

平成21年度 修士論文

論文題目

意味情報ネットワーク構築に適した
Contextのモデル化

神奈川大学 工学部 電気電子情報工学科

学籍番号 200770083

山下 祐平

指導担当者 木下宏揚 教授

目次

第1章	序論	4
第2章	基礎知識	8
2.1	非文字資料	8
2.1.1	只見町民俗資料調査カード	9
2.1.2	民具の分類	13
2.2	意味論とその隣接分野	15
2.3	Ontology	16
2.3.1	Top level Ontology	18
2.4	Topics map	18
2.4.1	Topics map の主要な構成要素	21
2.4.2	Topics map 利用のメリット	22
2.4.3	Topics map における Ontology 表現	22
2.5	Context	22
2.5.1	一般的な Context	23
2.5.2	非文字資料における Context	24
第3章	Context 表現の提案モデル	25
3.1	Context の関係性	25
3.1.1	Context の軸と抽象度	25
3.1.2	Context の組み合わせ	26
3.2	Context のデータ構造	27
3.2.1	データ構造分析	27
3.2.2	Context 間の関係性の表現手法	30

3.2.3	Context の類似度	32
3.3	Context の検索利用例	34
第 4 章	結論	36
	謝辞	38
	参考文献	39
	口頭発表	42

目 次

2.1	第 1 次民俗資料整理カード	9
2.2	第 2 次民俗資料整理カード	10
2.3	Topics map の概観図	19
2.4	Topics map の概観	20
3.1	Context の軸と抽象度	26
3.2	スコープを用いた Context 表現	28
3.3	隣接する主題を用いた Context 表現	29
3.4	Context 用トピックを用いた Context 表現	30
3.5	Context 間の関係を示す Context 表現	31
3.6	提案する Context 構築のためのデータ構造	32
3.7	Context 間の類似度	33
3.8	類似した Context から仮説を立てる例	34
3.9	Context の検索経路の優先度	35

第1章

序論

WWWをはじめとする今日の情報技術の急速な進展によりキーワード検索などにより、簡単に情報に到達で切るようになってきている。しかしながら、技術の向上を大幅に上回る速度で情報資源が増大しているために、所望の情報を得るための労力が減少しているとは言えない。むしろ、情報雑音が増えているために、単作の労力が増加している場合もあると考えられる。

何故この様な問題が生じるかというのと記号化された文字をパターン検索しているからである。パターン検索では、ユーザが入力したクエリと検索対象を記号として比較し、一致するか、含むかなどの真偽で所望の情報であるかどうかを判断する。したがって、記号が指示する概念がクエリと検索対象の間で一致している間はうまく機能するが、両者で齟齬が生じれば生じるほど所望の情報が得られる可能性が急激に下がってしまうのである。この問題を一般化すると、記号システム内のシンボルがどのようにして実世界の意味と結びつけられるかという問題である。この問題は、Steavan Harnadによってシンボルグラウンディング問題または記号接地問題と命名された。

では、代わりにどのような手段を用いればよいかというのと情報システムに意味を取り扱う枠組みを設けることにより解決できると考えられている。情報理論では一般に、メッセージの重要性には立ち入らず、データの質を扱わずに確率論的に扱えるデータの量だけを扱うものとする場合が多い。しかし、意味を扱うシステムが提案されていないわけではない。具体的な例をあげる

と，人工知能分野などで Ontology[9, 10, 8, 11] が議論，提案されている．

Ontology は，意味を扱う枠組みの中では有力なものであると考えられている．Ontology は，そもそも哲学用語に由来し存在論のことを指す．人工知能分野に派生し，「概念化の明示的な仕様」などと定義されている．先に述べたパターン検索に由来する問題のように，従来の文書検索の方法では，単語単位での一致か，よくても類義語を含む文書を検索するのが限度である．ここに，Ontology を導入すると検索対象となる文書が単なる単語の集まりとしてではなく，文書全体で大きな意味を持ったデータとして扱われ，各文書について統一的な付加情報を持たせることができる．このように，メタデータと Ontology の技術を用い，文書の意味に即した処理を計算機が行うことができる Web を Semantic Web と呼び，次世代の検索技術が実現されることなどで期待されている．

さて，ここで本研究の主題を述べる．過去に，神奈川大学が「人類文化研究のための非文字資料の体系化」と言うテーマにおいて 21 世紀 COE プログラムが採択されていた [1]．ここから，着想を得て民俗学研究のための情報システムの枠組みを調査したところ，デジタルアーカイブ [2] が有力な手段であることがわかった．なぜなら，一般にテキストだけではなく，画像や動画などのマルチメディアデータを扱うほうが非文字資料を記述するには適していると考えられる．近年，さまざまなデジタルアーカイブが登場し，普及し始めている段階である [3]．

既存の情報システムでは，非文字資料を体系化するときに記述力の面で根本的に問題があることが分かった．それはどのような問題であるかということと民俗学及びそれを包含する人類学などの学問がどのような研究を必要としているのかを分析することで理解できる．

民俗学では日常生活文化の歴史を，民間伝承を主な資料として再構成しようとする学門である．風俗や習慣，伝説，民話，歌謡，生活用具，家屋など古くから民間で伝承されてきた有形，無形の民俗資料をもとに，人間の営みの中で伝承されてきた現象の歴史的変遷を明かにし，それらを通じて現在の生活文化を相対的に説明しようとしている．このような研究目的に適した研究手法は，分析的というよりも記述的であり，対象へのインテンシブな調査を

元に厚い記述を目指す，いわゆる質的研究の一つに位置づけらるとされている。

厚い記述とは，ギルバートライルによって提唱された概念である．例えば，我々は誰かから目配せをされても，文脈が分からなければそれがどういう意味か理解できない．愛情のしるしなのかもしれないし，密かに伝えたいことがあるのかもしれないというように文脈が変われば目配せの意味も変わる．この厚い記述という概念を人類学その他の領域において導入したのが，クリフォード・ギアツである．1973年の著作「文化の解釈学」[4]で自らの民俗誌記述の方法として提示したことから勇名になった以降，社会科学のみならず広範な学問分野で有力な方法論として強く指示されている．このように，民俗学研究においては，文脈，換言すると Context を詳細に記述することが求められている．ところが，この Context を詳細に記述するということはすなわち意味を記述することであるから，先に述べた Ontology を導入する方法がよいという結論になる．副作用として Ontology を用いて厚い記述をすることにより，様々な推論その他の情報の提案なども可能になる．

