OWLはオントロジーの課題を本当に解決するのか？

山口 高平
（静岡大学）

Semantic Web の9階層
オントロジーに関する
知識工学と次世代Webの流れ

知識工学のトレンド

- 90-: 概念化の明示的仕様
  (Tom Gruber オントロジーの定義)
- オントロジー記述言語(Ontolingua)
- 知識交換言語(KIF)
- Generic Ontology
- CYC, WordNet, EDR...
- PSM
- Task Ontology
- オントロジー構築方法論...

次世代Webのトレンド

- 95-97: XML as arbitrary structures
- 97-98: RDF
- 98-99: RDFS (schema) as a frame-like system
- 00-01: DAML+OIL
- 02-07: OWL

OWLへの道のり

- On-To-Knowledge が OIL を定義
- 重要EUプロジェクトのいくつかが OIL を採用
- DAMLプロジェクトとOILがリンク
- 定義: DAML+OIL
- 改定: DAML+OIL
- W3Cへ DAML+OIL を提出
- WebOnt Working Group の立ち上げ
  ⇒ OWL のドラフト

Jan’00
Med’00
Sep’00
Dec’00
Mar’01
Aug’01
Oct’01
RDF(S)とOWLの関連

- class-def
- subclass-of
- slot-def
- slot-of
- domain
- range

- class-expressions
  - AND, OR, NOT
- slot-constraints
  - has-value, value-type
  - cardinality
- slot-properties
  - trans, symm

DAML Ontology Library (Ontology's by Keyword)
http://www.daml.org/ontologies/keyword.html

<table>
<thead>
<tr>
<th>Keyword</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>academic department</td>
<td><a href="http://www.cs.umd.edu/projects/plus/DAML/onts/cs1.0.daml">http://www.cs.umd.edu/projects/plus/DAML/onts/cs1.0.daml</a></td>
</tr>
<tr>
<td>academic department</td>
<td><a href="http://www.cs.umd.edu/projects/plus/DAML/onts/cs1.1.daml">http://www.cs.umd.edu/projects/plus/DAML/onts/cs1.1.daml</a></td>
</tr>
<tr>
<td>Academic Positions</td>
<td><a href="http://www.daml.ri.cmu.edu/ont/homework/cmu-ri-employmenttypes-ont.daml">http://www.daml.ri.cmu.edu/ont/homework/cmu-ri-employmenttypes-ont.daml</a></td>
</tr>
<tr>
<td>access control</td>
<td><a href="http://www.w3.org/2000/10/swap/pim/doc.rdf">http://www.w3.org/2000/10/swap/pim/doc.rdf</a></td>
</tr>
<tr>
<td>acronym</td>
<td><a href="http://orlando.drc.com/daml/Ontology/Thesaurus/CALL/current/">http://orlando.drc.com/daml/Ontology/Thesaurus/CALL/current/</a></td>
</tr>
<tr>
<td>activity</td>
<td><a href="http://www.kestrel.edu/DAML/2000/12/OPERATION.daml">http://www.kestrel.edu/DAML/2000/12/OPERATION.daml</a></td>
</tr>
<tr>
<td>Actors</td>
<td><a href="http://opencyc.sourceforge.net/daml/cyc.daml">http://opencyc.sourceforge.net/daml/cyc.daml</a></td>
</tr>
<tr>
<td>address book</td>
<td><a href="http://www.w3.org/2000/10/swap/pim/contact.rdf">http://www.w3.org/2000/10/swap/pim/contact.rdf</a></td>
</tr>
<tr>
<td>agenda</td>
<td><a href="http://www.daml.org/2001/10/agenda/agenda-ont">http://www.daml.org/2001/10/agenda/agenda-ont</a></td>
</tr>
<tr>
<td>aggregation</td>
<td><a href="http://www.daml.org/ontologies/keyword.html">http://www.daml.org/ontologies/keyword.html</a></td>
</tr>
</tbody>
</table>
ソフトウェア開発プロセス
のためのオントロジー
開発(人手)経験
Issue
Real software processes change dynamically...
very difficult to maintain
need the facility to correspond to changing process dynamically

Approach
First,
develop software process ontologies through the domain analysis using two case studies

Afterwards,
design the system which generate software processes interactively with a user

