

Ontologyを用いた民具のデータベース化

研究の背景

- 非文字資料の体系化として神奈川大学21世紀COEプログラムが行われた。(2008年3月31日終了)
- 従来のデータベースでは民具間同士の関連性が分かりづらいといった欠点がある。(例: チョウナは「木を削る」。錐は「木に穴をあける」という目的があるが、この両者には「木を加工する」という共通点がある等)
- そこでデータベースの項目をOntologyで体系化、関連性を明確にし、只見民具カードを使ったデータベースの構築を提案する。

研究の目的



- Ontologyを導入することにより、文書が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。(キリで検索すると「ミツメギリ」「ヨツメギリ」などが検索されないが、Ontologyを使うことで検索可能となる。)
- 民具カードに記載された使用目的をOntologyを使って構造化し、他の民具との関連性を明確にする。
- 最終的には民具カードの内容をOntologyでデータベース化することを目指す。

目標とする検索例

- 先の例で挙げた「キリ」で検索してもミツメギリやヨツメギリが検索できる。
- 「穴をあける」といった共通項から歴史的背景を探ったり、新たな関連性の発見。
- 「木材を加工する」で検索するとき、大工道具のキリやチョウナだけでなく、関連道具として、印をつけるスミサシも検索される。

民俗資料調査カード



調査号 Y-19 M 725 分類番号 28116-1

資料名 (種別名) (標準名)	キリモエ	写真・形状・寸法等	 
寄贈・借出年月日	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	木 針金		
使用年代	昭和から 現在まで、現在も使用中		
使用目的	木の穴をあける道具		
収蔵場所	旧朝日公民館		
備考			
調査年月日	平成 2 年 6 月 29 日		
調査員	横山哲夫		

只見町教育委員会

民俗資料調査カード



調査号 Y-19 M 1411 分類番号 28116-2

資料名 (種別名) (標準名)	キリモエ	写真・形状・寸法等	 
寄贈・借出年月日	昭和・平成 59 年 6 月 9 日		
寄贈者 (住所) 只見町大字 石佐 番地			
所有者 (氏名)	振倉玄野		
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	木の穴あけの道具		
収蔵場所	旧朝日公民館		
備考			
調査年月日	平成 2 年 12 月 7 日		
調査員	横山哲夫		

只見町教育委員会

民俗資料調査カード

調査号 Y-19 M 2522 分類番号 28116-3


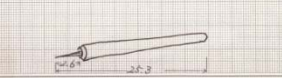
資料名 (種別名) (標準名)	ミンメギ	写真・形状・寸法等	 
寄贈・借出年月日	昭和・平成 10 年 7 月 10 日		
寄贈者 (住所) 只見町大字 岩井原 番地			
所有者 (氏名)	目黒唯夫		
材質	木と鉄		
使用年代	昭和から 現在まで、現在も使用中		
使用目的	物の穴を明す道具		
収蔵場所	朝日		
備考			
調査年月日	平成 10 年 7 月 2 日		
調査員	横山哲夫		

只見町教育委員会

錐(キリ)の民具カード

民俗資料調査カード

調査号 Y-19 M 2523 分類番号 28116-4

資料名 (種別名) (標準名)	ミンメギ	写真・形状・寸法等	 
寄贈・借出年月日	昭和・平成 10 年 7 月 10 日		
寄贈者 (住所) 只見町大字 岩井原 番地			
所有者 (氏名)	目黒唯夫		
材質	木と鉄		
使用年代	昭和から 現在まで、現在も使用中		
使用目的	物の穴を明す道具		
収蔵場所	朝日		
備考			
調査年月日	平成 10 年 7 月 2 日		
調査員	横山哲夫		

