

## 共同研究

### 非文字資料の効率的な検索と安全な流通

#### 非文字資料の効率的な検索と安全な流通

木下宏揚 佐野賢治

能登正人 松澤和光 宮田純子

小松大介 鈴木一弘

KINOSHITA Hirotsugu SANO Kenji

NOTO Masato MATSUZAWA Kazumitsu MIYATA Sumiko

KOMATSU Daisuke SUZUKI Kazuhiro

## 1 研究目的

本共同研究は、非文字資料を研究者間および専門家以外の人との間で情報の提供、共有などを行うために必要な基盤技術を構築し、実際の資料や研究者などを対象とした実証システムにより、その有効性を検証することを目的とする。

上記の目的達成に必要な基盤技術の提案を行い、その後、只見民具カードを対象に基本的なシステムを構築する。

具体的な研究目的は以下の項目からなる。

[目的 1] 非文字資料に特化した Ontology を構築し精度の高い検索と新しい知見のマイニングを行うシステムを構築する。

[目的 2] 非文字資料の Ontology 構築、検索などに適したユーザインタフェースを構築する。

[目的 3] 非文字資料の検索、流通時に個人情報や機密情報を保護し、著作権の調停を自律的に行う流通システムを構築する。

[目的 4] 非文字資料の資料の作成、データ処理、資料の流通などを円滑に行うために地域通貨的決済手法を提案する。

本稿では、主な研究成果のうち、以下の項目について報告する。

### 1. 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 [目的 1]

専門的な分野について情報検索する場合、専門的な知識を持たなければ関連性が深い項目を見逃すなど有効な情報検索が行えない。只見町インターネット・エコミュージアムでは、民具の公開を行っている。しかしながら、そこにある検索機能はユーザの意図した検索キーから正確な情報を出力しないという問題点がある。そこで、非文字資料の Ontology に Web Ontology 言語 OWL を用いることで、明示されていない関係を導出する。それにより、新たな関係を発見することができ、本当に必要

な情報を的確に検索することが可能となる。

## 2. 自己組織化可能な群知能を用いた情報リソースの管理 [目的 2]

ビッグデータやクラウドで知られている様々な情報データは、ファイルの持つ複数の属性により、管理が複雑になる。また、現在は木構造がファイル管理の主流であるが、直感的にファイルの位置を把握することが難しい。一方で、多数のファイルであっても、類似性の高いファイルごとに大きく分類されていれば、直観的にユーザがデータを扱えると考えられる。そこで本研究では、各データを、自己組織化可能な群知能データとみなし、それらの群知能データを制御することで、視覚的にデータを管理可能にする。提案管理手法は、データ管理の過程に応じて自己組織化する手法にも適用可能である。

## 3. ACO を用いた検索過程を重視した検索手法 [目的 2]

インターネットで検索をするときに、比較的簡単に一定レベルの検索結果は得ることができる。しかし、検索の質は常に十分とは限らない。もしその情報について更に深く理解しようとするれば、関連した情報についても調査する必要がある。そのとき、その分野に詳しい人の検索結果やその過程を参考にできれば、より詳細で多くの情報を検索過程から得ることができる。本稿では、検索過程の質を考慮した検索手法を提案する。群知能の ACO を用いた手法と「推薦システム」と呼ばれるシステムを用いた手法の二つを提案する。

## 4. 価値交換システムにおけるゲーム理論的解析 [目的 4]

近年、地域のコミュニティで物やサービスを循環させる目的で地域通貨が注目されているが、効率よく取引相手を発見したり、同じコミュニティ内でも価値観が異なると流通しない可能性がある。取引をもっと円滑にするために、地域通貨の「独自に価値を決める」という性質を活かし、多様な価値を表現可能な価値の交換システムを提案する。ここでいう価値とは、単なる金銭的な価値ではなく、地域通貨的に多様な価値を提案し、サービスに対して地域通貨的価値を付与した上で価値の交換システムを作ることを目的としている。これまでに、異なる価値観を持つ二者間の価値交換システムについて検討がなされてきた。しかし、これまで提案されてきた異なる価値観を持つ二者間の価値交換システムでは、各ユーザが満足する効用が得られるかどうかは未解決であり、複数人の場合にそのまま適用することができない。そこで、本稿ではゲーム理論を用いて特定の条件下で、 $n$  人の各ユーザにおける効用が満足するようなモデルを提案する。

