

平成 25 年度卒業論文

論文題目

Ontology を用いた非文字資料の検索

神奈川大学 工学部 電子情報フロンティア学科
学籍番号 201002748
羽生 敏英

指導担当者 木下宏揚 教授

目次

第1章 序論	3
1.1 背景	3
1.2 問題点	4
1.3 目的	4
第2章 基礎知識	5
2.1 非文字資料	5
2.2 民具	5
2.2.1 民具カード	5
2.3 Ontology	7
2.3.1 Ontology の役割	8
2.3.2 Ontology の構成要素	9
2.4 Context	10
2.5 OWL (Web Ontology Language)	10
2.6 Protege	11
2.6.1 Protege4.3 の使い方	11
第3章 提案	14
3.1 提案モデル	14

図目次

2.1	ユリノハの民具カード (表)	7
2.2	ユリノハの民具カード (裏)	7
2.3	is-a 関係構造図	9
2.4	part-of 関係構造図	9
2.5	attribute-of 関係構造図	10
2.6	instance-of 関係構造図	10
2.7	Protege4.3 起動画面	11
2.8	Protege4.3 クラス編集画面	12
2.9	Protege4.3 オブジェクトプロティ編集画面	13
2.10	Protege4.3 データプロティ編集画面	13
3.1	民具 Ontology	14
3.2	可視化した民具 Ontology	15

第1章 序論

1.1 背景

従来、各博物館に収蔵される博物館資料について検索する場合、直接博物館に問い合わせ目録を調べ、実際に現地へと赴く必要があった。近年、博物館資料の情報を、インターネット上で公開する試みがいくつか物の博物館で行われている。しかし、各博物館が持つ台帳の情報は膨大な量であり、多くの博物館にてその全てを有効に活用できていないのが現状である。そうした情報をデジタル資料として整理するとともに、インターネット上で公開する際には効率的に情報検索を行う必要がある。

そうした大量の情報の中で、ユーザにとって有効な情報検索を行うためには高度な知識処理が不可欠である。現在の WWW はリソース間のハイパーリンクによって形成されているが、これは人間に焦点を当てたものであり、人間が読むためのものである [1]。つまり、機械は内容を読み取ることはできても理解することはできない。見かけ上コンピュータは情報を蓄積し分類し表示するが、それらの情報をデータとして扱うだけで、情報の意味するもの自体の理解を必要とする処理を行うことはできない。与えられた情報を機械可読なかたちに変換することを目的とし、セマンティック Web が考案された。

セマンティック Web ではメタデータを用いて大量の情報を効率的に、高い精度で情報検索を行うことができる。また、異なる知識源の相互運用を保証する役割を担うものとして Ontology が構築される。標準化された概念を超えて知識交換や検索を行う場合には諸概念同士の関係を規定する必要がある。こうしたメタデータ記述の意味的な関係が Ontology で定義される。

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」による民俗学研究資料の情報共有・情報流通を背景とし、本研究では非文字資料に焦点を当て研究を行う。民族文化を対象とする非文字資料では、一般的には同じものを指す場合でも地域や年代によって相違が生じることがある。そのため、非文字資料の情報検索にはメタデータをを用いた意味検索が求められる。

本研究室では福島県只見町に伝わる民具についての情報が記載された民具資料カードを用いて研究が行われてきた。この県只見町には「只見町インターネット・エコミュージアム」というホームページがある。ここでは、ユーザに向けて民具の公開を行っている。

前年度の研究では、法造による只見町インターネット・エコミュージアムの「キー

ワード」検索のために「キーワード」の増加が行われた [2]。

本研究では、ユーザが意図する検索結果をより正確に出力するために、デジタル化された非文字資料のデータベースから Ontology 構造を用いた知識ベースを構築し、曖昧な検索キーに対して推論を行うことで検索性の向上が必要である。

