

## Context 表現のための手法の考察

山下 祐平<sup>†</sup> 木下 宏揚<sup>†</sup> 森住 哲也<sup>††</sup>

<sup>†</sup> 神奈川大学工学部, 神奈川大学ハイテクリサーチセンター  
221-8686 神奈川県横浜市神奈川区六角橋 3-27-1  
<sup>††</sup> ネットズエスアイ東洋株式会社  
253-0198 神奈川県高座郡寒川町小谷 2-1-1

E-mail: †{yamayu,kino}@cs.ee.kanagawa-u.ac.jp, ††moriz@olive.ocn.ne.jp

あらまし 情報で溢れている現代社会において、Web システムなどで扱う情報が膨大になり、情報検索や知識生成などの処理は限界を超えようとしている。そこでセマンティックウェブなどの意味情報に基づいたシステムが注目を浴びている。しかし、意味情報を探索する方法では今までと比較にならないくらいの処理量になる。これらの解消のために意味グラフを Context に分割する方法が提案されてきた。しかし、この方法には、適切な分割を行わないと重要な検索経路が分断されてしまい、分割していないときよりも検索の精度の悪いものになってしまう。したがって、Context に分割する際には適切に分割することが非常に重要になる。本稿では、実例を基に Context の特性について考察し、適切な分割のための指針を提案する。

キーワード Context, Ontology, トピックマップ, 非文字資料

## Study of method for expressing context

Yuhei YAMASHITA<sup>†</sup>, Hirotsugu KINOSHITA<sup>†</sup>, and Tetsuya MORIZUMI<sup>††</sup>

<sup>†</sup> Faculty of Engineering, Kanagawa University  
Rokkakubashi, Kanagawa-ku, Yokohama 221-8686, Japan  
<sup>††</sup> TOYO NETWORKS & SYSTEM INTEGRATION CO.,LTD.  
Kotani, Samukawa-machi, Koza-gun 2-1-1, Japan

E-mail: †{yamayu,kino}@cs.ee.kanagawa-u.ac.jp, ††moriz@olive.ocn.ne.jp

**Abstract** In the contemporary society that overflows with information, it becomes difficult to treat the flood of information. Then, the system based on the semantic information such kind of the semantic web is remarkable. However, the method of searching for the semantic information is needed incomparably more processing power than keywords based system. The method of the division of contexts has proposed for being solved these problems. But, inapropriate division causes deterioration of the precision of the retrieval because of the division of the search path. Therefore, when dividing of contexts, appropriately dividing becomes very important. In this text, the characteristic of contexts is considered, and it proposes for appropriate method for division context based on an example.

**Key words** Context, Ontology, Topic Maps, Nonwritten Cultural Materials

## 1. はじめに

近年の情報社会において、ウェブ上に存在する情報は膨大となり、その情報量はより急速に増え続けている。現在、このような状況において情報を検索する方法として、一般に普及しているのはキーワードに基づいたテキストベースの方法である。しかしながら、キーワードに基づいたテキストの情報検索では、同じキーワードでもユーザが所望する意図がそれぞれ異なり、ユーザが満足するものを提供することができない。

これを解決するために意味情報に基づいた情報ネットワークの構築に関する研究が盛んに行われている。

意味情報に基づいた情報ネットワークは、意味を考慮した検索であるから、その精度は疑いの余地がない。しかしながら、正確な意味情報ネットワークを実現するための構築コスト、グラフ探索であるが故の探索のコストなどの難問もあり、一般には普及していない。

以上の問題を改善するためには、意味ネットワークの構築コストと探索コストの削減が不可欠となる。そこで、これらを実行するために Context という概念を定義し、適切な意味ネットワークへ分割し、分割した Context を弱い紐帯で結ぶといった手法が有効であると考えられる。この手法は、Context を分断してはいけないつながりまで分けてしまうと探索そのものができなくなるなど問題もある。しかしながら、所望する情報の意味同士が厳密につながっているとわかっているときには、Context に基づいた比較的狭い範囲を探索すればよいと考えられる。また、Context 間の意味関係を定義してあげると、弱い紐帯の役割を果たし、連想のような検索も行う事ができると考えられる。

著者は、21 世紀 COE プログラム「人類文化研究のための非文字資料の体系化」[1]の一環として、民俗文化を知的文化遺産とし、研究成果を電子化、情報発信することを目指している。Dublin Core [2] を参考にした非文字資料に適したメタデータの提案を最初に、意味情報を加味した非文字資料のデジタルアーカイブ化に関連する研究 [3] [4] [4] を行ってきた。その中で、Context の記述にはトピックマップが適しているとしてきた。[6]