この厚い記述のために必要な Context は，すべてを表現するとなると組み合わせ爆発を起こしてしまう．そこで，本研究では Context を扱うフレームワークを設けて組み合わせに制約を加えることで解決することにした．ただし，Context を扱うフレームワークの設計を誤るとそのフレームワークを使わないほうが機能する場合が多くなってしまう．このことを考慮しながら適切にフレームワークを設計することとする．

Context のデータ構造としては，Topics map[18, 19, 20, 21]を採用することにした．

RDFなどのデータ構造でも同値の意味を記述できるが，構造が複雑になりすぎるといった問題などがあるために，通常の Ontology とは別に記述した方がよいと判断したからである．ただし，Ontology において Dublin Core などのように Context の記述のための定義をする方法も考えられる．したがって，将来において Topics map を用いた方法が有力であるとは限らない．

本論文では Topics map を用いて Context を記述する手法を提案した．また，それにより，得られるメリットなどを分析している．

以下，2章では，基礎知識について述べる．次に，3章では，Contextの提案システムについて述べる．4章では最後に5章でまとめを行う．

第2章

基礎知識

本論文での研究試料である特に民具などの非文字資料およびその学術領域である民俗学について簡単に述べる。

2.1 非文字資料

非文字資料とは、文字媒体として記録することのできない情報のことである。資料そのものだけでなくメタデータにも重要な価値のある資料があり、メタデータを含めた資料の全体像が真のデータであるとすれば、真の情報を表現することを目的としている。これらの情報は単純に文字媒体では表すことはできないが、情報発信のためには、資料情報をマシン上に表現しなければならない。そこで、真のデータを再構成可能であるような既存の媒体への写像・変換のモデル化を行う必要が生じる。対象の真の情報を包括的にモデル化する際に必要となるのが、Ontologyである。こうすることで、今まで表現できなかった情報の多くが固定可能となり、情報の次元が表面的な語彙のみによるものから厳密な意味に基づいたものへと飛躍的に高まる。

2.1.1 只見町民俗資料調査カード


只見町民俗資料整理カードとは、図のように只見町およびその近郊における調査で収集した民俗資料のデータを項目にまとめたものである。

生年用貝山穂田辰木祝用集仁アカ5船動 112.1.17

図説掲載
重要

通番号 1916
分類番号 28117-7
21122-8

民俗資料調査カード

資料名	(地方名) ケウナ	(標準名)	写真・形状・寸法等 
寄贈・借用 年月日	昭和・平成 年 月 日		
寄贈者	(住所)只見町大字 字 番地		
所有者	(氏名)		
材質	鉄 雑木		
使用年代	明大 年頃から 明大 年頃まで・現在も使用中		
使用目的	屋根の削りに使った		
收藏場所	旧五十嵐家		
備考	ちゅうな		
調査年月日	平成 2 年 10 月 26 日		
調査員	目黒鶴吉		

長 50 及び 9.5

只見町教育委員会

図 2.1 第 1 次民俗資料整理カード

このカードは、仕様が一度改訂されているので、第一次のものを民俗資料調査カード、第二次のものを只見町民俗資料調査カードと言う。カード項目を列挙すると次のようになる。

第 1 次民俗資料整理カード

1. 通番号
2. 分類番号
3. 資料名(地方名)


只見町民俗資料調査カード		第1次分類番号	第2次分類番号
		通番号 T・A・ <u>1560</u>	分類番号 分類 <u>9</u> 番号 <u>237</u>
標準名		<input type="checkbox"/> 写真・略図・寸法・重量など <input type="checkbox"/> 紀年名・焼印・墨書など	
地方名	カリアガリ ユッコギ (女) 別名		
採集年月日	昭和・平成 年 月 日 採集者 只見町教育委員会		
採集地	只見町大字 景取 字		
旧所有者	(住所) 只見町大字 景取 字 (氏名)		
入手経路	(旧所有者) 自家製・購入・譲え・もらう・交換・その他() 代価 円 (只見町) 寄贈・購入・製作・その他() 代価 円		
材料名	入手方法 採取・栽培・ <u>購入</u>		
製作地	製作者		
製作年代	明・大・ <u>30</u> 年頃まで製作 現在も製作中		
採取・処理・製作方法	<ul style="list-style-type: none"> ・ 加茂織・青木織などを購入し、自家製。 ・ 呼び名は各地で、又は家庭ごとにそれぞれ違った呼び方をしている。 <ul style="list-style-type: none"> ・ カリアグユッコギ ・ ホソユッコギ ・ サルツバカマ ・ カリアガリユッコギ ・ ホソツバカマ ・ モンベ (腰にゴム紐を使ったもの) ・ 男物は小便マチがあり女物にはない。 		
使用地	使用者		
使用年代	明・大・ <u>30</u> 年頃まで使用 現在も使用中 使用時間		
併用民具	保存・管理場所		
使用方法	<ul style="list-style-type: none"> ・ 山仕事、田畑、狩猟等屋外の仕事に着用した。 ・ また、細くて軽く、動き易いため、屋内屋外を問わず仕事着として使用するに便利である。 		
他の使用方法・由来・俗信・禁忌使用地域など		話者名 (住所) 只見町大字 字 (生年月日) 明・大・昭 年生	
		話者名 (住所) 只見町大字 字 (生年月日) 明・大・昭 年生	
	ネガ番号 <u>15-9-2</u>	調査年月日 平成 <u>24</u> 年 <u>7</u> 月 <u>8</u> 日 調査員 馬場 惇	

図 2.2 第2次民俗資料整理カード

4. 資料名(標準名)
5. 寄贈・借用年月日
6. 寄贈者・所有者(住所)
7. 寄贈者・所有者(氏名)
8. 材質
9. 使用年代
10. 使用目的
11. 収蔵場所
12. 備考
13. 調査年月日
14. 調査員
15. 写真・形状・寸法等