1.2 問題点

ユーザの意図する検索結果において、「検索条件を満たす」という点だけを考えると、全ての概念を検索結果として返すべきである。しかし、ユーザの意図を考慮すると、検索結果として返すべき範囲が異なる。また検索結果の中にユーザが求める情報が含まれていたとしても、膨大な検索結果が返ってきた時などには、求める結果の有無がユーザにとって不鮮明である場合も考えられる [3]。また、ユーザにとって検索に対し専門知識が必要となる場合、「一般的なことば」と「専門的なことば」において専門性が壁となる。専門分野の「ことば」を知らないとうまく検索をかけることができないことになる。加えて、概念体系が認識にバイアスを与える問題がある。これは、専門分野の認識と市民の認識とが異なっている場合である。異なるコミュニティでは、それぞれの概念体系に価値判断に基づく系統的なバイアスが掛かっている可能性がある。そういった「認識の枠組み」を考慮しなければならない [4]。

1.3 目的

本研究では、Protege 4.3 によって OWL を用いた民具 Ontology の作成を試みる。博物館資料は減るということがなく、増え続けていく。そのため、容易に情報の追加が可能であることが好ましい。Ontology を導入することにより、文書はただの単語の集まりではなく、文書全体で意味を持ったデータとして扱われる。これにより、本当に必要な情報を的確に検索することが可能となる。また、視覚的に各情報の持つ関係を捉えることが可能となる。そして、最終的には専門的な知識を持たなくても高度な検索を行えるシステムの構築を目指すものである。

第2章 基礎知識

2.1 非文字資料

非文字資料とは、ただ単に「文字資料ではないもの」というだけでなく、文字媒体として記録することなく受け継がれてきた感覚や経験といった文化に関する民族学的資料を指す。文字に表現されていない人間の観念・知識・行為は抽象的で人類文化研究のためには一般的理論化を図る必要があった。

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」では、図像資料、写真・映像にとらえられた景観、姿勢・動作を含む身体技法の3分野を設定し、体系化を行った [5]。

2.2 民具

本稿では、非文字資料の一例として民具を取り上げる。民具とは人々が生活の必要から製作使用してきた伝統的な器具造形物の総称である。いわば日用品とほぼ同義の言葉であるが、一般的には機械によって大量生産された製品を除外する。民具の成り立ちや民具同士の関連性から、当時の人間の営みや生活を知ることができる。

2.2.1 民具カード

民具カードとは福島県只見町に残されている民具情報を記録したカードである。民具を実際に使用した人が直接カードにきろくするという点で、学術的な研究対象としても評価が高く、只見方式と呼ばれ国の有形文化財に指定されている。

多くの民具整理作業では、調査者が使用者から民具に関する情報を聞き取り、それをカード化して整理する手法が取られているが、この方法だと調査者の見解が含まれてしまい、その民具独特の情報が捨てられてしまう危険性がある。只見町では、使用者＝調査者になることで、細かい民具の情報までがカードに記入され、今まで研究者が着目してこなかった民具の情報が盛り込まれている。「只見町方式」によって整理された民具は4417点にのぼり、1992年に『図説 会津只見の民具』（只見町史編さん委員会1992）という報告書にまとめられている。

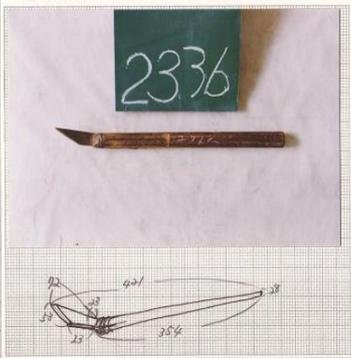
それ以降も継続して整理作業が進められ、現在では8000点以上の民具が収蔵・整理されている。そして、2005年には、「会津只見の生産用具と仕事着コレクション

ン」という形で、2333点の民具が国指定重要文化財に指定された（只見町教育委員会 2005）。「会津只見の生産用具と仕事着コレクション」では、只見町という山村に特化した民具である「ゼンマイ採り用具」、「水田稲作用具」、「畑作・焼畑用具」、「狩猟用具」、「漁撈用具」、「山樵用具」、「麻糸製造用具」、「マタタビ細工用具」、「屋根葺き用具」、「仕事着」という10分類の民具が選ばれている。

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」では、只見町の民俗とともにこの民具をデータベース化しWEB公開する計画を立て製作を行った。このシステムは「只見町インターネット・エコミュージアム」と名づけられ、只見町の俯瞰画像から只見町の民俗を提示し、只見町の山村生活を表したイメージ図から生業を理解することができるシステムになっており、その中で、民具データベースは、各民俗や生業に関する民具を表示する形になっている [6]。