そこで本稿では、トピックマップによる Context の記述を押し進め、意味情報ネットワークを構築する際に Context の適切な分割をどのように行えばよいのか、Context の特性を考察し実際にデータ構造を与えてみることでどのような利用法があり利点があるのかを示した。

また、弱い紐帯を利用した連想のような検索を可能とするには、Context 間の結びつきをどのようにモデル化するかを考察し提案した。

本稿のアウトラインは次のようになる。2 章は、意味に基づいた情報ネットワークの現状を説明した後にキーワードベースの情報検索システムでは起こらなかった問題点を示し、改善すべき課題を示す。

3 章は、Context についての考察を行い適切な Context の分割のための指針を提案する。

4 章は、提案した方法を適用した場合の利点を示す。

5 章は今後の課題を示し、6 章にまとめを行う。

## 2. 意味に基づいた情報ネットワーク

現在主流になってくる検索の方式は、知りたいことについてキーワードで入力する方式である。この方式では、基本的にパターンマッチングとブーリアンモデルの組み合わせを取っており、大量の情報に対して処理が必要な検索システムにおいて計算量が急激に増えないという利点がある。キーワード方式では、ブーリアンによる組み合わせによって知りたいことを限定できたとしても意味関係までは考慮されない。したがって、仮にブーリアンモデルで知りたいことを表せたとしても結果にごみが入る可能性も十分にある。

既存の情報システムではその性質上解決できないこれらの問題を根本から解決するために、意味に基づいた情報ネットワークの構築に関する研究が行われている。その中でも特に注目すべきなのが、Semantic Web と呼ばれる既存の Web に意味情報を統合したものである。

Semantic Web では、情報リソースを明確記述する RDF を用いて既存のデータにメタデータを付与する。これを行うことで、既存の対象データがどのような意味をもっているかが明示的に示される。このメタデータを利用して情報を扱うにはこれらのメタデータがどのような意味をもつ概念であるのか記述されていなければならない。それを行うのが RDFS や、OWL といった言語である。RDFS は RDF のスキーマ言語であり RDF のクラスを記述し、上位下位概念などの基本的な関係を記述することができる。OWL は RDF よりもさらに高度な関係を記述でき、さらに細かな制約条件なども記述できる Ontology 記述言語となっている。RDFS/OWL によって概念関係を記述することにより、実体データと概念関係の

これらの仕様は W3C により勧告されている。また、それらの応用である SPARQL などのクエリ言語なども最近に勧告となり Semantic Web の環境は次第に整備されつつある。

Topic Maps は ISO によって標準化されている RDF に似た知識表現の枠組みである。しかし、情報を見つけることを重視しているという特徴がある。Topic Maps についても、情報にアクセスするための API が存在するなど環境の整備は進んでいる。

これらの意味に基づいた情報ネットワークはその性質上、意味関係の異なった情報を引き出してしまふ可能性が少ない。そのため、既存の情報システムとは異なり、最適なシステムの構築の補助や情報の活用方法を考えるなどの人間が手動で行わなければならない動作をコンピュータに任せることが期待できる。

### 2.1 Ontology

Ontology は、概念と概念間の関係について記述するものである。ここでいう概念とは、メタデータの意味構造を定義するためのものである。具体的には実在する人物、動植物、さらには実在しないフィクションのものまで百科事典に載るような項目はすべて定義可能である。ただし、オントロジで世界全

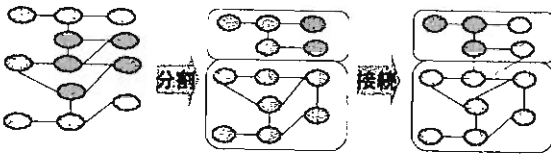


図1 ドメイン Ontology の Context による分割と接続

体を記述することは不可能に近いので、各々のドメインでの Ontology を構築しそれを相互利用するという方式が進められている。Semantic Web においては、RDF など記述されたメタデータの Subject, Object, Predicate がそれぞれどのような意味を持つものであるかを定義するを定義する。

一般的には、is-a (上位/下位) や part-of (全体/部分) といった概念間の関係が定義されているが、記述可能な関係などは特に限定されているわけではない。例えば、「一郎さんが鉄を所有している」という文を RDF で記述する場合には、「人間という型を持った一郎」と「道具という型を持った鉄」が「所有している」という述語 (プロパティ) によって接続されることになる。この「人間」や「道具」といった型や、「所有している」という述語がどのような概念であってそれらがどのような関係にあるかを記述するのが Ontology である。