## 5. ベジエ曲線を用いた電子透かしの提案 [目的 3]

現在、パソコンおよびインターネットの普及、小型端末の登場により、ネットからダウンロードして、パソコンや携帯端末などで読むことのできる電子書籍の市場が広がり始めている。小説などの文章だけでなく、漫画コンテンツなども電子書籍化が始まっている。また、ネット上には様々な種類の画像投稿サイトも存在しており、アマチュアの絵描きも自由な投稿が行われている。それに付け加え、漫画を書く際にアナログではなく、パソコン上で原稿を作成する漫画家も増え始めた。しかし、デジタルデータは複写、保存、加工が容易に行えるという特徴を持っており、ペイントソフトを使えばトレースなども容易に行うことが出来る。そのために、デジタルデータのコンテンツが著作権者の意思に反して利用されることが起こりうる。このような問題を解決するために著作権保護技術として電子透かし等の技術が存在している。しかしトレース画像において主に使用されている二値の線画像

についての有効な手法はあまり検討されていない。漫画など主に二値の線により構成されている画像に対して、画素数に対する情報量が少ないので品質を保ったまま必要な情報を埋め込むことは困難である。また、画素値に情報を埋め込む手法ではトレース等の手書きコピーに対して、情報の復元をすることが出来ない。そこで本稿では、二値の線画像の著作権管理に関して、ベジェ曲線を用いて、画素値ではなく形に情報を埋め込む電子透かしの手法を提案する。

## 6. 群知能を適用したアクセス制御システム [目的 3]

情報創発の支援とともに、その「行為を守る」システムを提示することである。そのためにアクセス制御を「言語ゲーム」とみなす。そして、群れの概念と遺伝の概念によって変動する環境に適應する群知能を提案する。群知能の構成要素「エージェント」はアクセス行列の中で定義される行為の連鎖 (Path channel) である。Path channel オブジェクトが群れを作るとき、「似ている行為」という概念、即ち「家族的類似」という概念を導入する。提示するマルチエージェントシステムにより、自己の行為に近い群れが、群れ全体として創発活動を支援し、かつ情報漏洩を一定量に維持調和させる新たなシステムの可能性が期待される。

## 2 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

### 1. まえがき

コンピュータが Web などのネットワーク上に存在する情報を蓄積、表示、分類する際は、情報を単にデータとして扱うだけで、情報の意味の理解を要するような処理をすることは困難である。このような高次元の処理を行うために、情報が表す内容を知識として扱う必要性が高まってきている。<sup>[1]</sup>

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、民俗学研究資料の情報共有・情報流通を目指していた。<sup>[4], [8]</sup> 非文字資料とは、文字媒体として記録されることなく受け継がれてきた民俗文化を対象とする民俗学研究資料である。今までの文化研究では文字に記録された事象に専ら関心が集中してきた。しかし、文字に表現されない人間の観念・知識・行為ははるかに幅広く、質量ともに大きい。それは文字で表現された事象とは比較にならない。

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」は、これらの中から (1) 図像、(2) 身体技法、(3) 環境・景観の 3 つに絞って、それぞれの事象について資料化する方法を開発し、その結果として資料を蓄積し、蓄積した資料を分析して発信することを目的としていた。本研究資料は、民俗文化をベースとしていることから、同じものを指し示す場合でも、地域や年代によって相違が生じる。そのため、非文字資料の情報共有・情報流通には情報資源に関する情報、すなわち、メタデータを用いた意味情報検索が求められる。現在の Web では、Web 上に散在している情報資源を表現するメタデータを利用した高度な処理を行うことはできない。