民具カードは客観的に実測された記録であると同時に、使用者による主観的な情報も含んでおり、只見地方の民具資料として詳細に記述された貴重なデータである。また、経験や知恵を伝承していくうえでも、資料価値の高い文化的価値を持つ。民具カードは表裏に記載されており、民具の用途などが書かれている。

図 2.1、2.2 に民具カードの一例を示す [6]。

民俗資料調査カード			通番号	分類番号
			2336	28122-1
資料名	(地方名) ユリノハ	(標準名)	写真・形状・寸法等	
寄贈・借用年月日	昭和(平成)6年5月 日			
寄贈者	(住所)只見町大字大谷	番地		
所有者	(氏名)吉澤 巧			
材質	木と金			
使用年代	昭和から	現在まで・現在も使用中		
使用目的	大工道具 湾曲の部分と削る			
収蔵場所	朝日(m)			
備考				
調査年月日	平成10年7月23日			
調査員	横山 哲夫			

只見町教育委員会

図 2.1: ユリノハの民具カード (表)

その他	
	
年 号	
焼 印	
馬 書	
年 号	
等	

只見町教育委員会

図 2.2: ユリノハの民具カード (裏)

2.3 Ontology

Ontologyとは哲学用語で存在論のことである。情報工学の分野では、セマンティック Webや人工知能等にて利用され、知識の共有のため確立されている。また、オントロジーは単なる「語彙」ではなく「概念」の体系である。情報工学の立場からは「概念化の明示的な記述」と定義される。

Ontology は共通語彙（概念）を提供する体系化された辞書のようなものである。Ontology の最も基本的な利用法として、Ontology で定義された概念を、知識を表すための共通の語彙（概念）として利用するという形態がある。

知識を計算機に格納して知的な処理を行おうとする際には、単なる自然言語での記述ではなく、何らかの計算機が処理可能なフォーマットで表すことが重要である。

しかし、そこで知識の記述に用いられる語彙が統一されていないと、せっかく計算機に格納した知識を、共有し活用することができない。そこで知識を記述する際に用いる語彙を Ontology としてあらかじめ定義しておき、それらを知識記述の際に共通して利用することで、知識の共有・再利用性を向上させることが可能となる。セマンティック Web においては、メタデータを記述する際の共通語彙を提供するために Ontology が用いられる。このような意味で、Ontology は辞書のような働きをするといえる [7]。

2.3.1 Ontology の役割

例えば「ドーピング」という言葉はスポーツ界においては選手が薬物を用いる不正行為を指すが、材料分野では材料に添加物を加えて材料の特性を変えることを意味する。しかも同じ材料分野においても、このドーピングという言葉は金属やセラミックの領域と半導体分野などでは概念の捉え方が、変わる。このようにバックグラウンドにある暗黙的な情報の違いにより、語彙やそれによって記述された知識の意味が変わってくる。そのような暗黙情報を明確にすることが、Ontology の果たす役割でもある。そのため、Ontology では表面的にどのような語彙を用いるかというラベル（概念の名前）の問題よりも、その概念がどのような意味を持つか、という概念定義の問題を重視する。

その結果として、Ontology に基づいて知識を記述することによって、その知識が表している内容が明確になり、Ontology は相互理解を助けることができる。これは知識を処理する複数の計算機システム間でのやり取りにおいては知識の相互運用性の向上につながる [8]。

2.3.2 Ontology の構成要素

Ontology は対象世界を説明するのに必要な概念「概念クラス」と、それぞれの概念間の関係「意味リンク」から構成される。

- is-a 関係 … 下位概念 B と上位概念 A の間には「B is-a A」という関係が成立する。例えば「昆虫」と「害虫」の間には害虫 is-a 昆虫という関係が成立する。

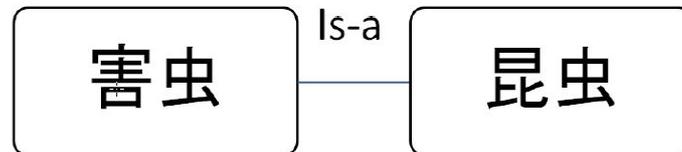


図 2.3: is-a 関係構造図

- part-of 関係 … ある概念と、その概念を構成している部分に当たる概念との間の全体-部分関係を表す。例えば「トンボ」とその構成要素である「複眼」との間には複眼 part-of トンボという関係が成立する。

