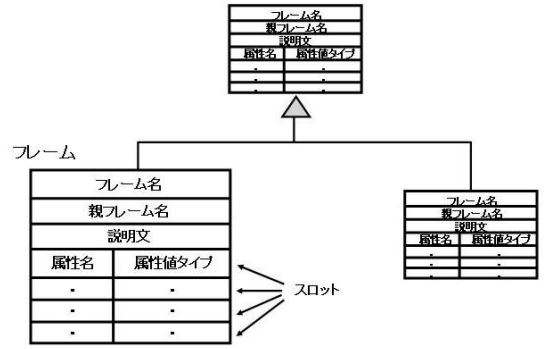


図 5: オントロジーモデル

以上の考察に基づき，本稿では図 5 に示すフレーム形式を用いたオントロジーモデルを採用する．すなわち，オントロジーはフレームの集合から構成され，フレーム間は継承関係を有している．フレームはフレーム名，親フレーム名，フレームの説明文（annotation）と複数のスロット（属性名と属性値タイプのペア）から構成される．そして，フレームのインスタンスがメッセージのメタ情報となる．

なお，オントロジーモデルがこのような階層構造を有していると，例えば階層的にマルチキャストパス（spanning-tree）を張るといったルーティング方式の実現も容易となる．

4. オントロジー自動変換方法

第 3. 節で定義したオントロジーモデルに限定しても，なお概念化（conceptualization）の違いに基づく構造化の仕方や語彙の選択における差異が発生する．提案するオントロジー自動変換方法は，このような差異を解消することを目的とするものである．そして，本提案方法は基本的には GW において利用されることを前提としている．すなわち，GW においてメッセージ内のメタ情報のオントロジー変換を行うことによって，ドメインを跨っても正しいメッセージ配送が実現できることを目指すものである§．

本提案方法は「(1) オントロジーの取得」,「(2) オントロジー対応情報の自動生成」,「(3) メタ情報のオントロジー変換」という，大きく分けて 3 つのステップから構成される．つまり，まず変換元オントロジーと変換先オントロジーを各ドメインから取得し，オントロジー対応情報を自動生成する．ここで，オントロジー対応情報とは，具体的には各オントロジーを構成しているフレームおよびスロットに関する対応関係を記述した「フレーム対応」および「スロット対応」の集合のことである．そして，そのオントロジー対応情報を用いてメタ情報のオントロジー変換を実現する．

本提案方法の特徴として，一度（1）と（2）のステップを経ることによってオントロジー対応情報を生成しておけば，そのオントロジー対応情報を用いてステップ（3）のみによって複数のメタ情報のオントロジー変換が実現できることが挙げられる．すなわち（1）と（2）のステップに関しては，最初に一度実行しておけば，あとは変換元オントロジーあるいは変換先オントロジーに変更があった場合にのみ実行すれば良い．

なお，本提案方法は第 2. 節で説明したどの方式に対しても適用することが可能である．

以下，各ステップの詳細を説明する．

(1) オントロジーの取得

始めに変換元（source）オントロジーと，変換先（target）オントロジーを各ドメインから取得する．オントロジーとは，具体的にはフレームの集合である．

(2) オントロジー対応情報の自動生成

本ステップは，大きく分けて「(2.1) スロットのマッピング」と「(2.2) フレームのマッピング」から構成される．

(2.1) スロットのマッピング

スロットのマッピングは，変換元スロットと変換先スロット（スロット対）の類似度を定義することにより，トータルとしての類似度が最も高くなる組合わせを求めるという「組合せ最適化問題」に帰着できるが，ここではより簡明なアルゴリズムを用いる．

(2.1.1) スロット対類似度の定義

スロット対類似度を以下の式によって定義する．

$$\text{「スロット対類似度」} = \min(\text{「属性名の類似度」}, \text{「属性値タイプの類似度」}). \quad (1)$$

まず「属性名の類似度」に関しては既存のシソーラスを用いて算出することも可能だが，本稿では国語辞書と英和辞書から構築した概念ベース [4] を用

§メッセージ内のデータに関しては，サービスアプリケーションが規定するもの（CBN の上位レイヤの問題）であり，オントロジー変換は適用しない．

いて算出する．すわなち，各属性名の概念ベクトルを $\boldsymbol{V_s}$,$\boldsymbol{V_t}$ とすると，

$$\text{「属性名の類似度」} = \frac{\boldsymbol{V_s} \cdot \boldsymbol{V_t}}{|\boldsymbol{V_s}||\boldsymbol{V_t}|}. \qquad (2)$$

