論文

非文字資料を対象とした Ontology データベースに対する RDF 推論の適用

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　まえがき

近年における情報化社会の発達は，社会全体の情報化と情報利用の変化をもたらした．従来では限られた世界・領域において情報の供給者と消費者の関係が明白であったが，現在の情報流通においては，多方向性と他領域性が特徴としてあげられる．扱われる情報と利用されるメディアは多種多様となり，それらに対応した情報技術の発達は，ネットワーク上での様々な活動を可能にしている．情報の電子化技術とネットワークの発展・普及により，膨大な情報資源が電子化される一方，利用者の目的に応じて，必要な情報を獲得・利用するためには，情報処理技術の発達だけでは解決できない問題も起きている[1]．情報網拡大に伴い膨大な情報があふれる環境において，有効な情報検索を行うためには，高度な知識処理が必要とされる．

しかしながら，コンピュータは Web に存在する情報を蓄積・表示・分類したりするが，それらの情報を単にデータとして扱うだけで，情報が意味するものの理解を要するような処理をすることはできない．Web 上の情報を理解し利用するためには，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[2]．

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族学研究資料である．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．

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 上に散在され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．

そんな中，本研究室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[3]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ことを示した．しかし，具体的な推論機構については未解決であった．そこで，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 の有意性を実証する．

2　非文字資料と民具カード

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．

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，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[4]．

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1，2に民具カードの一例を示す．[5]

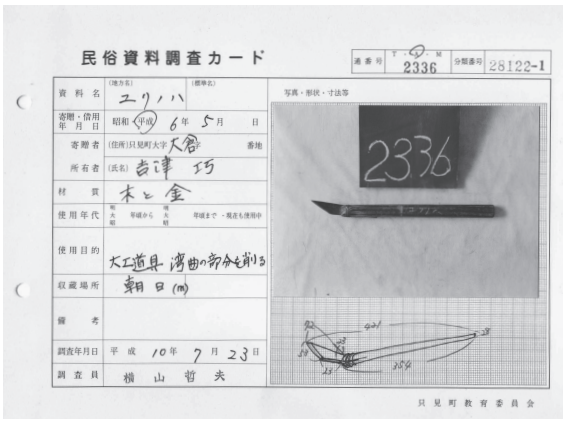

図1　ユリノハの民具カード1

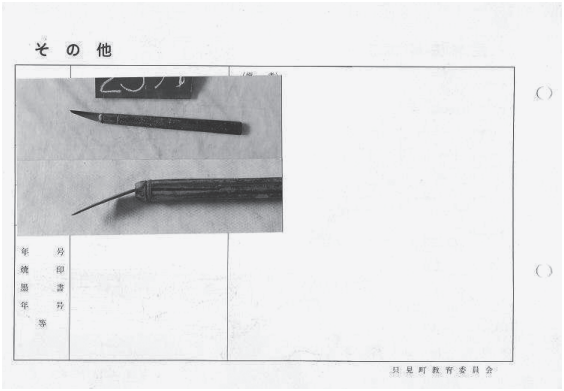

図2　ユリノハの民具カード2

3　Ontology

3.1　Ontologyの概要

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．[6], [7]

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．[8]

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．[9]

3.2　Ontologyの役割

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．[10]

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]．

3.3　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．
- 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．
- 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．
- 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．