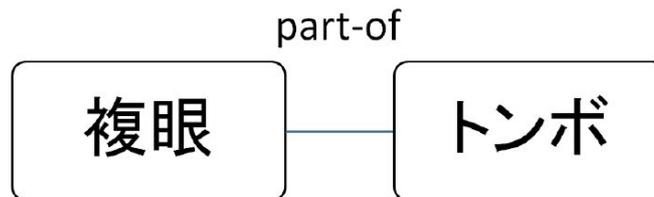


図 2.4: part-of 関係構造図

- attribute-of 関係・・・ある概念を構成している属性情報（色、形状等）を表す。例えば「トンボ」の構成要素である「複眼」の属性情報は丸い (is-an) attribute-of 複眼 (which is-a) part-of トンボという関係が成立する。



図 2.5: attribute-of 関係構造図

- instance-of 関係・・・概念とその具体例との間の関係を表す。例えば「害虫」の instance である「蚊」は蚊 instance-of 害虫という関係が成立する。

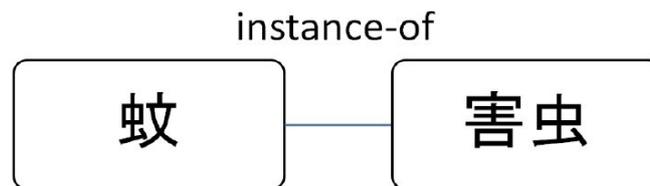


図 2.6: instance-of 関係構造図

2.4 Context

非文字資料には同じ「こと」を異なる視点から見て、異なる語彙で指し示したり、逆に同じ語彙で異なるものを指し示すことがある。これは共通語彙が示す Context の違いによるもので、Context を考慮することは、民具間の新たな関連性の発見に繋がる等重要な要素である。

2.5 OWL (Web Ontology Language)

Web Ontology Language の略であるで、本来であれば WOL となるはずであるが、発音のしやすさと智慧の象徴フクロウとの語呂合わせなどから、OWL という名前となっている。

2.6 Protege

Protegeとは、Ontologyを構築する支援ツールである。スタンフォード大学のメディカルインフォマティクスに所属するマーク・ミュセンらのグループによって開発された。protegeは構築のフェーズよりもむしろ、利用フェーズに重点が置かれている [9]。また様々なプラグインによって機能拡張が可能である。

2.6.1 Protege4.3の使い方

Protegeはクラスとインスタンス、プロパティをそれぞれ設定し、知識ベースを作成することができる。また、プロパティはオブジェクトプロパティ、データプロパティ、アノテーションプロパティをそれぞれ設定することができる。