第2次民俗資料整理力ード

1. 第一次分類番号
2. 第二次分類番号
3. 標準名
4. 地方名
5. 別名
6. 採集年月日
7. 採集者
8. 採集地
9. 旧所有者
10. 入手経路
11. 材料名

12. 入手方法
13. 製作地
14. 製作者
15. 製作年代
16. 採取・処理・製作方法
17. 使用地
18. 使用者
19. 使用年代
20. 使用時期
21. 使用民具
22. 保存・管理場所
23. 使用方法
24. 他の使用方法・由来・俗信・禁忌使用地域など
25. 写真・縮図・寸法・重量など
26. 紀年名・烙印・墨書など
27. 話者名(住所)(生年月日)
28. 調査年月日
29. 調査員
30. 備考

これらを比べると，第一次と第二次のカードは概ね同じ項目を持っている．この二つを簡単にまとめると民具そのものの情報，民具の使用に関する情報，民具の製作に関する情報，民具の調査に関する情報に分けられる．したがって，只見町民俗資料調査については最低限これら4つの Ontology を構築する必要がある．

2.1.2 民具の分類

民具の分類は、只見町民具分類総括表によると以下のとおりである。

1. 生活用具

衣類 衣類【上衣】，衣類【下衣】，衣類【その他】，雨具，防寒具，履物類【冬用】，履物類【冬用外】，かぶりもの，その他

食生活用具 炊事用具，調理用具，加工用具，保存用具，飲食器，嗜好品用具，その他

住居用具 いろいろ用具，暖房用具，照明用具，家具・調度，寝具，除雪用具，建築・儀礼用具，その他

その他

2. 生産用具

農耕用具 耕作用具，灌漑用具，施肥・除草・病虫害防除用具，収穫・脱穀用具【畑作】，収穫・脱穀用具【稲作】，調整・選別用具，貯蔵用具，その他

山樵用具 木挽用具，搬出用具，炭焼用具，その他

漁撈用具 ヒキカギ類，ヤス類，筌類，網類，釣具，その他

狩猟用具 ヤリ・ワナ類，鉄砲用具，加工用具，運搬用具，衣類，飲食用具，信仰用具，その他

畜産用具 飼育用具，牛馬附属用具，その他

養蚕用具 桑摘用具，飼育用具，上簇用具，その他

染織用具 苧引き用具，整糸用具，地機用具，高機用具，機織関連用具，その他

手工用具 藁加工，つる細工用具，その他

諸職用具 大工用具，桶・樽職人用具，畳職人用具，屋根葺職人用具，木羽屋職人用具，鍛冶屋職人用具，石工職人用具，その他

その他

3. 交通・交易用具

運搬用具 背負用具，腰掛用具，腰下用具，人力運搬用具，畜力運搬用具，その他

旅行用具

通信用具

商業用具

計算・計量用具

看板・広告類

手形・貨幣類

その他

4. 社会生活用具

共有道具

防災・避難用具

家印・印判

武家用具

戦時用具

その他

5. 信仰・年中行事用具

神体・偶像類

神事・仏事用具

奉納・祈願・縁起物類

年中行事用具

信仰関連用具

その他

6. 芸能・娯楽用具

芸能用具・衣装類

仮面

人形

楽曲

遊戯具

その他

7. 人の一生用具

育児用具

成長の祝い用具

婚姻・婚礼用具

葬送・墓制用具

その他

8. 民俗知識用具

教育用具

医療・医薬用具

暦・計時用具

その他

9. その他

その他

2.2 意味論とその隣接分野

ノームチョムスキー要する生成文法論者が言語学者の中の主流であったが、この主流パラダイムに対抗するように認知言語学を打ち立ててきた。なぜな

ら，認知言語学者の多くは，生成文法学者としても研究を進めており，生成文法の欠陥を十分に認識し始めたためであると言われている。

生成文法モデルでは，言語システムが言語固有の普遍文法に基づくものであるとしているものである。これに対して，認知言語モデルでは，上述の普遍文法の可能性は否定しないものの，それは認知に依存しているものであると考えたものである。つまり，言語へのアプローチの方法が普遍文法モデルから認知モデルへと移り変わって言ったのである。言語学では，伝統的に，言語の役割は外界の要素を言語形式にマッピングしたものであると考えられていた。この考えに基づけば，世界は各部分に切り分けることが可能であり，その各部分が言語と一対一に対応していると考えられる。これに対し，認知言語学では世界の要素は様々なとらえ方が可能とする。すなわち，要素それ自体が言語と一対一に対応しているものではなく，要素をある主観を持って概念化するものが言語であるとするものである。

この主観を極力排除する道具が自然科学であり数学である。しかし，主観を積極的に排除し続ければ，信頼性は高まるが客観性を持たないものについて言及することができない。

この主観を含めて概念を記述することが，現状のテキストベースの情報処理を打開するものとなりうるはずである。

2.3 Ontology

Ontology とは，哲学用語の存在論 (ontology) に由来する。哲学における存在論とは，存在そのものの根源を論じる学問である。そこから派生して，計算機科学や情報科学の分野においては，ドメイン内での概念集合および概念間の関係性の形式的表現をするものである。これは，あるドメインの特定のプロパティについて推論することに使われる。また，ドメインを定義するためにも使われることがある。以降では，哲学用語のオントロジと区別するためにこれを Ontology ということにする。

工学的には，Ontology は「形式的な，明示的な共有概念化の仕様」である。Ontology は共有語彙を提供し，それによってドメインをモデル化するために

使うことができる。現在，Ontology は世界またはその一部についての知識表現の一形態として，人工知能，Semantic Web，システム工学，ソフトウェア工学，医療情報学，図書館学，その他多くの分野で用いられている。

中でも Semantic Web [6, 7] は Ontology によって人工知能を Web 上に展開する野心的な試みがなされている。他の分野と同様に Ontology を概念化と概念間の明示化に利用している。Semantic Web の特徴は，Web にメタデータを付与しそれらのメタデータの意味を Web Ontology で定義することである。書誌情報などで使われるであろうメタデータについては，すでに Dublin Core[5]などで定義されている。