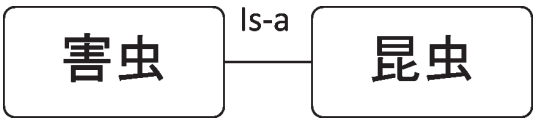

図 3　is-a 関係構造図

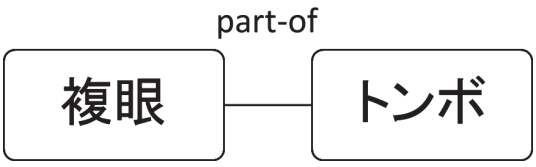

図 4　part-of 関係構造図

図 5　attribute-of 関係構造図

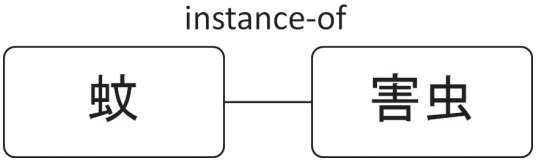

図 6　instance-of 関係構造図

3.4　基本項目のOntology

民具カードは以下に示す3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）

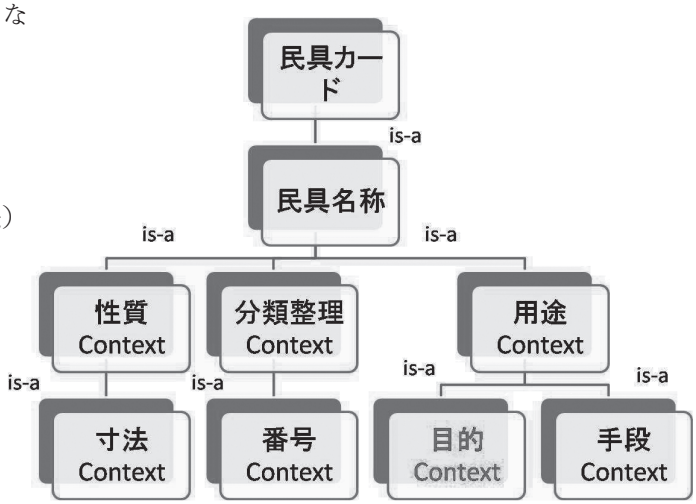

図7　Ontology構造図

4　RDF

4.1　RDFの概要

RDF（Resource Description Framework）は，WWW上で資源に関する情報を表すための言語である．タイトル，ウェブ・ページの更新日，ウェブ・ドキュメントの著作権及びライセンス情報，ある共有資源に対する利用可能スケジュールなどのようなウェブ資源に関するメタデータの表現を特に目的としている[11]．しかし，RDFは「ウェブ資源」の概念を一般化することによりウェブでは直接検索できないがウェブで識別できる事物に関する情報を表すために使用できる．例えば，オンラインショッピング機能で入手できるアイテムに関する情報（仕様，価格，入手可能性に関する情報など）や，情報発信に対するウェブ・ユーザの嗜好に関する記述が含まれる．

RDFは人間に表示するだけでなく，アプリケーションが情報を処理する必要のある状況を目的とする．この情報を表現するための共通の枠組みを提供するため，意味を損なわずにアプリケーション間で情報交換が行える．共通の枠組みであるためアプリケーションの設計者は共通のRDFパーサや処理ツールを有効利用できる．異なるアプリケーション間で情報交換できるということは，情報が元々作成された以外のアプリケーションでその情報を利用できることを意味する．

RDFはウェブ識別子（URI）を使用して事物を識別し，シンプルなプロパティとプロパティ値で資源を記述するという考えに基づいている．これにより資源を表すノードとアークのグラフや，そのプロパティと値として資源に関するシンプルなステートメントを提供できるようになる[12]．

4.2　RDFの推論

RDFの推論は，RDFのデータには直接含まれていない情報をOntologyやスキーマといった規則に従って発見する処理である．

5　Jena

5.1　Jena の概要

Jena とは，Java による Semantic Web アプリケーション開発のためのフレームワークである．Semantic Web はコンピュータにとって理解可能なウェブを構築しようという試みである．Jena は RDF で表されるデータ（知識）を処理し利用するための様々な機能を提供する．Jena の主な機能は次のようなものがある．[13]