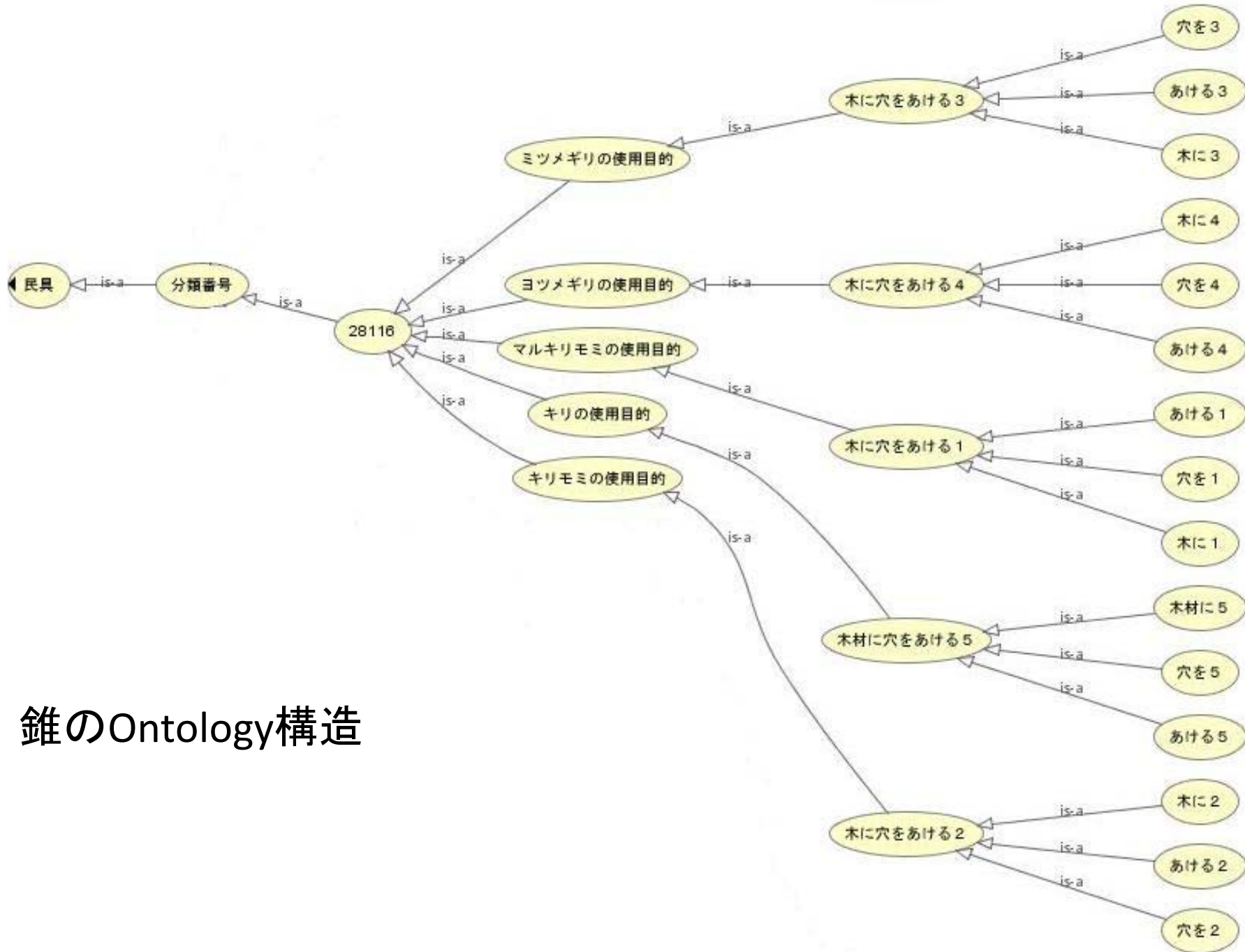
只見町教育委員会

民俗資料調査カード

調査号 Y-19 M 2788 分類番号 28116-5

資料名 (種別名) (標準名)	キリ	写真・形状・寸法等	 
寄贈・借出年月日	昭和 (年) 年 月 日		
寄贈者 (住所) 只見町大字 岩井原 字 番地			
所有者 (氏名)	長谷川及一		
材質			
使用年代	昭和から 現在まで、現在も使用中		
使用目的	木の板の穴をあける道具		
収蔵場所	旧朝日公民館		
備考			
調査年月日	平成 10 年 10 月 14 日		
調査員	横山哲夫		