一方、Semantic Web では、情報資源間の関係を構造的に表現したメタデータから、意味的な検索や推論・演算といった知的な処理を提供することができる。したがって意味情報検索を主体とする非文字資料の情報共有・情報流通に適すると考える。

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、福島県只見町に古くから伝わる民具についての情報が実測されて記載された民具資料カードについて研究を



図1 只見町インターネット・エコミュージアム



図2 検索の失敗例（服）



図3 検索の失敗例（仕事着）

行ってきた。この成果は福島県只見町において、図1に示すように Web 上で「只見町インターネット・エコミュージアム」として、一般ユーザーや民具の研究者を対象に公開されている。しかし、現在実装されている検索機能では、ユーザーの指定した検索キーから必ずしも意図した結果を出力しないという問題点がある。例えば、図2、図3に示すように「服」「仕事着」といった、一般的な検索キーワードを入力しても検索結果が得られないことがある。

本研究は、非文字資料研究センター第二期共同研究課題「非文字資料の効率的な検索と安全な流通」の一環として、非文字資料用に作成した Ontology に Web Ontology 言語 OWL を適用することで、只見町インターネット・エコミュージアムの「キーワード」検索を改善する。

近年、Ontology を用いた研究では次のようなことが行われている。総合モデルのソフトウェアによる実現の試み、総合モデリングツールと総合モデル閲覧ツールおよび OntoGearServer の試作を用いた SOFC システムの設計支援への応用の<sup>[18]</sup>研究、Ontology 構築の「ルール」と「関係」に関する基礎的な考察から得た知見に基づいて設計・開発を行った Ontology 記述環境の研究、自

然言語解析とリンク構造解析を利用することで、キーワード検索でインターネット上の評価をもとに検索結果を出力する研究などが行われている。<sup>[19]</sup>

しかし、非文字資料は個人的に管理・保管しているものが多く、また、特有の性質を持つ資料のため上記の研究内容では、民具の検索機能には活用できないと考えられる。

本研究では福島県只見町に古くから伝わる民具についての情報が実測されて記載された民具資料カードを用いた「Ontology を用いた民具のデータベース化」の研究で、非文字資料の Ontology を構築し、研究者に対し意義のある新たな知見の提示が可能なることを示した。そして、非文字資料の Ontology に Jena を用いた RDF の推論を導入することで、明示されていない関係を導出する。それにより、新たな関係を発見することができ非文字資料の Ontology の有意性を実証した。

しかし、先行研究では RDF データや推論のルールを手動で記述する必要があったため、構築時や修正時に時間と手間が掛かるなどの問題点があった。そこで、提案手法では Ontology エディタの法造を用いることにより、プログラムの自動構築による修正の手間の改善手法を提案する。

まず、民具を Ontology 化するプログラムの作成をする。次に、キーワードにシソーラスを用いる

プログラムの作成をする。そして、この二つのプログラムを用いた「キーワード」検索のプログラムを作ることを提案する。それにより、新たな関係を発見することができ、本当に必要な情報を的確に検索することが可能となる。

本研究では、民具の Ontology を生成するために、法造と OWL を用いて、民具間の関係を推論する。これにより明示されていない関係を導出し、柔軟な検索キーワードの指定が可能となる。

## 2. 基礎知識

### 1. 非文字資料と民具

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し、工夫して編み出し使用してきた古風な器具や造形物の総称である。民具、民具同士の関連性を知ることにより当時の人間の営みや生活を知ることが可能になる。

### 2. 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記録するという点で、学術的な研究対象としても評価が高く、「只見町方式」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」という形で、2333 点の民具が国指定重要有形民俗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。このシステムは図 1 に示すように、「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、また、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。<sup>[10]</sup>

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上でも、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 4、図 5 に民具カードの一例を示す。<sup>[9]</sup>