Semantic Web においては，RDF などで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，「一郎さんが鋤を所有している」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鋤」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」という述語がどういう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。

Semantic Web においては，Web Ontology 言語 (OWL) を用いて記述することが多い。OWL に限らず Ontology を記述する場合は，正確性が重要になることからツールで記述をしたほうがよい。本研究においては，著名なオントロジ

エディタである protg[12] を用いて記述することにした。このツールは、OWL で記述可能なクラスやプロパティ、制約条件やインスタンスなどの入力時に間違いが起こりづらいように考慮されたインターフェースが採用されている。また、プラグインによる拡張でさまざまな機能を拡張することができる。そのプラグインに OWLVis[13]、OWLVis[14] などがあり、記述した概念を有向グラフとして可視化してくれるので全体像を把握するためによく利用することができる。

2.3.1 Top level Ontology

まず、Top level Ontology を大きく分けて2つに分類できる。抽象的な側面を重視して構築する形式オントロジ (formal Ontology) と、実体的な側面を重視して構築される形式オントロジがある。前者には Sowa、などが、後者には DOLCHE などがあり、それぞれに特徴を持っている。

現在、有力な上位 Ontology として考えられているものについて比較してみるとどれも他を圧倒して優れているわけではないので、どれを選ぶかはドメイン Ontology の記述に適しているかに依存するが、今回は、Universal な属性については考慮せず、民具資料の実態についての Particular の属性についてのみ定義することにした。

2.4 Topics map

RDF などの知識表現を使って意味を記述することができるものの表現が複雑になりすぎるという問題がある。そこで、Topics map という RDF などに類似するデータ構造を取り入れることで、Context の表現方法を簡潔にすることができる。

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 Topics map がある。図 2.3 は Topics map は Topics map の概略を示したものである。Topics map は ISO 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

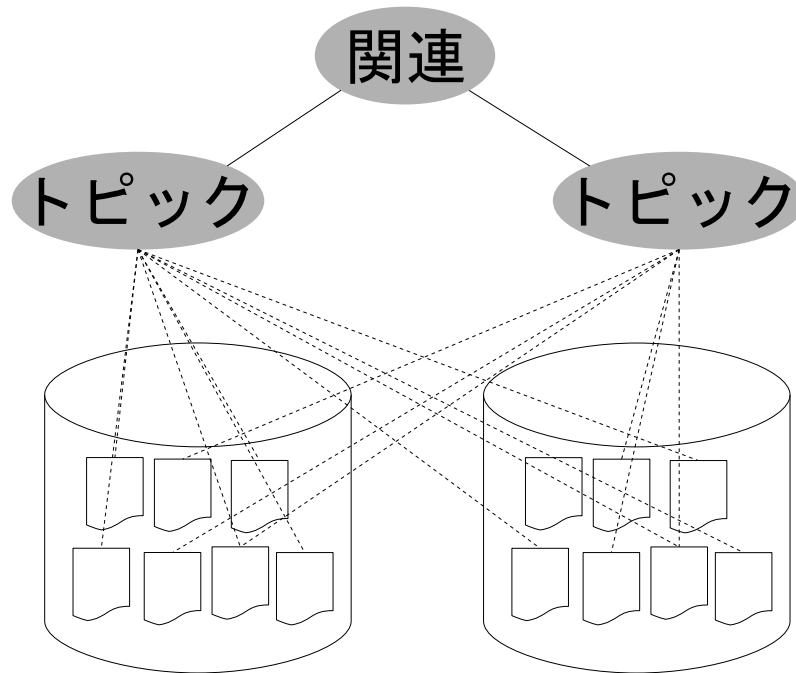


図 2.3 Topics map の概観図

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、Topics map 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 Topics map でオントロジを記述し、インスタンスを登録するといった方式の Topics map 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に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と Context 部分について Topics map 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。

Topics map 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。

本章では，本論文の提案に手段として利用する Topics map について説明する。

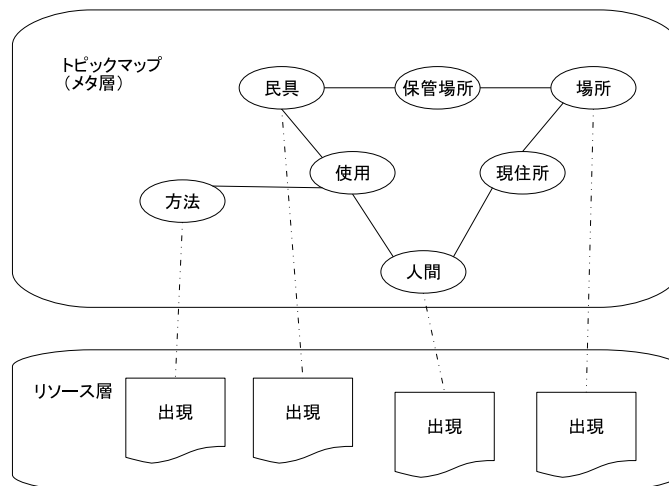


図 2.4 Topics map の概観

Topics map は，知識を記号化し，この記号化された知識を関係がある情報リソースに結び付ける技術である。図 2.4 に示すように，Topics map は，論議の主題 (Subjects) を表現するトピック (Topics)，主題間の関係性を表現する関連 (Associations) 及び主題と主題に関連する情報リソースを結び付ける出現 (Occurrences) によって組織化される。ISO/IEC JTC1 SC34 WG3 で策定された標準 (ISO/IEC 13250 Topic Maps) であり，現在では広く活用されている。タクソノミーやシソーラス，Ontology などと同じく知識表現の手段の一つで

ある。主な特徴を挙げると、情報リソースと明確に分離されており、情報リソースに依存しないという利点や、自由に関連性を定義可能であり、自由度が高いという利点がある。また、Topics map には標準で主題の有効範囲を設定できるスコープという機能や、同一の主題と同定できるものについては併合できるという機能が盛り込まれているのでネットワーク上で、セマンティクスネットワークを共有する際に優位であるという利点もある。