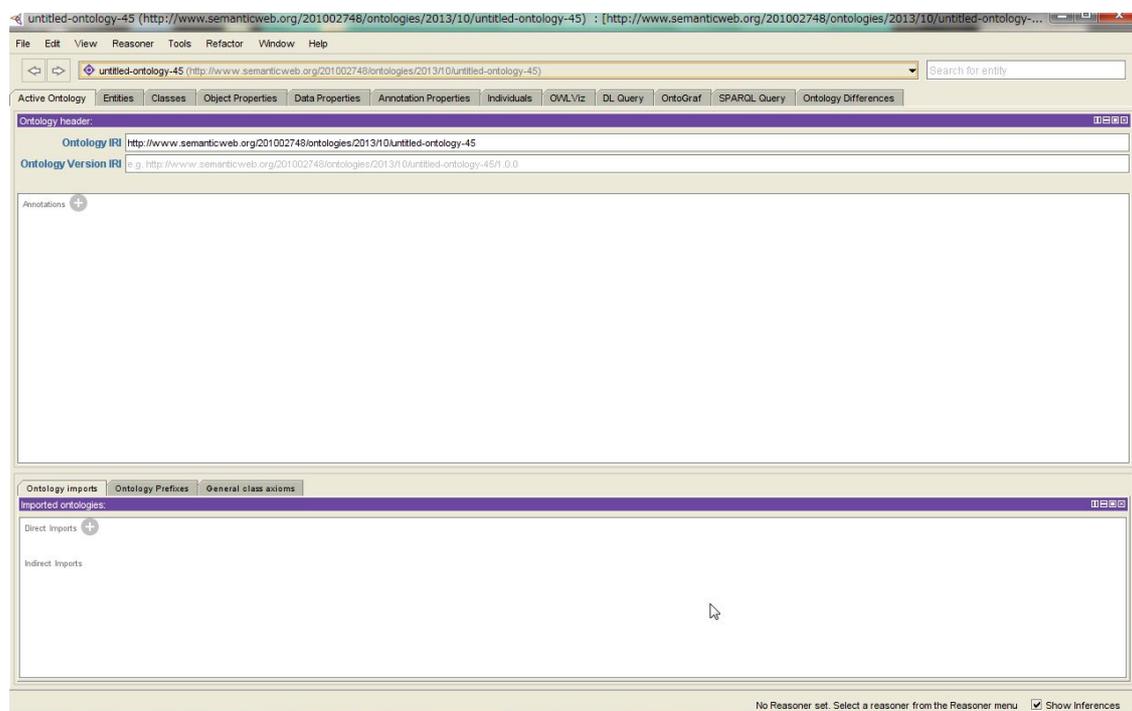


図 2.7: Protege4.3 起動画面

クラス編集画面では、クラス名を定義する。定義するクラスは下位クラス、兄弟クラスを選択することができる。兄弟クラスは同じレベルのクラスを追加する際に用いる。また、すべてのクラスは「Thing」クラスの下位クラスとなる。

クラス名を修正する場合、右クリックでは修正できない。まず、修正したいクラスを選択し、メニューにある「Refactor」から「Rename entity…」を選ぶ必要がある。