只見町教育委員会



錐のOntology構造

研究内容

- 民具カードの内容をExcelを使い整理する。

研究内容

- 民具カードの内容をExcelに記述する。
- Excelをxmlに変換する。(タグが自由に決められるため)

研究内容

<?xml version="1.0" encoding="Shift_JIS" ?>

- <民具>

- <資料>

<ファイル名>tadami1-28116-1-</ファイル名>

<通番号>A725</通番号>

<分類番号>28116-1</分類番号>

<地方名>マルキリモミ</地方名>

<標準名>なし</標準名>

<寄贈借用>なし</寄贈借用>

<寄贈者所有者>なし</寄贈者所有者>

<氏名>なし</氏名>

<材質>木、金</材質>

<使用年代>なし</使用年代>

- <使用目的>

- <あけるを修飾>

<名詞助詞>木に</名詞助詞>

<名詞助詞>穴を</名詞助詞>

</あけるを修飾>

- <道具を修飾>

<動詞>あける</動詞>

</道具を修飾>

<名詞>道具</名詞>

</使用目的>

<収蔵場所>旧朝日公民館</収蔵場所>

<備考>鎌</備考>

<調査日>平成2年6月23日</調査日>

<調査員>横山哲夫</調査員>


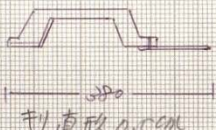
</資料>

- <資料>

<ファイル名>tadami1-28116-2-</ファイル名>

<通番号>A1641</通番号>

<分類番号>28116-2</分類番号>

民 地方名 カード		通番号	T・A・M	分類番号
資料名 (地方名) マルキリモミ (標準名)		725		28116-1
寄贈・借用年 月 日	昭和・平成 年 月 日	写真・形状・寸法等		
寄贈者 (住所) 只見町大字 字 番地	所有者 (氏名)			
材 質 木 針金	使用年代 明大 年頃から 明大 年頃まで			
使用目的	木に穴をあける道具			
収蔵場所	旧朝日公民館			
備 考				
調査年月日	平成 2年 6月 23日			
調査員	横山哲夫			