2.4.1 Topics map の主要な構成要素

図に示すように、Topics map はトピック、関連、出現の3つの主要な構成要素からなる。

トピック (Topics) 人間が認識する具体的または抽象的な主題/概念を表す要素。

関連 (Associations) 一つ以上の主題間の関係の表現。関連は役割を持っている。

出現 (Occurrences) 主題及び情報資源間の関係の表現する。出現は型(主題)を持ち、出現によってリンクされた主題及び情報資源間の関係の性質を記述することが可能。

公開主題識別子 (Published Subjects Identifier) トピックや関連といったメタ層に属するノードを Subjects と呼んでいるが、それらはweb上に公開することができ、サブジェクトを共有できる。さらに、Topics map 自体がもつサブジェクトのマージ機能により Topics map 同士が接続され、シームレスなセマンティクスネットワークが形成される。

有効範囲 (Scope) Topics map のトピック名、関連、出現についての有効範囲を設定できるものである。この有効範囲が同じ集団の Topics map を集めることで、任意の Context についての Topics map を抽出できると考えた。

以下に、スコープの典型的な利用例を挙げる。

国を表すトピック名：English(”country”), Japanese(”国”), Norwegian(”land”)

異形名をもつトピック：(”ニナワ”, ”ニナ”);

関連が成立するドメイン：subject area (history, biography, culture, politics);
historical period (ancient, classical, medieval, modern; past, present)

2.4.2 Topics map 利用のメリット

Topics map は、意味ネットワークを形成するメタ層とデータを保持するリソース層が明確に分離されている。したがって、データモデルの変更なしにオントロジの拡張が可能となる。

RDF などと異なり、その構造を知らなくても機械的に処理すればよいので、情報の多様性の表現力や変化に対する頑強性が高い。Topics map のメリットは、柔軟性が高いということに尽きる。マージ機能やスコープを用いたフィルタリングを利用できる Topics map を利用した方が、メリットが多い。

述語論理の表現からはやや逸脱するので推論可能性、知識表現の柔軟性が高い。

2.4.3 Topics map における Ontology 表現

Ontology を「概念と概念間の定義と体系化」とするならば、Ontology における概念を”トピック”に、概念間の関係を”関係”に写像することで、Ontology を Topics map で表現可能といえる。したがって、Ontology で記述可能な知識表現は、Topics map でも同様に可能であるといえる。具体的には表 2.1 のような対応関係になっている。

2.5 Context

ここでは、非文字資料に依存した Context を述べる前に一般的な Context の概念について説明する。

表 2.1 Ontology と Topics map の要素間の対応

OWL/RDF	Topics map
主体	トピック
述語	関連
客体	トピック/出現

2.5.1 一般的な Context

一般に、Context とは「文脈」あるいは「前後関係」、「背景」と直訳される。文字通りに、文中における論理関係の脈絡を表す。すなわち、コミュニケーションの場で使用される言葉や表現をて意義付ける背景や状況そのものを指している。相対的に定義が異なる言葉の場合は、コミュニケーションをとる2者の中でその関係性、背景や状況に対する認識が共有・同意されていることがコミュニケーションの成立の条件となる。このように、コミュニケーションを成立させる共有情報を Context というが、以上の共有情報をマシン上にモデル化する上で、知識処理における Context は言語的なコミュニケーションのみならず、

例えば、同じ時速 10km でも、徒歩 Context ならば速い、自転車 Context ならば普通、自動車 Context ならば遅いというように評価が異なる。これは、速度という思考対象が、移動手段という Context に依存していることを表している。

主要な Context

Context は様々な前後関係、背景について表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。

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、本論文で主に述べている 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

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，知りたい知識には前提条件がある．

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．

2.5.2 非文字資料における Context

ここで，本論文にの重要な主題である非文字資料における Context について説明する．Context とは，知識表現を行う対象世界において，普段われわれが暗黙的に考慮しているものごとと考えられる．その暗黙的に考慮されている Context の違いによって，概念化を行う際に視点依存性などとして現れる．ここで，Context の違いから以下の問題が生じる．

- 計算機的な意味の一貫性を保つことが困難である．
- 異なる対象世界（領域）での相互理解・利用が困難である．

これらの問題は質的研究の一部である非文字資料を対象とした場合に顕著に現れる．質的研究においては厚い記述をすることが求められているのでさまざまな Context によって定義が異なることも多いからである．

第3章

Context 表現の提案モデル

これまで見てきたような意味を扱う場合に特に見られる問題点を解決するような Context モデルを提案する。提案する Context モデルの構成要素，Context 同士の関係性，Context の軸，抽象度，組み合わせなどを先に述べながら全体のデータ構造などを説明する。また，本モデルの利用例も述べる。

3.1 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。そうすると 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。

3.1.1 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，Contextet 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 3.1 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。

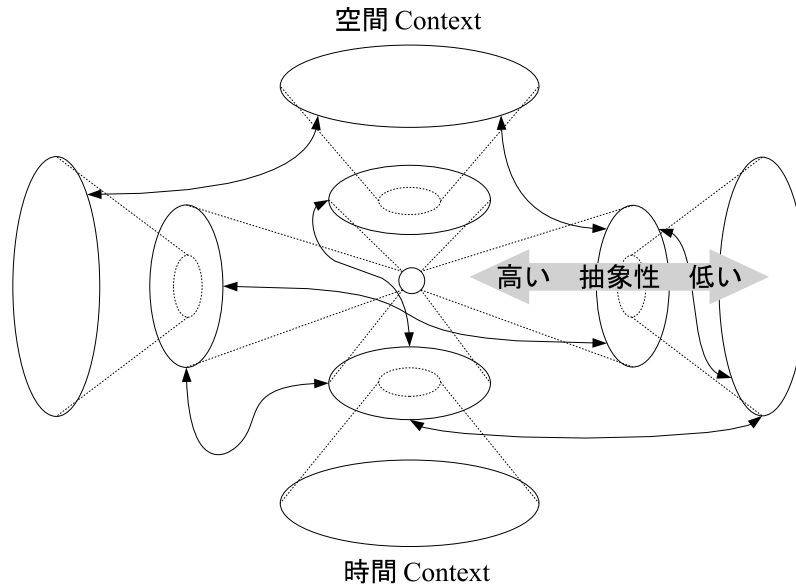


図 3.1 Context の軸と抽象度

例えば、「江戸時代の食文化」といった Context があったとする。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ことが必要になる。