Ontology を使って定義された意味構造は、意味情報ネットワークシステムによって参照され、記述された意味を少しでも利用する場合に利用される。

既存の情報システムに対する Ontology によって概念を記述することのメリットは、意味情報をシステムから分離できることによる他のシステムの情報がシームレスに利用可能になること、厳密な概念に基づいた情報探索や推論が可能になることが期待できる。その一方で、Ontology 構築の必要条件となる厳密な概念の記述が容易なことではなく、さらに概念の変化にも対応させなければならないなど構築コスト面での問題が大きい。

## 2.2 意味情報に基づく検索の問題点

意味情報に基づく検索が理想的であるにもかかわらず既存の情報検索が普及した理由として、データ自体を構造化することなく単純な処理に抑えることによる応答性能の高さがある。一方、意味情報に基づく検索はパターンマッチングではなくグラフ探索という計算量が大いものとなってしまう。

それらの問題を解消するために、図1にあるように Context によって分割する方法が提案された。これは、意味情報が Context に依存しているという特長を活かし、探索範囲を Context 内部にとどめることで計算量が大幅に削減されるためである。この方法では、クエリに対してどの Context を探索範囲とするかについての情報が必要となる。そのために、Context の違いに関する情報を持った Context 間を接続する特殊なひもが必要になる。これは、主に先にも述べた共通語彙の接続にて行われる。

## 2.3 トピックマップ

RDF と RDFS/OWL などと同様に意味ネットワークを記述するものにトピックマップがある。図2はトピックマップはトピックマップの概略を示したものであるトピックマップは ISO

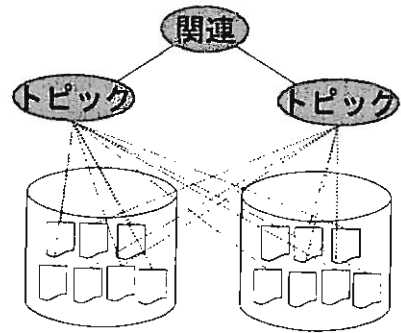


図2 Topics map の概観図

によって標準化されているという強みがある。データにトピックという索引をつけそれらの関係性を記述することで情報へのアクセスを容易にするという目的から生まれた。

意味理論上は厳密な定義が可能である RDF には、記述が冗長、RDF トリプルの表現力が乏しい、リソースの具体化に関する記述が曖昧であるなどの問題があった。記述が冗長である点はフォーマットに依存する関係でありあまり変わらないが、トピックマップではトピック同士の関係性を役割を含めて記述するための基本の枠組みが定められており、表現力の問題や、リソースの具体化に関する記述の問題などが軽減される。

Ontology はトピックを用いて記述することも可能であり、実際にトピックマップでオントロジーを記述し、インスタンスを登録するといった方式のトピックマップのツールも存在する。2つ以上の関係性も記述可能である設計になっているので、複数の概念間の関係を記述するのに適しているといえる。また、オントロジー記述言語ほど厳密な概念関係でなくてもよいので比較的容易となる。これは、多様性のある Context の記述はなるべく多くの人が参加する必要があるという要求を満たす。以上から、現状ではオントロジーと Context 部分についてトピックマップを使ってそれぞれについて記述し、システムを構築することが望ましいといえる。

トピックマップにはデファクトスタンダードの API が存在するので、それらを利用した Web アプリケーションの構築が可能となっている。また、ISO によるスキーマなどの標準化も進められている。

## 3. Context

Context は、直訳すると文脈といわれるように前後関係や背景、暗黙の前提条件などといった意味合いのものである。人工知能分野においては、意思伝達、言語学、形而上学などに属する部分と深い関係がある。

人間は、このような暗黙の了解としての Context を考慮することが得意である。この Context について両者が合意していないとコミュニケーションもままならないことになる。

一方でコンピュータは、苦手である。例えば、検索を行う際に共通語彙が異なる概念として現れる場合も多いが、それは属する Context が違うためといえる。この場合、Context を考慮

せずに処理を行うと見当違いの情報を返してしまう。

Context が分割されていれば、探索の際に行われる計算も少なくて済む。それだけでなく、Context 間の関係についてメタデータを記述しておけば探索経路の優先度の決定に活用したり、Context は違っても意味構造がほぼ一致している意味グラフから推論を行い仮説を立てることなどが可能である。しかし、不適切な Context を与えると意味的に切り離してはいけない部分を分けてしまったり、本来探索が不要である Context を探索することが起こりうる。