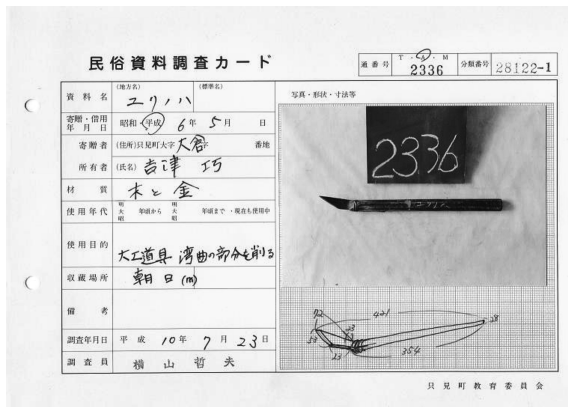


図4 ユリノハの民具カード1

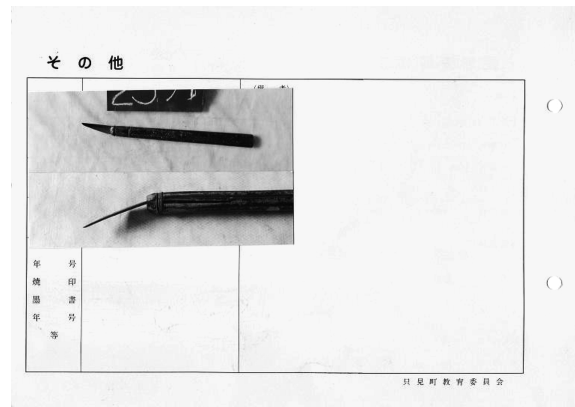


図5 ユリノハの民具カード2

### 3. Ontology

Ontology とは本来哲学用語であり、「存在に関する体系的な理論（存在論）」という意味である。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 とは共通語彙（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。

Semantic Web においては、Web 上で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。<sup>[3]</sup>

今 Ontology は概念と意味を処理する Ontology 工学として、Semantic Web、人工知能、自然言語処理、人間工学などの情報科学を貫く原理として注目されている。<sup>[6]</sup>

Ontology はコンピュータという道具を使い、人間の知識の構造を明らかにする。<sup>[4]</sup>

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。<sup>[11]</sup>

その結果、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。<sup>[3], [5]</sup>

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6)

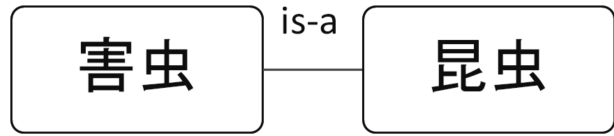


図6 is-a 関係構造図

- 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7)

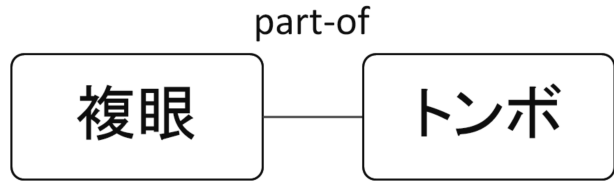


図7 part-of 関係構造図

- 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8)



図8 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9)