・RDF データの読み込み，出力

・RDF モデルの編集，マージ，問い合わせ

・RDFS, OWL, DAML＋OIL などの Ontology の操作

・SQL データベースを利用した永続的な利用

・問い合わせ言語 RDQL による検索

・Ontology などのルールに基づいた推論，検証

5.2　Jena による推論

Jena には Ontology などのルールに基づく推論が実装されている．Jena の推論で最も重要なのは com. hp. hpl. jena. rdf. model. InfMod というモデルのサブインタフェースである．生成段階で推論が行われるため，InfModel のインスタンスは推論実行後のモデルを表す．推論によって情報の発見を行うため，これには推論の元となったモデルに含まれなっかたステートメントも含む．元のモデルに存在したステートメントと同様に操作したり，検索したりすることができる．[13]

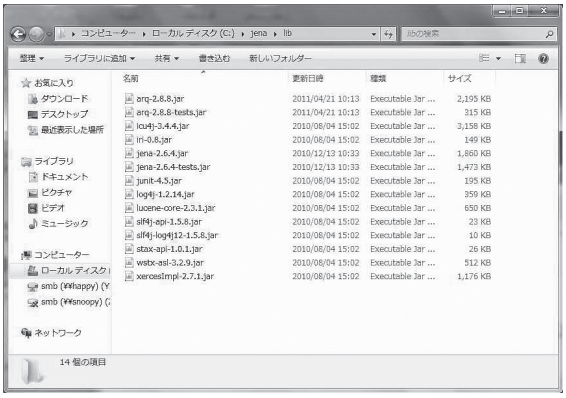

図 8　Jena の lib フォルダに入っている jar ファイル

5.3　Reasoner の役割

Reasoner インタフェースを実装したクラスはどのような規則に従って推論を行うかの情報が入れられる．実際に推論が行われるときには Reasoner の情報に従ってスキーマ，Ontology が解釈され新たな情報の発見が行われる．独自の規則を定義することもできるがいくつかの有用な Reasoner が実装されているのでそれを利用することができる．これらの Reasoner は Reasoner Registry クラスを利用することで取得することができる．

RDFSやOWLに関するReasonerは一般にOntologyやスキーマとともに使用する．ある特定のOntologyやスキーマについてバインドを行い，専門のReasonerを作ることもできる．それにはReasonerのbind-Schema（Model model）メソッドを使用する．引数にはOntologyやスキーマの情報の入ったモデルを渡す．注意が必要なのはbind-Schema（Model model）は新たにオブジェクトが生成されて戻されることである．[13]

5.4 Jenaの使い方

ダウンロードしたJenaのZIPファイルを解凍するとソースやAPIDocsなど様々なファイルが得られる．このうちlibというフォルダに14個のjarファイルが入っている．Jenaを利用する場合はこれらのjarファイルをクラスパスに含むようにする．含んだ状態で作成したデータをコンパイルし実行する．

Jenaを使用したRDFの推論を行うために，民具カードに記載されているデータをRDFとして記述する．

図11は推論に使用するRDFデータの1つである．7行目のPurposeOfUseは使用目的の意味である．8行目のrangeは目的語，9行目のdomainは主語を表す．これにより使用目的の主語はKIRI（錐），目的語はHole（穴をあける）ということを表すRDFデータとなる．

図12も推論に使用するRDFデータの1つである．7行目から9行目で「MITSUMEGIRI（ミツメギリ）はKIRI（錐）のLocalName（地方名）である.」を表し，11行目から13行目で「Hole（穴