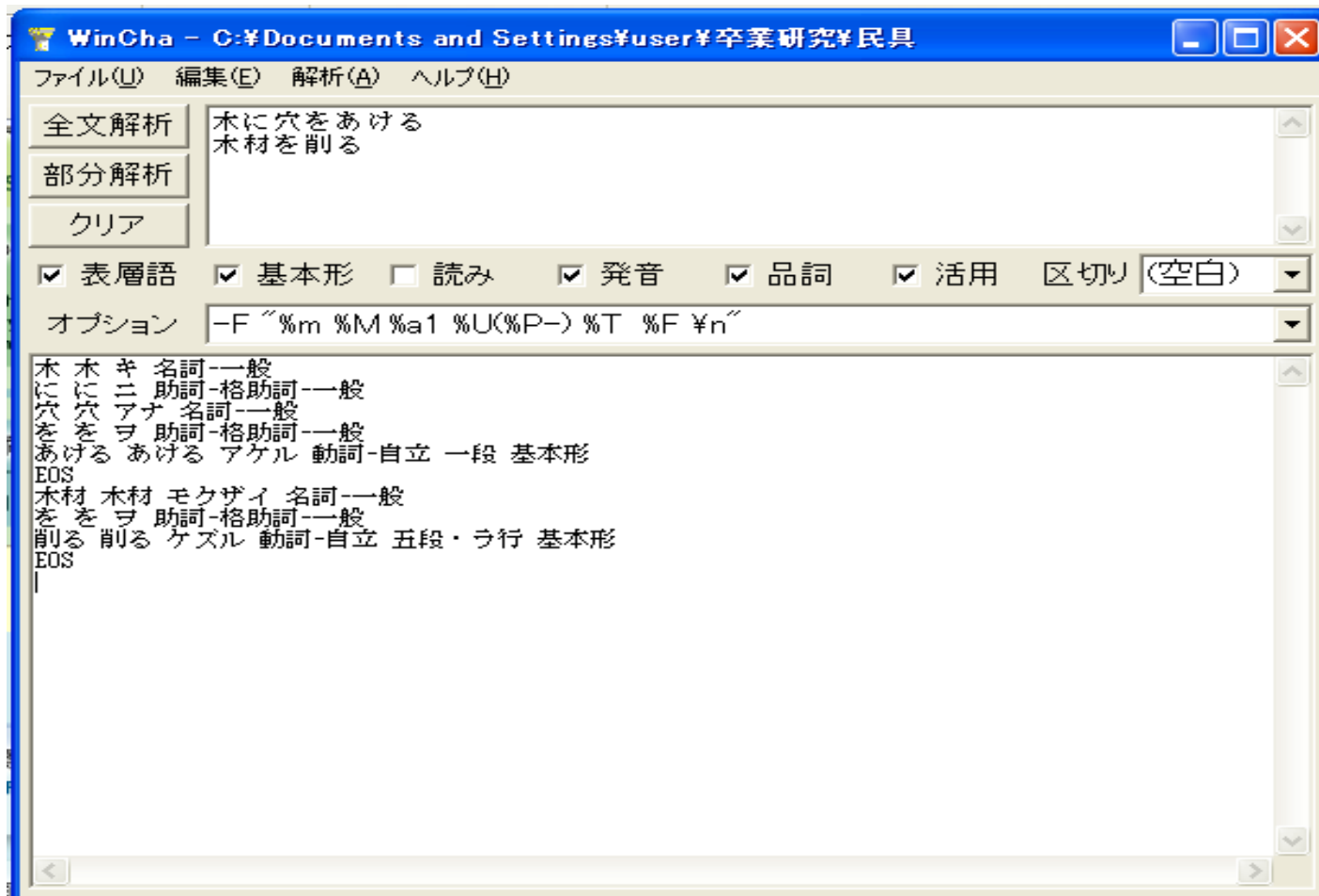
只見町教育委員会

前処理

- 民具カードの内容をExcelを使い整理する。
- Excelをxmlに変換する。(タグが自由に決められるため)
- 民具名等とは違い、使用目的や備考は自由に書かれているためchasen、cabochaといった形態素解析器を使い分解する。