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。

3.1.2 Context の組み合わせ

図 3.1 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 を作るのとほぼ同じことになる。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．

3.2 Context のデータ構造

3.2.1 データ構造分析

Context 間の関係性の表現手法

Context を表現するには，いくつかの手法が考えられるが，本論文ではトピックマップの主題に Scope で直接有効範囲を与える Context 表現及びトピックマップの隣接したノードを有効範囲とする Context 表現，Context のトピックによる Context 表現について説明する．

- scope の利用して Context を判断する方法 (図 3.2)

この方法は，共通の scope を持っている主題を Context とする方法である．図中の主題に注目するとわかるが，主題は Scope をもっている．図ではカッコ内の英字で表記されている．Scope は空白文字で連結することで，任意の個数保持することができる．Scope A に着目した場合の一般的な Context は，図中の枠で囲われた context A 内の Subject 集合である．

この方法は，最も自由度が高い方法であり，Scope の集合と Context の集合の関係の定義もいくつか考えられる．[22] (集合が等しくなければならない，Scope の集合が Context の集合サブセットあるいはスーパーセットでなければならないなど) トピックマップのつながりに依存せずに，任意の主題の集合を Context とすることができ，表現力が非常に高いという利点がある．一方で，

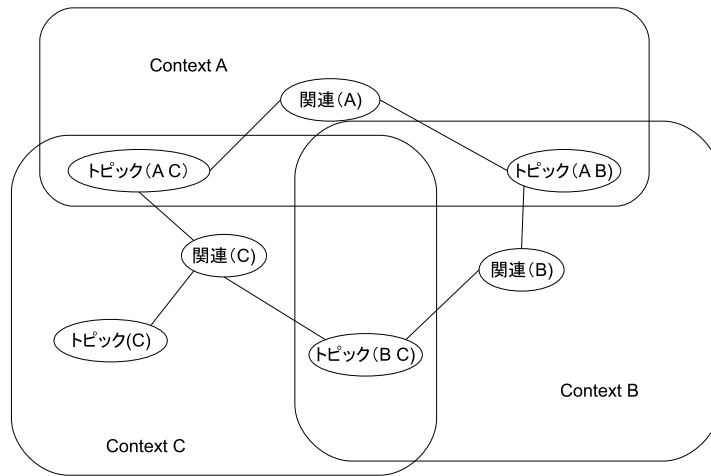


図 3.2 スコープを用いた Context 表現

Scope はあらゆる視点や背景などが対象となりうるので，Scope の記述量はすぐに現実的な量を超えてしまうという欠点がある．

したがって，Scope の使用例ででてきた言語を指定するといったようなグローバルな Context については，Scope による記述が望ましく，ドメインに依存している Context などは，他の記述方式を用いた方がよいといえる．

- 隣接している主題から Context を判断する方法 (図 3.3)

あるトピックに接続されている出現，あるいは関連およびその先にあるトピックを Context とする方法である．図中の Subject に着目すると関連に注目したときは，その隣接する Subject であるトピックの集合トピックに注目したときは，その隣接する関連およびその関連の隣接サブジェクトの集合全体を Context とするものである．

トピックマップの構造のみを利用している方法なので，新たに特別な操作が必要ないという利点がある．しかし，これはデータモデルとしてはトピックマップそのものであり，クエリや推論エンジンが隣接する Subject を Context

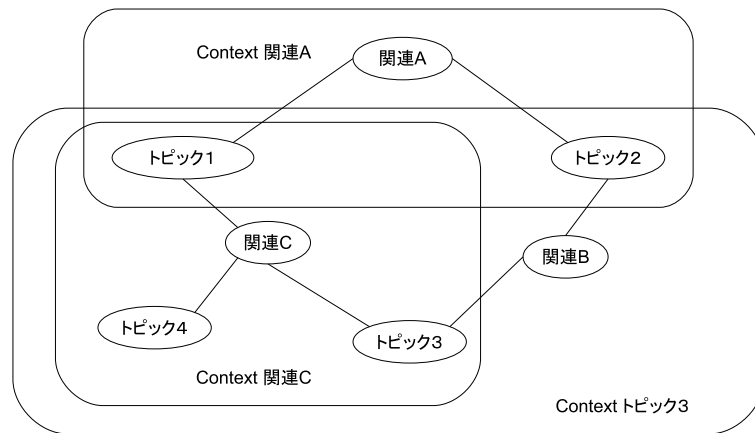


図 3.3 隣接する主題を用いた Context 表現

として扱うように考慮して設計されたものでなければならないという欠点がある。また，Context の表現力は，今回提示する手法の中で一番低い。

- Context 用トピックから Context を判断する方法 (図 3.4)

Context を表すトピックに接続されている関連及びその先にあるトピックを Context とする方法である。図中のトピック，関連から Context を示すトピックへとリンクが張られているが，これは抽象的なもので，関連を含みそれを経由してリンクされている。また，Context のトピックから関連につながっているが，これも抽象的なものであり，関連を具体化し，その具体化したトピックと Context を示す関連で接続されている。

この方法は，Context 自身をトピックとすることで，text 間の関係性を記述できるという点が他のモデルと異なる特徴である。Context を表すトピックは Context の関連で接続され，Context が，どのトピックとどのような関係にあるのかわかるようになっている。