Goal
A Process-Centered Software Engineering Environment Using Ontologies

Development of Software Process Ontologies

Step1. Extract real software processes and objects from the two cases

Step2. Organize extracted processes and objects into distinct groups such that there are more similarity in meaning, while defining the software process scheme

Step3. Give detailed definitions to each processes using software process scheme
Focus on input, output, and reference roles
Focus on reference, tool and agent roles

Process Ontology

Invent
Invent with references
Invent without references

Join
Join with references
Join without references

Append
Evaluate
Review
Examine

Extract
Extract with references
Extract without references

Software Process Ontologies (conceptual hierarchy)

Process Ontology including about 250 concepts

Object Ontology including about 400 concepts
Software Process Automation using ontologies

Retrieve processes satisfied with constraints

Process Ontology

Object Ontology

Start
Object

Goal
Object

Software Process Automation using ontologies

A Process-Centered Software Engineering Environment

Process Hierarchy

Activity

Extract with references
Extract without references

Insert with references
Insert without references

Join with references
Join without references

Examine
Review
Evaluate

Extract with references
Extract without references
Case Study

A query about SEP

What SEP between the following input and output?

Input: Requirement Specification from S/W Subsystems
Output: Detailed Design of Each Component of S/W Subsystems
Interim Software Process Plan Candidates

Final Software Process Plan Candidates
Evaluation

Applicability

It turned out that the environment generated SPP candidates good for the user’s query with user interaction

Issues and Future work

• The environment can’t allocate human and time resource
• The environment has not the facility that supports the user to add user’s specific processes
• The methodology for building ontologies has no theoretical aspects, such as soundness
How to Build up Domain Ontologies with Less Cost

• Taking Existing Information Resources

DODDLE Project
(a Domain Ontology rapiD DeveLopment Environment)
exploiting a MRD
Constructing up just a conceptual hierarchy

Why not taking MRD to build up domain ontologies?

• Good News:
  MRD have many concepts.

• Bad News:
  The concepts have been defined from the point of natural language processing (common sense) and so there are some concept drift between MRD and specific domain ontologies.
Concept Drift

Domain A

Domain B

Reusable Part

MRD

Domain

No reusable part because of concept drift

DODDLE

A Set of Input Terms

Domain Expert

Spell-Match

Spell-Matched Results

Selection of Best-match

Initial Model

Trimming

Trimmed Model

Matched-Results Analysis

Trimmed-Results Analysis

Extension of Modification

Concept Hierarchy

WordNet
Matched-Results Analysis

Trimmed-Results Analysis

User re-constructs the sub-tree.
Acquiring not only Conceptual Hierarchy but Concept Definitions

**DODDLE II Project**  
(Extending DODDLE)

Info. Resource: **Domain-Specific Texts**  
Technique: Co-occurrence information  
**WordSpace**
Extracting Concept Relationships from Domain-Specific Texts

**WordSpace** (Marti A. Hearst, Hinrich Schutze)

- Words and phrases in texts can be expressed by **vector representation containing co-occurrence statistics**
- Inner products among the vectors work as the similarity between the words and phrases.

**Context Similarity** between concepts $C_1$ and $C_2$

\[
\ldots W_i \ldots W_j \ldots C_1 \ldots W_k \ldots
\]

\[
\ldots W_i \ldots W_j \ldots C_2 \ldots W_k \ldots
\]

high value with similar words coming up around the concepts
Constructing WordSpace

1. Extract Word 4-grams with high-frequency
2. Collocation Matrix
3. Context Vectors
4. Word Vectors (WordSpace is a set of word vectors)
5. Vector representations of all concepts

Texts (4-gram array)

Texts (4-gram array)

Texts (4-gram vector array)

Constructing Concept Specification Template

Taxonomic Relationships from TR Acquisition Module

ancestor, descendant and sibling

a Set of Concept Pairs from non-TR Learning Module

Concept Specification Template

\[ C_i \text{ non-TAXONY? : } C_a \]

\[ \text{TAXONY : } C_b \]

\[ \text{non-TAXONY? : } C_c \]

\[ \text{non-TAXONY? : } C_d \]
A Case Study

The Target Domain

Contracts for the International Sale of Goods (CISG)

Input Terms to TR module
46 Legal Terms from CISG Part-II

Domain-Specific Texts to NonTR module
full text of CISG (about 10,000 words)
Results

The paths from Spell Matched Nodes to the Root of WordNet

Initial Model

Trimmed Model

Legal Ontology

Select Best Matches

Trimming

Modification

46 Terms

377 Terms

113 Terms

56 Terms

61 Terms

Taxonomic Relationships for input terms constructed from TR Acquisition Module
Support Ratio

Support ratio: How much is included in final domain ontology
the intermediate products at each DODDLE activity.