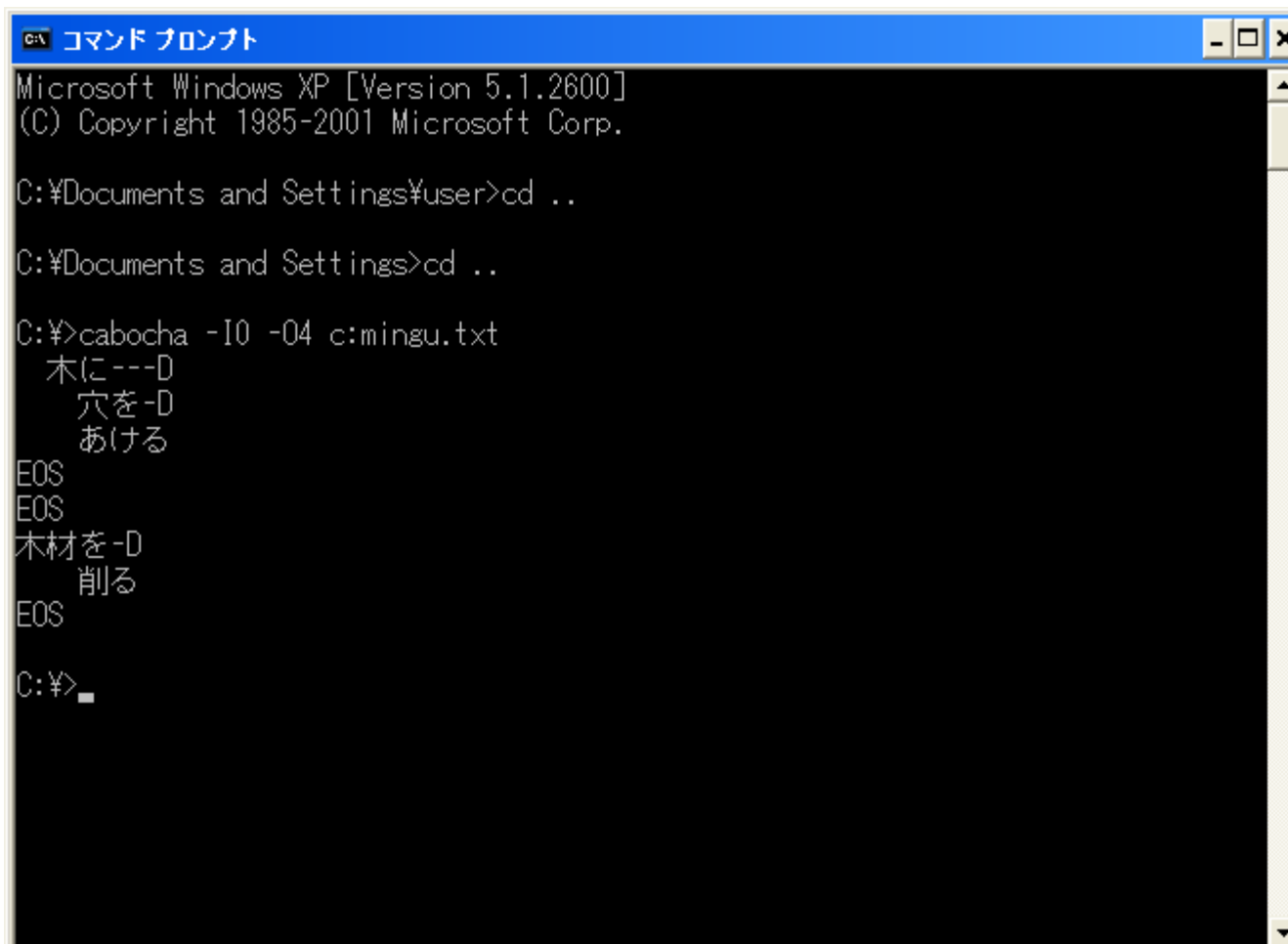
形態素解析

chasenを使い、使用目的「木に穴をあける」、「木材を削る」を形態素解析



統語解析

cabochaを使い、使用目的「木に穴をあける」、「木材を削る」を係り受け解析



```
コマンド プロンプト
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:¥Documents and Settings¥user>cd ..

C:¥Documents and Settings>cd ..

C:¥>cabocha -I0 -O4 c:mingu.txt
 木に---D
   穴を-D
   あける
EOS
EOS
木材を-D
 削る
EOS

C:¥> .
```

Ontologyの構成要素

Ontologyを構成する要素として以下の要素が挙げられる。

- 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。
- 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。
- 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。
- instance-of関係・・・概念とその具体例との間の関係を表す。

Ontologyの構築

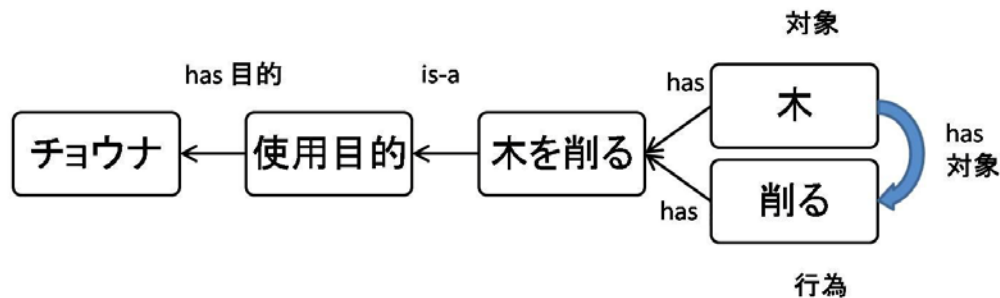
Protégéとは？

- Ontology言語であるOWLを書くための支援ツールである。
- 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。
- Protégéを使うことでOWLを容易に記述することができる。