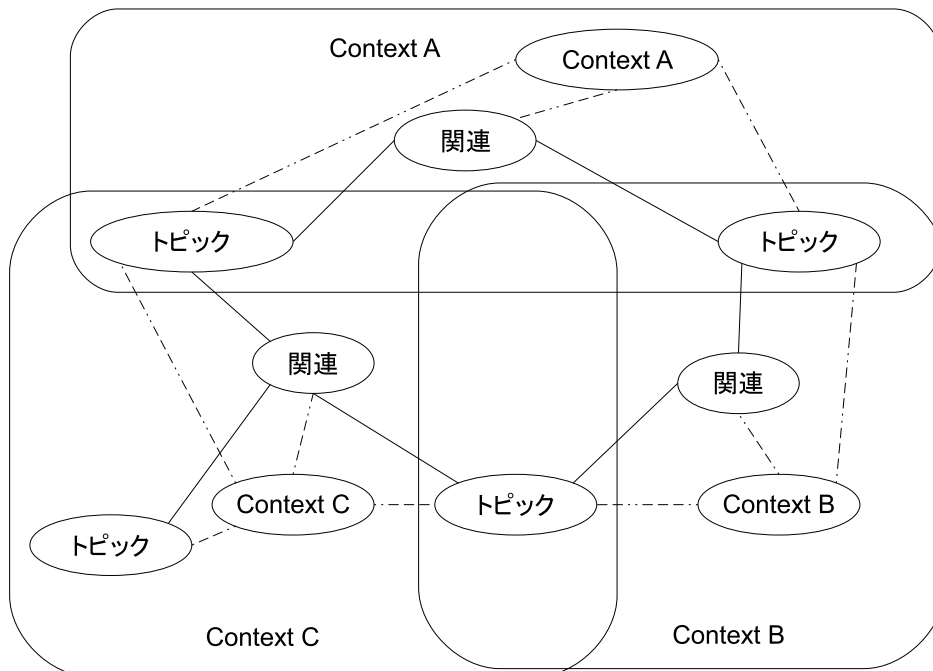


図 3.4 Context 用トピックを用いた Context 表現

3.2.2 Context 間の関係性の表現手法

これまで、Context の表現手法について述べてきたが、Context 間の関係性についても表現しなければ、トピックの単純なタクソノミーを生成したに過ぎない。Context 用トピックを用いて、Context 間の関係性を記述したものを図に示す。これは、データモデルとしては、トピックマップ自体の関係性を記述する上位の層に位置する。

このようにして表現した Context 間の関係によって、Context そのものを推論によって導くことを可能にすれば、資料に暗黙に隠されていた意図などが発見できると考えられる。図 3.4

図 3.6 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

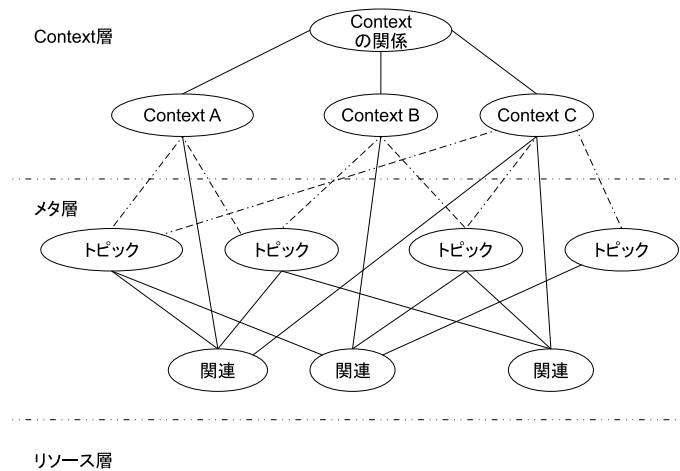


図 3.5 Context 間関係を示す Context 表現

述の例である。

まず、海に関係している Context と、川に関係している 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 3.1 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関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 関係を内部出現として保持する。

このように Context 間関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 間関係を表すノードを参照するだけでよくなる。Context の関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

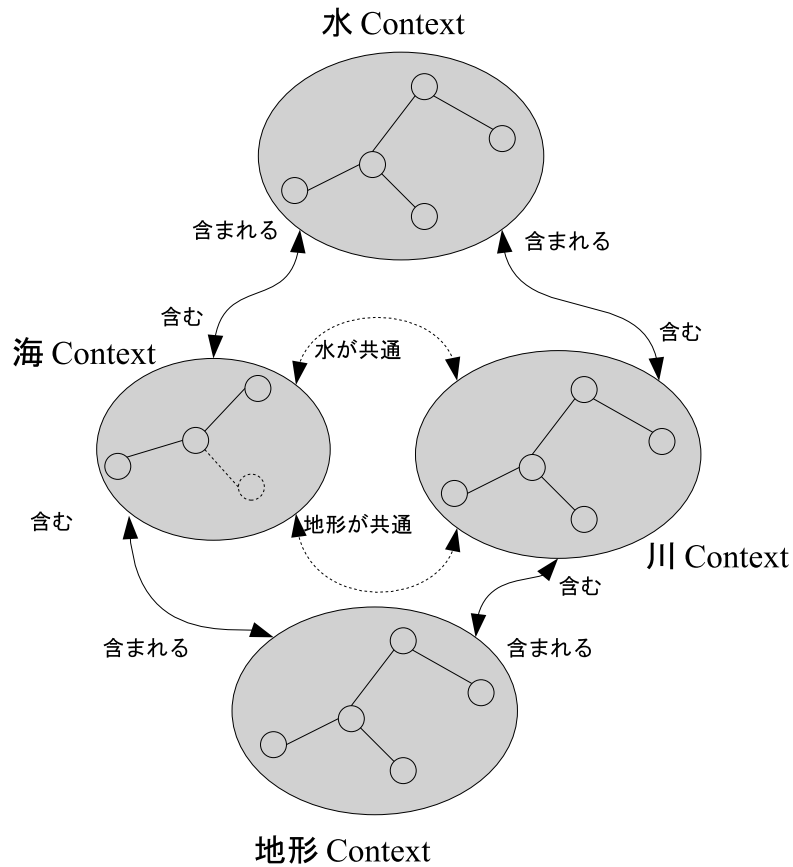


図 3.7 Context 間の類似度

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い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」とよいという仮説が得られる。

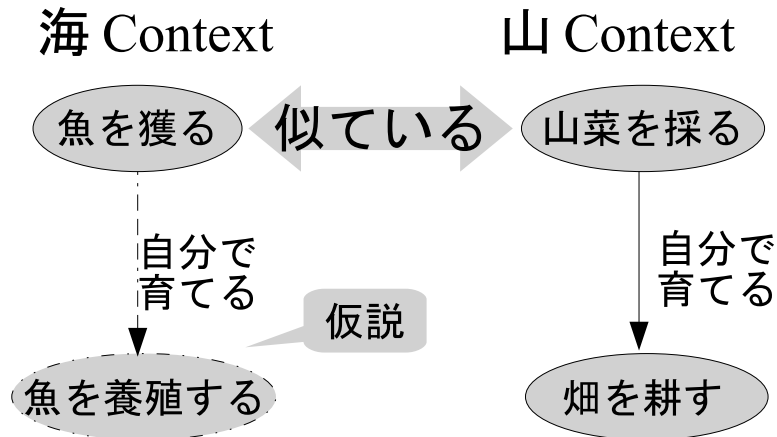


図 3.8 類似した Context から仮説を立てる例

このように，Context 同士の関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 等の関係がない 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．