このことから Context を適切に設定することは本方式においてきわめて重要な問題となる。ここでは、Context の主な種類、抽象度などについて考察

### 3.1 主な Context

Context は様々な前後関係、背景について変化したものであるのでいくつかの種類に分類される。Context の種類について以下に示す。

#### a) 意味グラフがもつ Context

意味情報ネットワークがもつノードが成立する前提条件を表す。この種類の Context は、本稿で主に述べている Context の分割に関係している。Ontology を構築するには、意味グラフが成立する複数の文について結合し記述して

#### b) クエリがもつ Context

クエリの意味グラフに直接含まれていない場合でも、知りた知識には前提条件がある。

#### c) ユーザの自身が持つ Context

ユーザ自身が持つ情報である。ユーザの言語、年齢、性別、所属などのユーザ自身の持つ属性などである。例えば、ある言語のことわざを他の言語に直訳した場合には、意味がまったく伝わらないことがあることなどがあげられる。

本稿ではユーザ自身が持つ Context については考えず、意味グラフが持つ Context とクエリがもつ Context について考察し、その利用例を挙げている。

### 3.2 Context の関係性

Context は多く定義できる方がよいが、関係性が複雑になると計算量が増大する。そうなると Context に分割した当初の目的である探索の軽量化ができない。

#### 3.2.1 Context の軸と抽象度

制約を加えないと Context は無数に定義できて混沌としてしまい、Context を用いていない状況と変わらなくなってしまう。そこで、図 3 に示すようにドメイン Ontology を多次元の軸に分離し、推移律が成立する Context 同士を従属関係にしたがって並べていくことでそれぞれの Context の抽象度を定義する。また、それらの組み合わせで任意の Context を表現することにより、それぞれの軸での特徴から Context の関係を導くことが可能となる。

例えば、「江戸時代の食文化」といった Context があつたとする。その場合、安土桃山時代後期や、明治時代初期は江戸時代隣接した時代であることから、所望する情報を含むかもしれない。また逆に、江戸時代の中でも特に限定した慶長、寛永などの年代間でも食文化との関係が濃い年代と薄い年代がわかるか

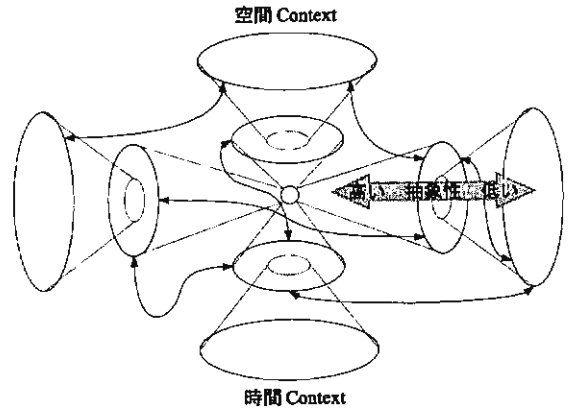


図 3 Context の軸と抽象度

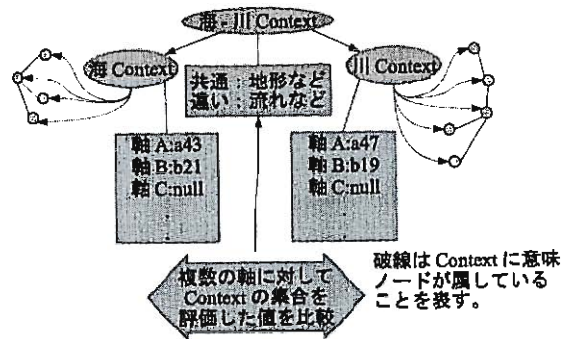


図 4 提案する Context 構築のためのデータ構造

もしれない。このように指定した Context の抽象度がクエリの Context に最適でない可能性を考慮して、抽象度を高くしたり低くしたりすることを可能としておくが必要になる。

軸と抽象度を考慮したモデルでは推移律が成り立つ物を想定しているので推論を行うときには、特に効果が現れる。

#### 3.2.2 Context の組み合わせ

図 3 の各軸に属する Context においてそれらを組み合わせることができれば、任意の Context を作るのと同様になる。ここでは、Context の組み合わせ方について簡単に述べる。