となる．なお，属性名が複合語である場合は，まず最長一致法，あるいは形態素解析プログラム等によって複合語をその構成語に分割する．そして，上記 $\boldsymbol{V_s}$（あるいは $\boldsymbol{V_t}$）の代わりとして，構成語の概念ベクトルの加重平均を用いる．

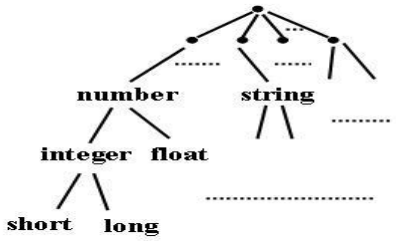


図 6: 属性値タイプのシソーラスツリー

「属性値タイプの類似度」に関しては，図 6 に示すような属性値タイプ（string，integer 等) の階層関係に基づいたシソーラスツリーを構築しておき，以下の式によって算出する．

$$\text{「属性値タイプの類似度」} = \frac{2d_c}{d_s + d_t}. \qquad (3)$$

ここで，d_s と d_t は各属性値タイプのツリーにおける深さであり，d_c は共通上位ノードの深さである．

(2.1.2）スロット対応の決定

以下（［1］～［2］）のアルゴリズムに従ってスロット対応を決定する．

［1］スロット対の類似度行列 S を作成する．ここで行列要素 S_{ij} は，i 番目の変換元スロット slot_i と j 番目の変換先スロット slot_j の類似度を式（1）を用いて算出した値をとる．

［2］スロット対応のペアが作れなくなるまで，以下を繰り返す．

［2.1］S_{ij} が最大値 $s_{\max}$ をとるときの i と j を求める．そして，

if （ $s_{\max} > \theta$(threshold） ） {

　スロット対応（$\text{slot}_i \rightarrow \text{slot}_j$）をオントロジー対応情報として追加する．

} else {

　ループを終了する．

}

［2.2］上記 i 行と j 列を削除する．

(2.2）フレームのマッピング

フレームのマッピングに関しても，基本的にスロットのマッピングと同様の方法を用いる．

(2.2.1）フレーム対類似度の定義

フレーム対類似度を以下の式によって定義する．

$$\text{「フレーム対類似度」} = \min(\text{「フレーム名の類似度」},\text{「説明文の類似度」},\text{「スロット対集合の類似度」}). \qquad (4)$$

ここで「フレーム名の類似度」は「属性名の類似度」と同様に概念ベースを用いて算出する．また「説明文の類似度」に関しても，各説明文中の自立語に対応する概念ベクトルの加重平均を計算することにより，式（2）を用いて算出する．そして「スロット対集合の類似度」は，ステップ（2.1）において決定されたスロット対集合を用いて，以下の式によって算出する．

$$\text{「スロット対集合の類似度」} = \sum \text{スロット対類似度} \,/\, \text{スロットの数}. \qquad (5)$$

この式は，対応するフレームは同様のスロットを保持しているという考えに基づいている．

(2.2.2）フレーム対応の決定

フレーム対応の決定はステップ（2.1.2）のスロット対応の決定と同様である．すなわち式（4）を用いてフレーム対の類似度行列 F を作成し，同様のアルゴリズムによってフレーム対応を決定する．

(3）メタ情報のオントロジー変換

メタ情報のオントロジー変換は，ステップ（2）において算出されたオントロジー対応情報を用いて行う（図 7 参照）．

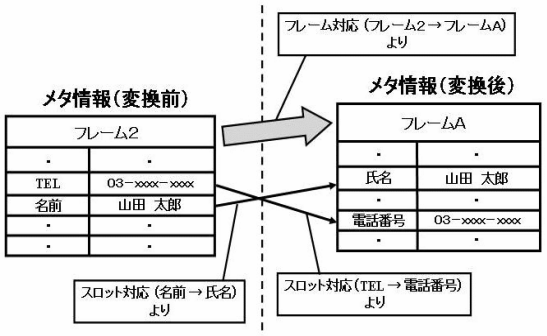


図 7: メタ情報のオントロジー変換例

すなわち，オントロジー対応情報に含まれているフレーム対応を参照することで，変換元メタ情報のフレームに対応する変換先フレームを用意し，さらにスロット対応を参照することで対応する変換