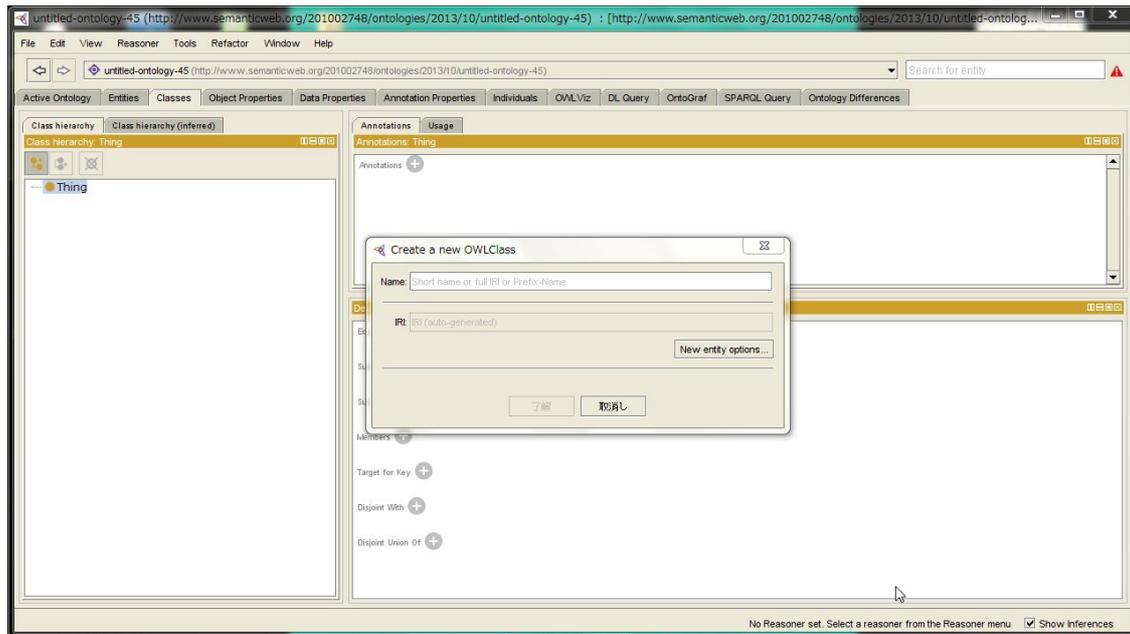


図 2.8: Protege4.3 クラス編集画面

オブジェクトプロパティ編集画面では、オブジェクトプロパティを作成し、定義域と値域を定義することができる。

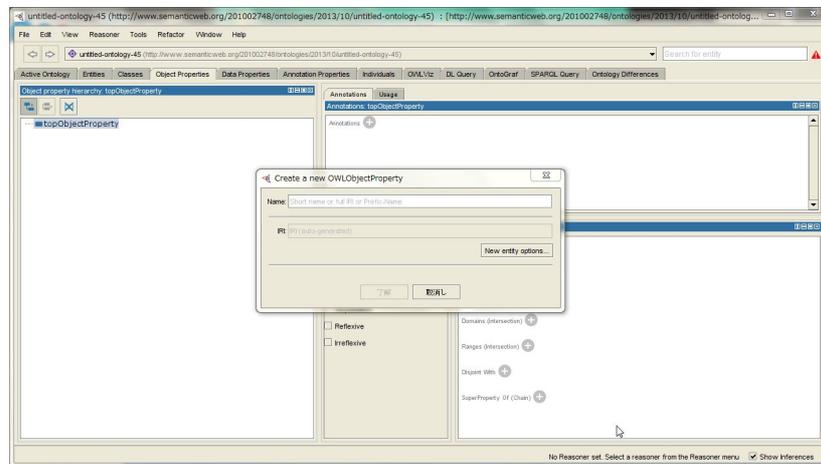


図 2.9: Protege4.3 オブジェクトプロパティ編集画面

データプロパティ編集画面でも、同様にデータプロパティを定義し、定義域と値域を定義することができる。

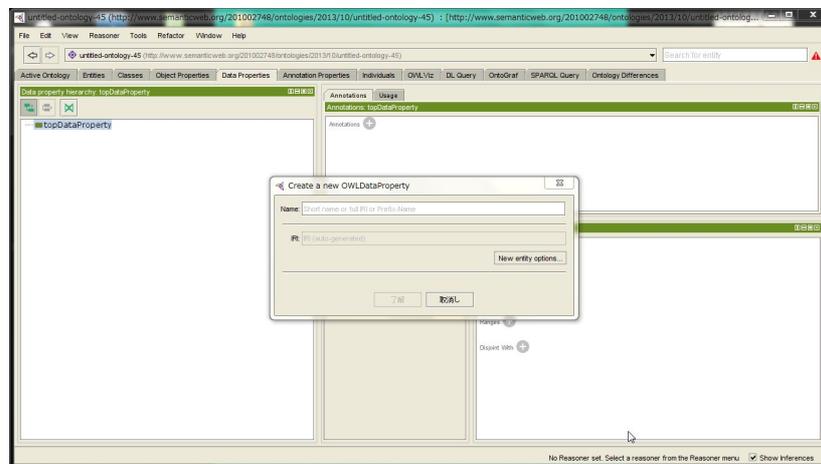


図 2.10: Protege4.3 データプロパティ編集画面

第3章 提案

3.1 提案モデル

本稿では、Protege 4.3 を用いて、民具に新たな関係性と検索キーワードの増加只を行い、見町エコミュージアムの検索システムの向上を目的とし、民具オントロジーの構築を目指すものである。

オントロジー間のクラス定義は、只見町エコミュージアムの分類に沿って行う。また、本来大分類には、生活用具の他に生産用具など8項目が存在するが、今回は簡略化のために生活用具に絞って民具 Ontology の構築を行う。