Ontologyの構築

- 民具カードの使用目的をProtégéを使いOntologyで構築。
Protégéを使うことで言語を簡単に書くことができる。

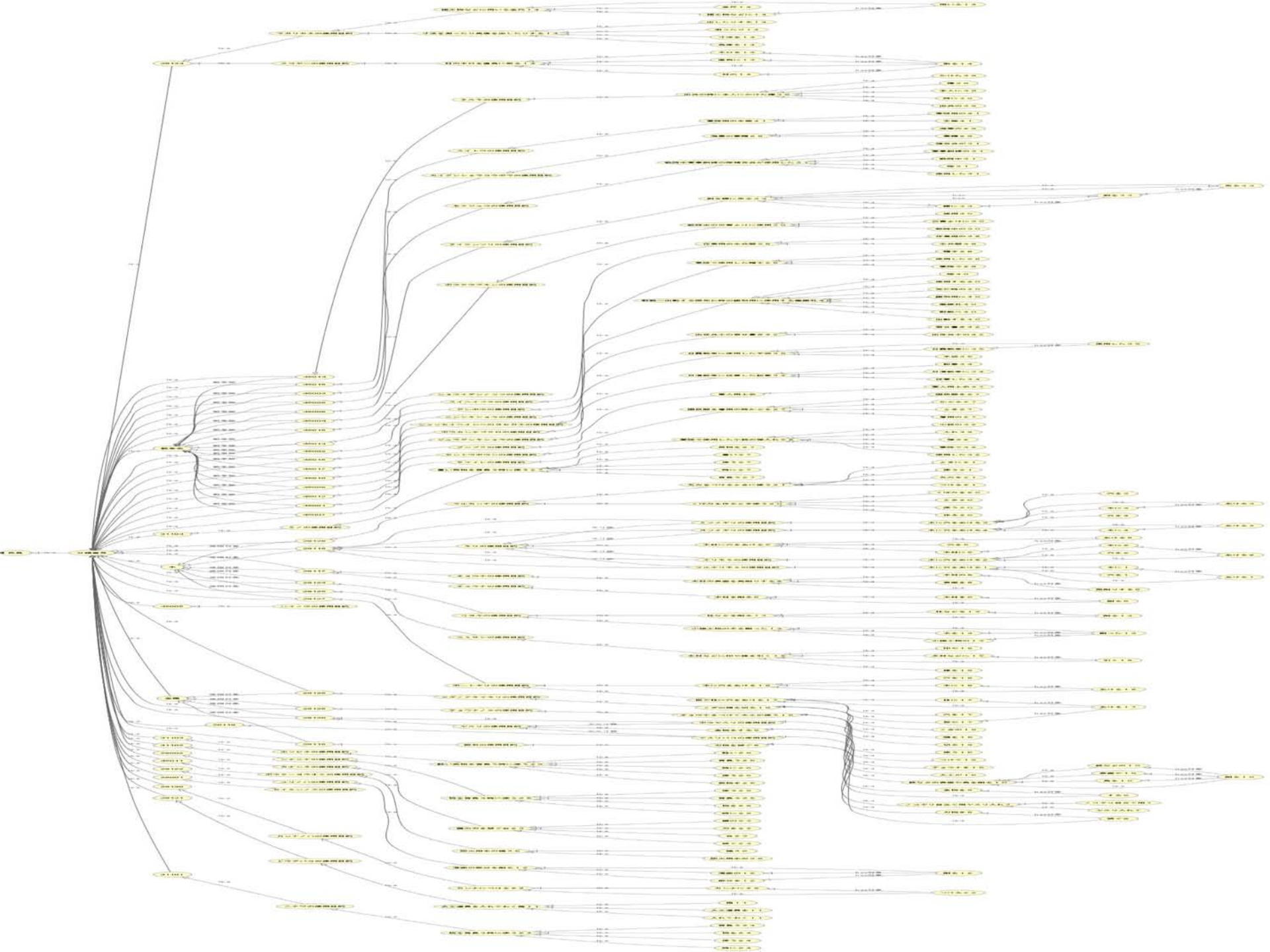
例: チョウナの使用目的「木を削る」をOntologyで分解



Protégéの実行内容

The screenshot displays the Protégé ontology editor interface. The window title is "OntologyDB.owl" with the URL "http://www.semanticweb.org/ontologies/2009/10/OntologyDB.owl". The menu bar includes File, Edit, Ontologies, Reasoner, Tools, Reflector, Tabs, View, Window, and Help. The toolbar shows navigation and search icons. The main interface is divided into several panels:

- Accepted class hierarchy:** Shows a tree view of classes. The selected class is "キリの使用目的" (Purpose of using Kiri), which is a subclass of "28116". It has several subclasses: "木材に穴をあける5" (Drilling holes in wood 5), "キリモミの使用目的" (Purpose of using Kiri-momi), and "マルキリモミの使用目的" (Purpose of using Maru-kiri-momi). Each of these has further subclasses.
- Class Annotations:** Shows the annotations for the selected class, currently empty.
- Description:** Shows the description for the selected class, currently empty.
- Object property hierarchy:** Shows a tree view of object properties. The selected property is "木工1" (Woodworking 1), which is a subclass of "has". It has several subclasses: "has対象" (has target), "キリ取" (Kiri-tori), "平スリ取" (Flat-slicing), "使用対象" (Usage target), "加工対象" (Processing target), and "製筆取" (Brush-making).