3.3 Context の検索利用例

ここまでに示してきた 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 3.9 に示す．

図 3.9 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る．

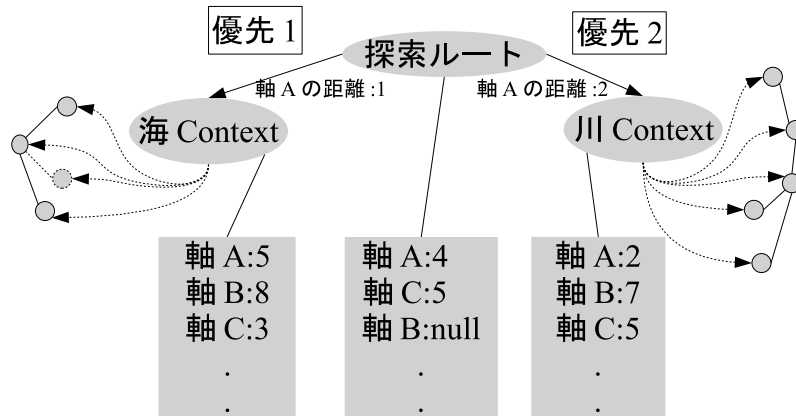


図 3.9 Context の検索経路の優先度

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の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。

第4章

結論

本論文では、非文字資料を主な対象とした Context の表現手法について述べた。その中で、Context の性質の分析からトピックマップを用いた表現手法が適していると考えた。

Context には、常民の隠れた意図や、研究者の無意識な視点の違いなどさまざまなものが考えられる。これらの Context を表現することは、研究者間の民俗資料の共有・相互利用を進めた際に起こりうる問題を解消することになる。社会的背景・歴史的背景などによって世界観が異なることによる視点の違いである。

したがって、Context を表現することが非常に重要となる。ISO により標準化されたトピックマップというメタモデルを表現する規格を活用し、Context の表現手法を明らかにした本研究により、非文字資料の Ontology において Context の表現が可能となり、よりいっそう民俗学の研究に適した情報の共有・相互利用の実現性が高まったといえる。

本論文では、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 について考察した。これまでは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、本論文では Context の特性を考察し、既存の情報システムに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.

今後の課題としては, 暗黙的な Context を導出および証明, Context を適切に生成や破壊をせたり, 自動で調整を行えるような仕組みを構築することなどがあげられるである.

謝辞

本研究を行なうにあたり，終始熱心に御指導していただいた木下宏揚教授に心から感謝致します。また，様々な面で数多くの有益な御助言をしていただいた鈴木一弘助手に深く感謝致します。さらに，公私にわたり良き研究生生活を送らせていただいた木下研究室の方々に感謝致します。

参考文献

- [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] 笠羽 晴夫: “ 1. デジタルアーカイブの歴史的考察 ”, 映像情報メディア学会誌, 61, 11, pp.1545-1548 (2007) .
- [3] デジタルアーカイブ推進協議会, “国内グッド Web サイト事例集:” ,
<http://www.dublincore.org/>
- [4] C. ギアーツ, “文化の解釈学 1 (岩波現代選書)”, 岩波書店, 1987, 324p
- [5] “Dublin Core Metadata Initiative” ,
<http://www.dublincore.org/>
- [6] “Semantic Web official website” ,
<http://semanticweb.org/>
- [7] “W3C Semantic Web Activity” ,
<http://www.w3.org/2001/sw/> 人工知能におけるオントロジーとその応用
- [8] 武田英明 “人工知能におけるオントロジーとその応用” , 情報知識学会研究報告会講演論文集 , 12(4) , pp , 559-569 , 2001 .
- [9] 溝口理一郎 , 池田満 : “オントロジー工学序説 : 内容指向研究の基盤技術と理論の確立を目指して” , 人工知能学会誌 , 9 , pp , 1-12 , 1997 .

- [10] 溝口理一郎, 池田満, 来村徳信: “オントロジー工学基礎論-意味リンク, クラス, 関係, ロールのオントロジー的意味論”, 人工知能学会誌, 14(6), pp, 87-100, 1999 .
- [11] 溝口 理一郎, 人工知能学会 . “知の科学 オントロジー工学” . 第 1 版 . 東京, オーム社, 2005, 280p
- [12] “The Protg Ontology Editor and Knowledge Acquisition System” ,
<http://protege.stanford.edu/>
- [13] “OWLviz” ,
<http://protegewiki.stanford.edu/index.php/OWLviz>
- [14] “The OWLPropViz plug-in” ,
<http://www.wachsmann.tk/owlpropviz/>
- [15] 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”
- [16] 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”
- [17] 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”
- [18] The TAO of Topic Maps
<http://www.ontopia.net/topicmaps/materials/tao.html>
- [19] ISO/IEC 13250-2 Topic Maps - Data Mode <http://std.dkuug.dk/jtc1/sc2/wg2/>
- [20] トピックマップ入門
http://www.knowledge-synergy.com/topicmaps/document/TopicMaps_Introduction.pdf

[21] XTM 1.0

http://www.y-adagio.com/public/standards/tr_xtm/xtm-main.htm

[22] Towards a General Theory of Scope

<http://www.ontopia.net/topicmaps/materials/scope.htm>

口頭発表

- [1] 山下祐平, 木下宏揚, 森住哲也, “トピックマップを用いた非文字資料における Context の表現,” 信学技報, vol.107, no.375, pp.33-38 (2007-12).
- [2] 山下祐平, 木下宏揚, 森住哲也, “Context 表現のための手法の考察,” 信学技報, vol.108, no.160, pp.103-160 (2008-7).