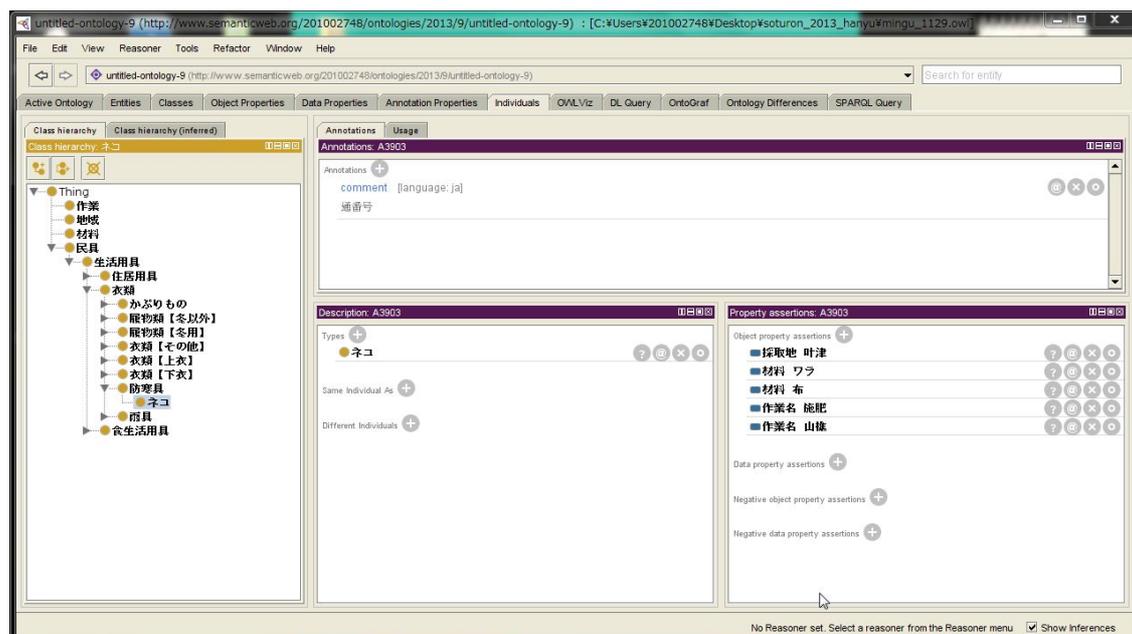


図 3.1: 民具 Ontology

民具 Ontology を可視化することで、その構造を視覚的に捉えられるようにする。図 3.2 は民具 Ontology の一部を可視化したものである。これにより通番号 A3903 という民具はネコという名称の防寒具で、布とワラを材料とし、叶津という地域で山樵や施肥といった作業に使用されていたことがわかる。

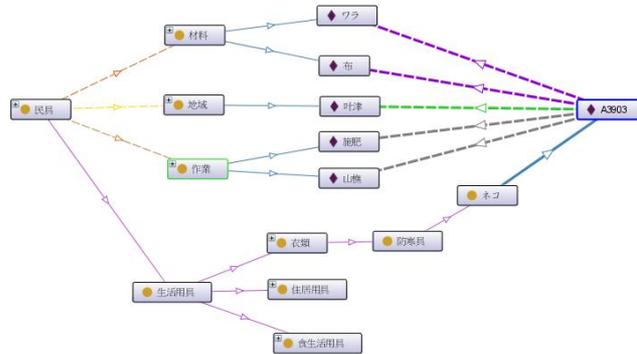


図 3.2: 可視化した民具 Ontology

参考文献

- [1] Vishal Jain、Dr. Mayank Singh:”Ontology Development and Query Retrieval using Protg Tool” (2013)
- [2] 小林夏生、木下宏揚:”只見町インターネット・エコミュージアムの「キーワード」検索の改善” 神奈川大学 2012 年度卒業論文
- [3] 北河祐作、古崎晃司:”大規模オントロジーの知的探索に向けた多段階展開型概念検索システムの開発” 人工知能学会研究会 (2013)
- [4] 尾暮拓也、古田一雄:”オントロジーを利用した分野特化型情報検索技術の社会的実装” 社会技術研究論文集 Vol.5 (2008)
- [5] 橘川俊忠:”「非文字資料の体系化」についての理論的諸問題” 神奈川大学 21 世紀 COE プログラム研究推進会議 (2008)
- [6] 小松大介:”福島県南会津郡只見町の民具のデータベース化とその問題点” 神奈川大学 21 世紀 COE プログラム研究推進会議 (2008)
- [7] 福島県只見町公式ホームページ
<http://www.tadami.gr.jp/>
- [8] 溝口理一郎:”オントロジー構築入門”
オーム社 2006 年 9 月 20 日 第一版代一刷発行
- [9] 溝口理一郎: 知の科学オントロジー工学
オーム社 2005 年 1 月 20 日 第一版第一刷発行

質疑応答

Q.

A.