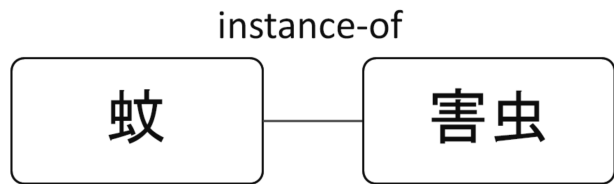


図9 instance-of 関係構造図

図10に民具カードのデータの Ontology の一例を示す。

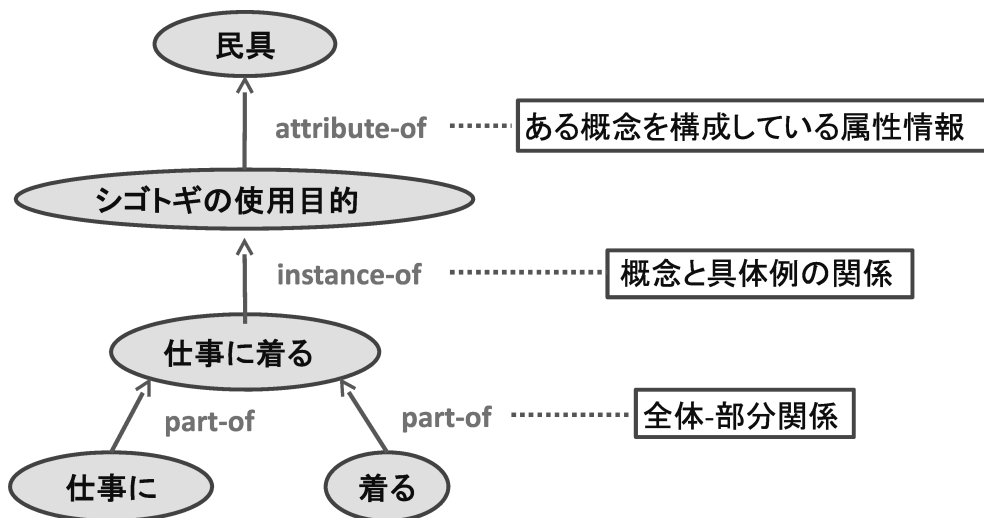


図10 Ontology

#### 4. 民具カードの基本項目の Ontology

民具カードは以下に示す三つの基本的な Context 情報から成立している。

- 民具の性質に関するもの（寸法）
- 分類・整理に関するもの（番号）
- 民具の用途に関するもの（目的・方法）

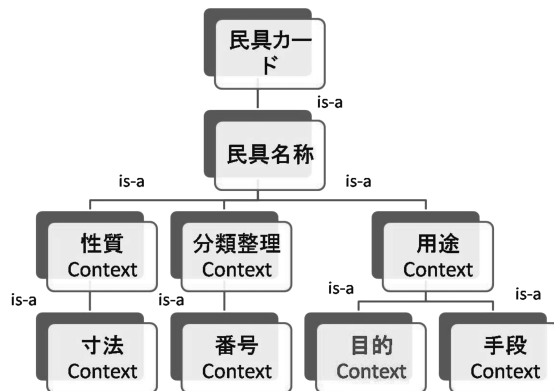


図 11 基本的な Context 情報

#### 5. 法造 — Ontology エディタ

「法造 — Ontology エディタ」は、Ontology の基礎理論に関する考察に基づいて設計された Ontology 構成内容の閲覧・編集をする記述環境である。本ツールでは上記の機能に加えて、大規模 Ontology 開発に向けて複数の Ontology を扱うためのプロジェクト管理機能、および複数のユーザがネットワーク上で Ontology を共有するための Ontology 分散管理機能を実現した。「法造」は、Ontology (=“法”) を構築する (=“造”) ための計算機環境で、「Ontology エディタ」、「概念工房」(Ontology 構築ガイドシステム)、「Ontology サーバ」、そして「Ontology マネージャ」の四つのシステムから構成されている。<sup>[13], [14], [15]</sup>