![Graph showing support ratio for different stages of DODDLE process]

Extracting Concept Relationships

the concept pairs extracted according to context similarity (threshold 0.9993)

<table>
<thead>
<tr>
<th>Concept</th>
<th>Acceptance</th>
<th>Communication</th>
<th>Offer</th>
<th>Indication</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>act</td>
<td>officer</td>
<td>accept</td>
<td>effect</td>
<td>payment</td>
<td>person</td>
</tr>
<tr>
<td>assert</td>
<td>officer</td>
<td>accept</td>
<td>effect</td>
<td>person</td>
<td>offer</td>
</tr>
<tr>
<td>contact</td>
<td>party</td>
<td>empty</td>
<td>communication</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>delay</td>
<td>delivery</td>
<td>offer</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>deliver</td>
<td>payment</td>
<td>quantity</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>dispatch</td>
<td>goods</td>
<td>price</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>effect</td>
<td>person</td>
<td>accept</td>
<td>contract</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>goods</td>
<td>dispatch</td>
<td>quantity</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>inten</td>
<td>intention</td>
<td>empty</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>offer</td>
<td>act</td>
<td>accept</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>party</td>
<td>contract</td>
<td>effect</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>payment</td>
<td>quantity</td>
<td>delivery</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>place</td>
<td>payment</td>
<td>delivery</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>quality</td>
<td>quantity</td>
<td>payment</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>empty</td>
<td>communication</td>
<td>empty</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>wirefrd</td>
<td>offers</td>
<td>accept</td>
<td>effect</td>
<td>empty</td>
<td></td>
</tr>
</tbody>
</table>

Setting value for WordSpace:

- Extraction frequency of 4-gram (times): 8
- Context scope (4-grams): 60
- Context scope (4-grams): 10
Domain Experts Modifying Concept Specification Templates

ex) non-Taxonomic Relationships for “assent”

<table>
<thead>
<tr>
<th>Concept Specification Template</th>
<th>non-Taxonomic Relationships from the template</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>assent</strong></td>
<td>non-taxonomic relationship: person, offer, withdrawal</td>
</tr>
<tr>
<td><strong>offeree</strong></td>
<td>taxonomic relationship: act, proposal</td>
</tr>
<tr>
<td><strong>offeror</strong></td>
<td>inheritance: offeror, offeree</td>
</tr>
<tr>
<td><strong>person</strong></td>
<td>unnecessary: effect, time</td>
</tr>
<tr>
<td><strong>effect</strong></td>
<td><strong>agent</strong> : person</td>
</tr>
<tr>
<td><strong>offer</strong></td>
<td><strong>legal-sequence</strong> : offer</td>
</tr>
<tr>
<td><strong>withdrawal</strong></td>
<td><strong>antonym</strong> : withdrawal</td>
</tr>
<tr>
<td><strong>time</strong></td>
<td><strong>taxonomy</strong> : offer</td>
</tr>
<tr>
<td><strong>proposal</strong></td>
<td><strong>non-taxonomy</strong> : offer</td>
</tr>
</tbody>
</table>

Recall & Precision of Concept Specification Templates

The number of extracted concept pairs
Conclusions

DODDLE II: A Domain Ontology Construction Support Environment using MRD and Domain-Specific Texts

Legal experts appreciate DODDLE II to some extent.

Future Work

- How to Decide Threshold for CS
- How to Identify NTR
- Another DM method instead of WordSpace
- Another Case Studies with Large Scale

OWLとオントロジー開発

- OWLとオントロジー開発 ⇔ UMLとソフトウェア開発
- OWL＋RDF(S)＋RDF ⇒中味のあるオントロジー
- オントロジー構築支援ツールが必要
- オントロジーの重量 ⇔スケーラビリティ