図9 jarファイルをクラスパスに含みコンパイル

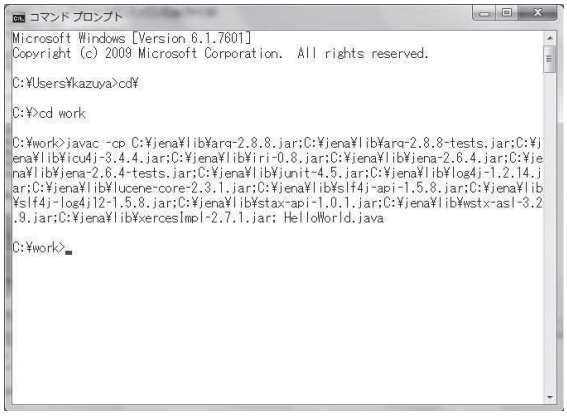

図10 jarファイルをクラスパスに含み実行

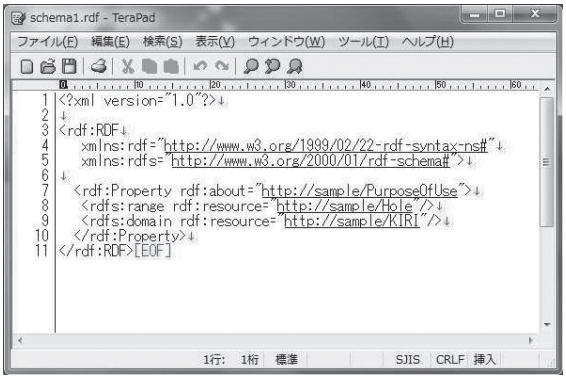

図11 推論に使用するRDFデータ1

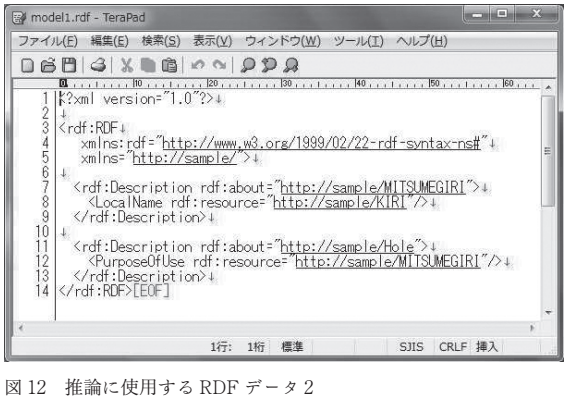

図12 推論に使用するRDFデータ2

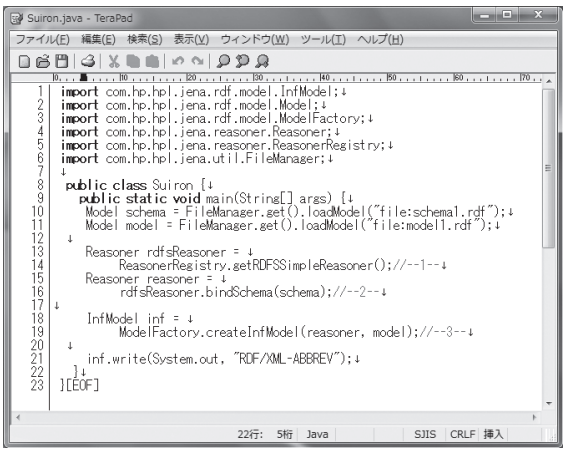

図 13　推論を行うためのプログラム

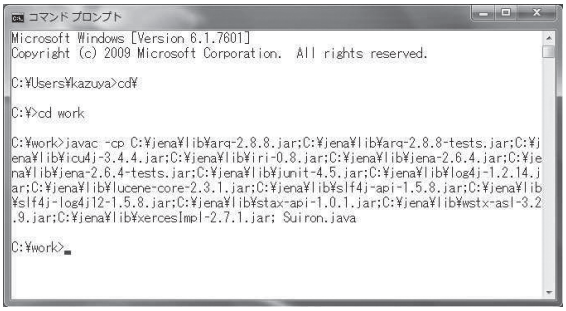

図 14　推論プログラムのコンパイル

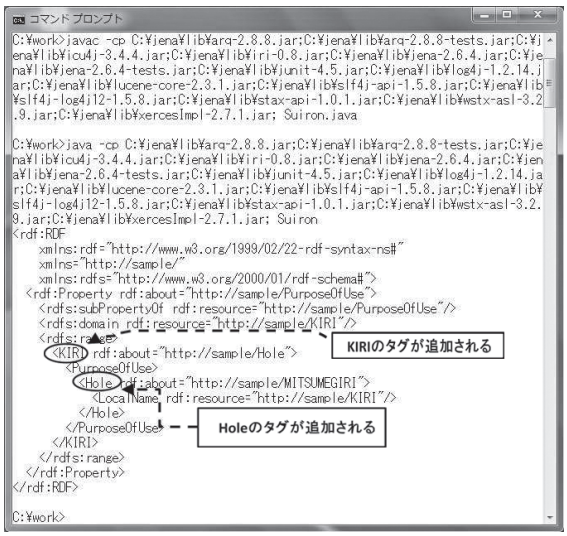

図 15　Jena による推論の実行結果

をあける）は MITSUMEGIRI（ミツメギリ）の PurposeOfUse（使用目的）である.」を表す RDF データとなる.

Jena を使用し図 11，12 のデータに対し推論を行うためのプログラムを java で記述する.

この推論プログラムの推論を行う基本的な流れは「13，14 行目の ReasonerRegistry. getRDFSSimpleReasoner（　）メソッドにより必要な Reasoner を取得する．15，16 行目で Resoner にスキーマのデータをバインドする．18，19 行目の ModelFactory. createInfModel（reasoner, model）メソッドにより InfModel オブジェクトを生成する」である.

この推論プログラムを図 11，12 のデータに対して行う.

Jena の lib というフォルダに入っている 14 個の jar ファイルをクラスパスに含むようにしてコンパイルする.

図 15 に示したような結果が得られる．図に書いてあるように Hole に KIRI のタグが追加され，MITSUMEGIRI に Hole のタグが追加される．このタグは元の図 11，12 の RDF データに直接含まれていない情報である．このことから推論を行ったことでタグが自動的に追加されたことがわかる.

6　むすび

本研究では，Jena を用いた非文字資料の Ontology の有意性を検証した．結果に表れたように民具カードに記載されていない情報をコンピュータが推論し導き出した．この記載されていない情報により民具間などの新たな関係が見えるようになった．しかし，民具カードに記載されている情報を