##### 1. 法造の使い方

メニューバーの [ファイル] — [新規作成…] を選択すると、メイン画面に新しい編集パネルが表示される。

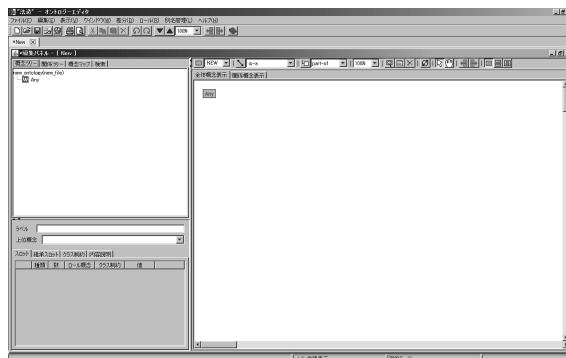


図 12 Ontology エディタ起動画面

概念を追加する。

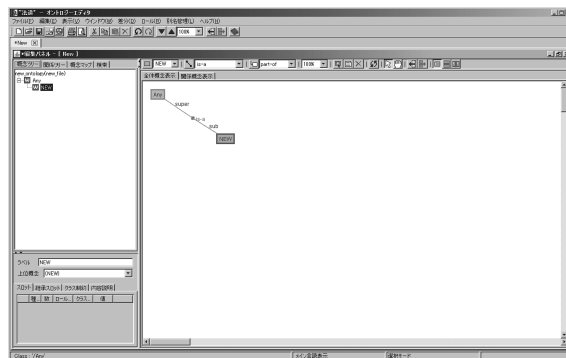


図 13 概念を追加



スロットを追加する。

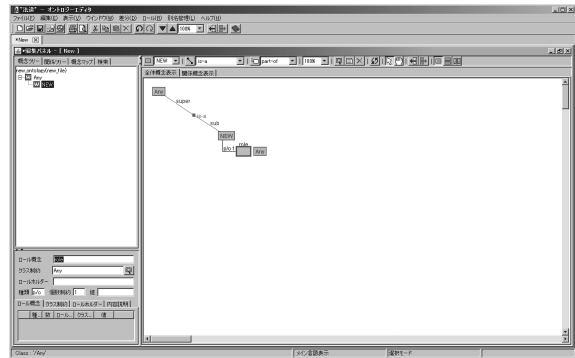


図 14 スロットを追加

クラス編成と Ontology が可視できる。

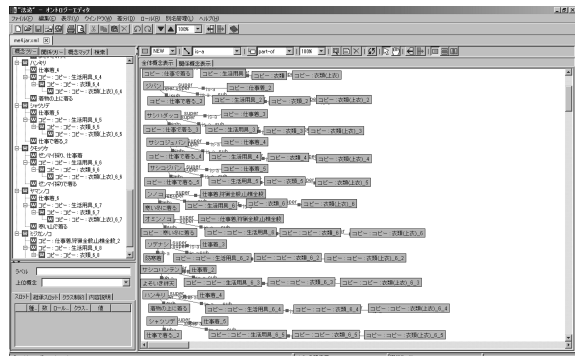


図 15 入力結果

法造で作成した Ontology を OWL に変換する。



図 16 Ontology の OWL 化

## 6. Web Ontology 言語 OWL

OWL は W3C における Semantic Web 活動の一環として、Web Ontology 作業グループによって作成された Web Ontology 言語である。

OWL は Web Ontology 言語という名が示すように、Web での使用を前提としている。スタンドアロンでの使用ができないわけではないが、Web 上での使用により適した言語使用になっている。<sup>[2], [7]</sup>