まず、組み合わせの方法は既存のモデルと同様の論理和、論理積などを使った組み合わせに限定する。これは、集合の側面をもつ Context の特性を生かしたものである。ここでなぜ意味関係を使用しないかという Context 自体がメタ的な概念であるから意味関係を考慮するよりも処理量の削減を重視すべきという理由からである。

Context を任意に定義した場合と同じようにさまざまな Context を定義可能だが、Context が増えていくことによる混乱は、Context の軸の数が Context の数に比較してかなり少ないことによって抑えられる。

### 3.3 Context のデータ構造

図 4 は、今まで述べてきた Context の特性を元にトピックマップによる記述の例である。

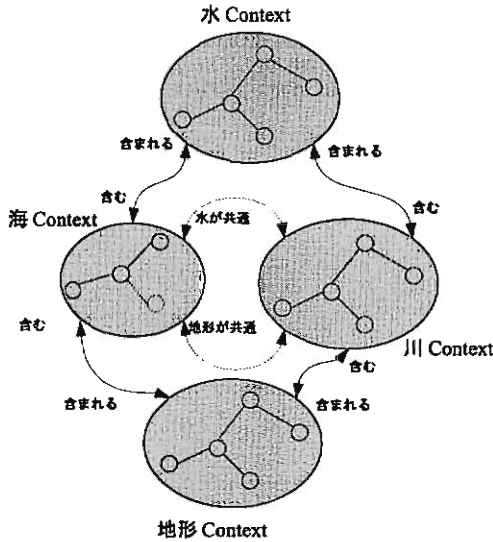


図 5 Context 間の類似度

まず、海に關係している Context と、川に關係している Context をそれぞれ Context のノードとして設定する。それらの Context が、図 3 の階層化された軸のどの Context を組み合わせて作成されているのかを調べる。組み合わせ元の Context を Context ノードの内部出現のリストとして保持する。これにより Context 自体は表現されたことになる。

次に、作成した Context の特性リストからある軸について一致している場合や、逆の特性を指している場合など特筆すべき關係を持っている場合のみ、Context ノードを接続し、各軸における重要な Context 關係を内部出現として保持する。

このように Context 間の關係を示すノードに記述しておけば、探索や推論などでうまくいかなかった Context を置き換えるときに Context 間の關係を表すノードを参照するだけでよくなる。Context の關係が既知の場合には、未知の場合に対して計算量は大きく削減されて探索の精度も向上するというメリットがあり、Context のノードを作成して軸に属している Context の組み合わせを保持することは割に合うといえる。

### 3.3.1 Context の類似度

軸に分離された Context の組み合わせによって作られた Context 同士の關係は、Context の軸の概念距離から導出することが可能である。特に、類似した性質を持っている Context については、連想検索などに利用することが可能である。

さらに、未知の知識に関する仮説を類似する Context から推論し提供することができる。図 5 は、Context 同士の關係が存在するときにそれがどのような構造として現れるかを示したものである。図では、海 Context と川 Context は水を含んでいるものでも地形であることを示している。それらの關係は Context の軸と抽象度で説明したように推移可能な Context 同士の概念距離やその共通祖先によって定められる。ここでは、海 Context と川 Context がともに共通の祖先である水 Context

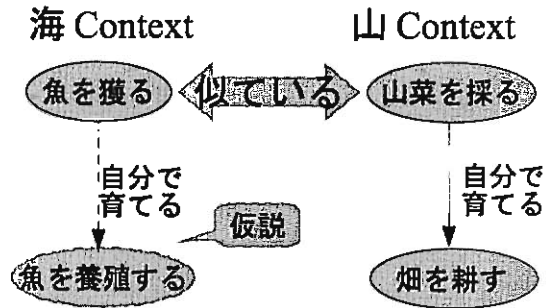


図 6 類似した Context から仮説を立てる例

および地形 Context をもつことから海 Context と川 Context には共通属性があるという Context 間の關係を示している。

図 6 は、類似する Context の關係から仮説のノードを作成する例である。海 Context と山 Context は、同じ地形という概念であるから地形 Context の抽象度を高くした場合に Context の概念距離が近いことがわかる。その中で、海 Context の「魚をとる」と山 Context の「山菜をとる」というグラフパターンが近いことがわかる。