実験方法

- 従来のデータベースとOntologyデータベースとで検索を行い結果を比べる。

実験結果

Microsoft Excel 2010 画面のスクリーンショット。背景には「民具DB - Microsoft Excel」というブックが開かれ、表が表示されている。前景には「検索と置換」ダイアログボックスが開かれ、検索条件として「戦争」が設定されている。ダイアログボックスの下部には、検索結果のプレビューが表示されている。

検索と置換ダイアログボックスの検索結果プレビュー:

ブック	シート	名前	セル	値	数
民具DB.xlsx	Sheet1		\$K\$170	日清戦争に従軍した記事	
民具DB.xlsx	Sheet1		\$K\$176	日露戦争時に使用した手袋	
民具DB.xlsx	Sheet1		\$D\$204	日清戦争の掛け軸	

3セルが見つかりました

背景の表のデータ:

寄贈者所有者	氏名	材質	使用年代	使用目的		
なし	なし	木、針は金	なし	木に 穴を あける		
石伏	堀金主計	木、金	なし	木に 穴を あける		
荒井原	目黒唯夫	木、鉄	なし	物に 穴を あける		
荒井原	目黒唯夫	木、鉄	なし	物に穴をあける	朝日	なし
黒取	長谷川友一	なし	なし	木や板に穴をあける	旧朝日公民館	なし
只見字新屋敷	菅家次正	木、鉄	なし	木材に穴をあける	なし	刃の断面が正方形(四つ目)
只見字新屋敷	菅家次正	木、鉄	なし	手でもんで穴をあける大工用具	なし	なし
只見字新屋敷	菅家次正	木、鉄	なし	木材に穴をあける大工用具	なし	先端部分が折損している。又
只見字新屋敷	菅家次正	木、鉄	なし	木材に穴をあける大工用具	なし	刃先が三角形(三つ目)
只見字新屋敷	菅家次正	木、鉄	なし	木材に穴をあける大工用具	なし	刃先が三角形(三つ目)
黒告	菅家刃吾郎	金	なし	大工道具	旧朝日公民館	なし
只見字	新国茂之	木、鉄	なし	木材の表面を荒削りする道具	旧朝日公民館	なし
なし	なし	なし	なし	なし	なし	なし
なし	なし	鉄	なし	気を割る大工用具	朝日	柄なし
なし	なし	なし	なし	なし	なし	なし
大倉字	吉津巧	木、金	なし	木材を削る大工道具	朝日	なし
なし	なし	なし	なし	なし	なし	なし
黒谷字	舟木千一	鉄	なし	木材を削る	旧朝日公民館	なし
なし	なし	なし	なし	なし	なし	なし
なし	なし	鉄、雑木	なし	屋根の削りに使った	旧五十嵐家	ちゅうな
なし	なし	なし	なし	なし	なし	なし
なし	なし	鉄、雑木	なし	屋根の削りに使った	旧五十嵐家	ちゅうな
なし	なし	なし	なし	なし	なし	なし