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:owl="http://www.w3.org/2002/07/owl#"
5   xmlns:owl1="http://www.w3.org/2002/07/owl# 1"
6   xmlns:owl2="http://www.w3.org/2002/07/rdf-schema#"
7   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
8   xml:base="http://www.hoco.jp/owl/SHIGOTOGI.owl">
9
10 <owl:Ontology rdf:about="">
11   <rdf:comment?>
12     H2O:OWL Export
13   </rdf:comment?>
14 </owl:Ontology?>
15
16 <owl:Class rdf:ID="RelationalConcept">
17   <rdf:label?>RelationalConcept</rdf:label?>
18 </owl:Class?>
19 <owl:ObjectProperty rdf:ID="hasPart">
20   <rdf:label?>hasPart</rdf:label?>
21 </owl:ObjectProperty?>
22 <owl:ObjectProperty rdf:ID="hasAttribute">
23   <rdf:label?>hasAttribute</rdf:label?>
24 </owl:ObjectProperty?>
25 <owl:Class rdf:ID="SHIGOTOGI_SHIGOTOGI">
26   <rdf:label?>SHIGOTOGI_SHIGOTOGI</rdf:label?>
27 </owl:Class?>
28 <owl:Class rdf:ID="SHIGOTOGI_SHIGOTOGI1">
29   <rdf:label?>SHIGOTOGI_SHIGOTOGI1</rdf:label?>
30 </owl:Class?>
31 <owl:Class rdf:ID="SHIGOTOGI_SHIGOTOGI2">
32   <rdf:label?>SHIGOTOGI_SHIGOTOGI2</rdf:label?>
33   <rdf:subClassOf?>SHIGOTOGI_SHIGOTOGI />
34 </owl:Class?>
35 <owl:Class rdf:ID="SHIGOTOGI_SHIGOTOGI3">
36   <rdf:label?>SHIGOTOGI_SHIGOTOGI3</rdf:label?>
37   <rdf:subClassOf?>SHIGOTOGI_SHIGOTOGI />
38 </owl:Class?>
39 <owl:Class rdf:ID="SHIGOTOGI_SHIGOTOGI4">
40   <rdf:label?>SHIGOTOGI_SHIGOTOGI4</rdf:label?>
41 </owl:Class?>
42 <owl:Class rdf:ID="UndefinedClass">
43   <rdf:label?>UndefinedClass</rdf:label?>
44 </owl:Class?>

```

図 17 OWL

## 3. 提案手法

### 1. フレームワーク

提案手法のフレームワークを図 18 に示す。「只見町エコミュージアム Web インタフェース」は Ontology を可視化して Ontology の作成を容易にする。「Wikipedia Ontology」は Wikipedia が構築している日本語 Ontology で、アプリケーションプログラムから利用可能な API が用意されている。これを利用することで、民具の Ontology には記述されていない知識を補うことが可能となる。「OWL による記述」では、対象とするデータの詳細な分類や推論メカニズムを表現するためのツールとして動作する。「Jena と OWL-API を利用した解析エンジン」は推論規則を生成するための記述である。

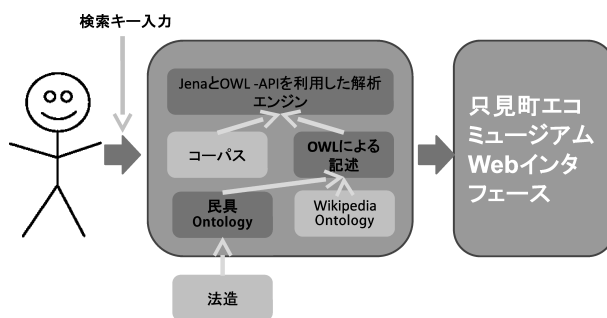


図 18 フレームワーク

ID	名前	...
831	只見町	...
832	只見町	...
833	只見町	...
834	只見町	...
835	只見町	...
836	只見町	...
837	只見町	...
838	只見町	...
839	只見町	...
840	只見町	...
841	只見町	...
842	只見町	...
843	只見町	...
844	只見町	...
845	只見町	...
846	只見町	...
847	只見町	...
848	只見町	...
849	只見町	...
850	只見町	...
851	只見町	...
852	只見町	...
853	只見町	...
854	只見町	...
855	只見町	...
856	只見町	...
857	只見町	...
858	只見町	...
859	只見町	...
860	只見町	...

図 19 只見カード Relational Database

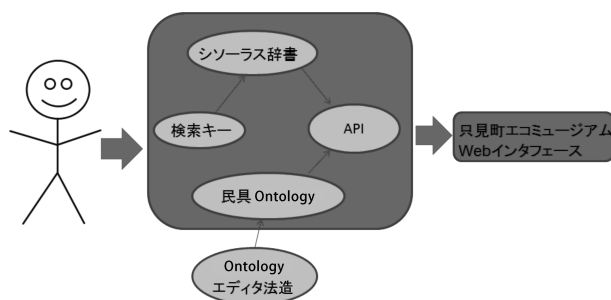


図 20 提案構造