RDFデータとして記述したり推論のプログラムを記述する際，手動で行うのでシステムの構築に時間が掛かったり，記述のミスがあったとき修正に手間が掛かるなどの問題点があった．

今後の課題としては，Excelなどに記述されている情報を簡単に利用できるRDFデータ記述システムや推論システムの開発，Ontologyを構築する際に同時に推論を行い新たな関係も一緒に導出するシステムの開発があげられる．

参考文献

［1］ 内藤求：セマンティックWebコンファレンス2005-RDFとTopicMapsで実現するSearnless Knowledge

［2］ 木下慶子　村上敦志　稲積泰宏　木下宏揚　森住哲也：“Context間の関連性を表現するメタOntology――民俗学研究のための情報発信――”情報処理学会研究報告，じんもんこん研究会，pp. 1-6,（2006-1）Meta-Ontology that express the content to relation of between Context

［3］ 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/

［4］ 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 http://www.himoji.jp/jp/publication/pdf/seika/401/02-033-040.pdf

［5］ 福島県只見町公式ホームページ http://www.tadami.gr.jp/

［6］ 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

［7］ AIDOS：セマンティック技術シリーズ　オントロジ技術入門ウェブオントロジとOWL　東京電機大学出版局　2005年9月20日

［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　第一版第一刷発行

［9］ 齋藤孝：社会科学情報のオントロジ　中央大学出版部　2009年2月25日　第一版第一刷発行

［10］ 齋藤孝：意味論からの情報システム　中央大学　2006年4月15日　第一版第一刷発行

［11］ 神埼正英：セマンティック・ウェブのためのRDF/OWL入門　森北出版　2005年1月7日　第一版第一刷発行

［12］ RDF入門 http://www.asahi-net.or.jp/~ax2s-kmtn/internet/rdf/rdf-primer.html

［13］ Jena-TECHSCORE http://legacy.techscore.com/tech/Others/Jena/index.html