実験結果

The screenshot shows a Microsoft Excel window titled '民具DB Ontology [互換モード] - Microsoft Excel'. The '検索と置換' (Find and Replace) dialog box is open, with the search criteria set to '戦争'. The search results are displayed in a table with the following columns: ブック (Book), シート (Sheet), 名前 (Name), セル (Cell), 値 (Value), and 数式 (Formula). The results list 22 items, all from '民具DB Ontology.xls' on 'Sheet1', with values ranging from '\$AQ\$748' to '\$AQ\$1086'. The value '\$C\$1082' is noted as '日清戦争の掛け軸' (A hanging scroll of the Sino-Japanese War). A message at the bottom of the dialog box states '22セルが見つかりました' (22 cells found).

ブック	シート	名前	セル	値	数式
民具DB Ontology.xls	Sheet1		\$AQ\$748	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$764	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$781	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$797	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$813	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$843	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$861	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$877	戦争系	
民具DB Ontology.xls	Sheet1		\$D\$887	日清戦争に	
民具DB Ontology.xls	Sheet1		\$AQ\$893	戦争系	
民具DB Ontology.xls	Sheet1		\$D\$903	日露戦争時に	
民具DB Ontology.xls	Sheet1		\$AQ\$914	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$941	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$956	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$986	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$1005	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$1022	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$1037	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$1055	戦争系	
民具DB Ontology.xls	Sheet1		\$AQ\$1073	戦争系	
民具DB Ontology.xls	Sheet1		\$C\$1082	日清戦争の掛け軸	
民具DB Ontology.xls	Sheet1		\$AQ\$1086	戦争系	

22セルが見つかりました

実験結果

- 従来のデータベースでは、使用目的に「軍人が用いた～」などと書かれている場合「戦争」で検索しても、検索漏れしてしまう。
- Ontologyデータベースでは“is-a関係”により上位概念の“戦争”が検索されるため検索漏れはしない。

実験結果

- 従来のデータベースではキリで検索してもキリであるミツメ“ギリ”、ヨツメ“ギリ”はキリの単語を含まないため検索されない。
- しかしOntologyデータベースではis-a関係によりキリでもミツメギリ、ヨツメギリが検索できる。

評価

- 従来のデータベースでは検索漏れしてしまう項目も検索できる。(戦争で使われていた水筒が、検索できるようになる。)
- グンプクや焼夷弾などの戦争に使われた道具がis-a関係により、まとめて検索できる。

まとめ

- Ontologyを使うことで検索漏れがほとんどなくなった。
- しかし短所として、二重検索などが発生。(使用目的の“戦争”とis-a関係における上位概念の戦争が検索される。)

今後の課題

- データベースで項目検索とタグ検索の切り替えが可能になれば二重検索を克服することができる。
- 今回のデータベースでは民具間同士の関連付けが上手いかなかったなのでそのあたりを改良。
- 備考についても同様の形態素解析を行う。