先スロットの属性値（スロット値）を設定する．なお，属性値に関しては，属性値タイプとして列挙（enumeration）型が指定されていた場合においてのみ，その変換を行う．また，ステップ（2）において算出された類似度をメッセージに付与しておくことで，本類似度をオントロジー変換の精度を示す指標として，例えば検索結果の順位付け等に活用することが可能である．

5. 実装

本提案システムを Linux 上で Java を用いて実装した．メタ情報の記述言語には XML [5] を用い，フレームの記述言語には RELAX-NG [6] を用いた（フレーム記述の一例を図 8 に示した）．記述言語に XML 技術を採用したのは，独自仕様を採用する場合に比べてユーザの習得負担の軽減化が期待でき，各種市販ツールの活用も可能となるためである．

```
<?xml version="1.0" encoding="EUC-JP" ?>
<element name="社員"
  xmlns="http://relaxng.org/ns/structure/1.0"
  datatypeLibrary="http://www.w3.org/2001/XMLSchema-
  datatypes">
  <element name="氏名">
      <data type="string"/>
  </element>
  <element name="年齢">
      <data type="integer"/>
  </element>
  <element name="住所">
      <data type="string"/>
  </element>
  <element name="電話番号">
      <data type="string"/>
  </element>
</element>
```

図 8: RELAX-NG を用いたフレーム記述の一例

なお，オントロジー対応情報に関しては現在独自仕様を用いているが，XSLT スクリプト [7] へのコンバータを用意することによって，第 4. 節のステップ（3）は市販の XSLT エンジンを用いて実現することが可能である．

6. 関連研究

ここでは，本提案システムと関連研究の比較を行う．

EBFM [8] は，フレーム形式において同じ属性を表現しているスロットを自動的に対応付ける事例ベースのフレームマッピング方法であり，基本的に本提案システムと同様の問題を扱っている．しかし，EBFM は異なる情報源において共通の対象を表現した事例を同定し，スロット値の類似をもとにマッピングを行うものであり，シソーラス等の言語リソースを用いることで，スロット値を利用することなくマッピングを行っている本提案システムとは，そのアプローチを異にしている．

XML のボキャブラリ（スキーマ）変換技術（例えば [9]）も，広義には本提案システムと同様の問題を扱っている．しかし，XML は文書とデータの混在を許す柔軟性に富むデータモデルであり，その適用領域も広い．そのため，汎用的にさまざまな適用領域（例えば商取引のトランザクション等）における変換を自動化することは容易でなく，これらの技術は人間による変換作業のサポートに主眼を置いている．それに対して本提案システムは，データ中心のデータモデルに対象を限定し，CBN という検索レイヤのみに適用領域を限定することで，自動化の実現可能性を追求するものである．

7. 結論

本稿では，CBN におけるオントロジー変換アーキテクチャを示し，3 つのオントロジー変換方式に関する比較を行った．そして，CBN におけるオントロジーモデルに関する考察を行い，オントロジー自動変換方法の提案を行った．今後，提案したアルゴリズムの評価と改良を行う予定である．

参考文献

[1] 星合 隆成，小柳 恵一，ビルゲ スクバタール，久保田 稔，柴田 弘，酒井 隆道：意味情報ネットワークアーキテクチャ，電子情報通信学会論文誌 B Vol.J84-B No.3 pp.411-424（2001 年）．

[2] Carzaniga, A., Rosenblum, D.S., and Wolf, A.L.: Design and Evaluation of a Wide-Area Event Notification Service, *ACM Transactions on Computer Systems* 19(3):332-383 (2001).

[3] 溝口 理一郎：オントロジー研究の基礎と応用，人工知能学会誌 Vol.14 No.6（1999 年）．

[4] 笠原 要，松澤 和光，石川 勉：国語辞書を利用した日常語の類似性判別，情報処理学会論文誌 Vol.38 No.7（1997 年）．

[5] http://www.w3.org/XML/

[6] http://www.oasis-open.org/committees/relax-ng/

[7] http://www.w3.org/TR/xslt

[8] 伊藤 史朗，上田 隆也，池田 裕治：分散情報源に対する情報エージェントのための事例に基づくフレームマッピング，電子情報通信学会論文誌 D-I Vol.J81-D-I No.5 pp.433-442（1998 年）．

[9] TIBCO XML Transform, http://www.tibco.com/solutions/products/extensibility/xml_transform.jsp