このような前提の中で「安定して魚をとる」ということを行いたくてもその方法がわからないときに、魚をとるというグラフが属している Context と概念が近い Context を探索し、山 Context の「山菜をとる」というパターンを発見する。このとき、「安定して山菜をとる」には「畑を耕す」こと、つまり野菜を育てることがよいことがわかっているとすると、海 Context でもそれを適用して魚を自分で育てるという「魚を養殖する」といふ仮説が得られる。

このように、Context 同士の關係を軸に分けてそれぞれの概念距離を与えることができれば、今までになかった知識である「魚を養殖する」といった仮説のノードを作成することができる。これを軸に分類しないで行おうとすると、火星 Context 等の關係がない Context の意味グラフパターンから仮説を立ててしまうなどの不都合が生じる。

こうした適用例からも Context を軸に分離し、各軸に属した Context を表現する方法が優れていることがわかる。

## 4. Context の検索利用例

ここまで示してきた Context の分割を活用した検索時の優先探索経路の決定アルゴリズムを図 7 に示す。

図 7 は、意味グラフのパターンマッチングを行う際に着目しているカレントノードの優先度の変化を示している。

まず、ユーザが生成したクエリの意味グラフからその Context の特性を各軸についてたどり、Context の特性リストとする。これを Context ノードが保持している Context の属性と比較する。その際、優先度の高い軸（図では軸 A）から順番に比較して概念距離が近い Context を重要な検索対象であると考えられる。したがって、クエリから、は少なくとも軸と軸の優先順位を導かなければならない。しかし、それは、意味ネットワークグラフと同じものを使うことができる野でそれほど問題にはなら

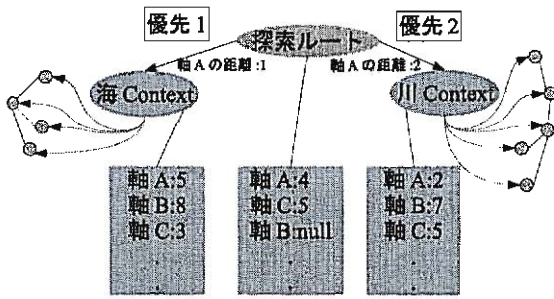


図7 Contextの検索経路の優先度

- [5] 木下慶子, 村上敦志, 稲積泰宏, 木下宏揚, 森住哲也  
デジタルアーカイブにおける Ontology の活用
- [6] 山下祐平, 木下宏揚, 森住哲也  
トビックマップを用いた非文字資料における Context の表現

ない。

意味に基づいた情報検索では、次に探索するノードの順番が非常に重要になる。例えば、探索中のノードの隣接ノードに解があるにもかかわらず膨大な隣接ノードをもつ別のノードを先に調べてしまい、それらのノードを探索し終えてから探索することでは探索コストが非常に異なる。このことから、曖昧な概念集合としての Context の探索の優先順序を簡単に導出できる提案方式のデータ構造は、探索の応答性をあげるのに優れているといえる。

## 5. まとめ

本稿では、意味に基づいた検索を行う際に有益であると思われる Context について考察してきた。これまで主として意味グラフのパターンマッチングにおける探索の計算量を削減することを主張してきたのに対して、本稿では Context の特性を考察し、既存の情報システムではなく意味情報に基づいた情報ネットワークに独自のグラフ探索や推論などに活用できる適切な Context 分割を行う方法について具体例を交えて示した。

まず、Context の適切な分割を目的とした Context の概念の軸と抽象度の導入である。Context の分類を推移律が成り立つ軸にたいして抽象度を変化させることができる Context の組み合わせで記述することで、Context 間の関係性がより明確になり、より効率的な検索や推論を行うことができる。

Context に軸と抽象度という仕組みを取り入れることで、さまざまな問題が解決されると考えられるが、Context の概念階層の設計には、Ontology の構築と同じように厳密な意味論理を損なわないように設計しなくてはならない。今後の課題は、現在ある Top Level Ontology などを参考にし、これらの Context の概念階層を構築することである。

謝辞 この研究は平成 20 年度科学研究費補助金 [基盤研究 (C)](17500048) の援助を受けた事に感謝の意を表する。

## 文 献

- [1] 神奈川大学 21 世紀 COE プログラム 人類文化研究のための非文字資料の体系化  
<http://www.himoji.jp/>
- [2] Dublin Core Metadata Initiative.,  
<http://www.dublincore.org/>
- [3] 木下慶子, 稲積泰宏, 木下宏揚, 森住哲也  
COE における非文字資料の共有と流通
- [4] 木下慶子, 稲積泰宏, 木下宏揚, 森住哲也  
非文字資料に適した Ontology 構築