The Knowledge Reengineering Bottleneck

Rinke Hoekstra
rinke.hoekstra@vu.nl

VU University Amsterdam/University of Amsterdam
Knowledge Engineering

“Critical scientific problem [...] successful applied AI requires that knowledge move from the heads of experts into programs”

- The lack of adequate and appropriate **hardware**
- Lack of cumulation of AI **methods** and **techniques**
- Shortage of trained **knowledge engineers**
- The problem of **knowledge acquisition**
- The **development gap**

**Problems of Knowledge Engineering**
Knowledge Acquisition Bottleneck

“The problem of knowledge acquisition is the critical bottleneck problem in artificial intelligence”

The Dark Ages
Knowledge Elicitation

Repertory Grids
Think Aloud Method
Cardsorting

...
Knowledge Elicitation
Repertory Grids
Think Aloud Method
Cardsorting

MYCIN and GUIDON
Knowledge Types
Knowledge Elicitation
Repertory Grids
Think Aloud Method
Cardsorting

CommonKADS
Engineering Methodology
Problem Solving Methods
Domain Models

MYCIN and GUIDON
Knowledge Types
Knowledge Elicitation
Repertory Grids
Think Aloud Method
Cardsorting

CommonKADS
Engineering Methodology
Problem Solving Methods
Domain Models

Ontolingua
“Explicit specification of a shared conceptualization”
Sharing ontologies

Mycin and Guidon
Knowledge Types
How to build the right ontology?
How to build the right ontology?

Methodologies
Middle Out Approach
Uschold & Gruninger
METHONTOLOGY
KACTUS
SENSUS
(KA)²
How to build the right ontology?

Methodologies
Middle Out Approach
Uschold & Gruninger
METHONTOLOGY
KACTUS
SENSUS
(\(KA^2\))

Ontology Types
Top
Foundation
Generic
Domain
Application
Core

Ontology Types
Top Ontology
Generic Ontology
Core Ontology
Domain Ontology
How to build the right ontology?

Methodologies
Middle Out Approach
Uschold & Gruninger

METHONTOLOGY
KACTUS
SENSUS
(KA)²

Specify Guidelines
Identify Purpose and Scope
Motivating Scenarios
Competency Questions
Ontology Capture
Ontology Coding
Ontology Integration
Evaluation

Ontology Types
Top
Foundation
Generic
Domain
Application
Core

Principles
OntoClean
Ontology vs. Epistemology
How to build the right ontology?

Methodologies
Middle Out Approach
Uschold & Gruninger
METHONTOLOGY
KACTUS
SENSUS
(KA)^2

Specify Guidelines
Identify Purpose and Scope
Motivating Scenarios
Competency Questions
Ontology Capture
Ontology Coding
Ontology Integration
Evaluation

Ontology Types
Top
Foundation
Generic
Domain
Application
Core

Ontology Reuse
Merging & Alignment
Modularization
Ontology Design Patterns

Principles
OntoClean
Ontology vs. Epistemology
2009: WebPIE

Throughput (Ktriples/sec) vs. Input size (Billions of statements)

- BigOWLIM
- Oracle 11g
- DAML DB
- BigData
- WebPIE

We are here!!

Monday 10 May 2010

2009: WebPIE

Throughput (Ktriples/sec) vs Input size (Billions of statements)

2011: QueryPIE

Backward-chaining inference at query-time, over 1B triples, in milliseconds, on just 8 parallel machines.

Pre-computation in 8-300sec against 1-3 hours in WebPIE

We are here!!

Monday 10 May 2010

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th>Terminological Closure</th>
<th>Full Closure</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>FactForge</td>
<td>862M</td>
<td>89 sec</td>
<td>2h45min</td>
<td>1:111</td>
</tr>
<tr>
<td>LinkedLifeData</td>
<td>649M</td>
<td>332 sec</td>
<td>1h5min</td>
<td>1:11</td>
</tr>
<tr>
<td>LUBM</td>
<td>1.1B</td>
<td>8 sec</td>
<td>1h15min</td>
<td>1:562</td>
</tr>
</tbody>
</table>

Knowledge
Executable Models

Knowledge
Task Independent
Domain Knowledge

Knowledge
“Semantic” Data

Knowledge Sharing
Reusable System Components
1980 - 1995

Knowledge Sharing
Ontology Reuse
1995 - 2005

Knowledge Sharing
Data Interoperability
2005 - now
Knowledge Reengineering Bottleneck

The difficulty of the correct and continuous use of preexisting knowledge for a new task
Challenge 1
Data Dependency
Design Patterns
“Data” Driven Knowledge Engineering
“Data” Driven Knowledge Engineering
Challenge 2
Complexity
Use

Reuse

Use = 1 - Reuse

Challenge 3
Limited Control
Data is Dirty
Data is Dirty
Verbose
Data is Dirty
Verbose
Inconsistent
Data is Dirty

Verbose

Inconsistent

Redundant
Data is Dirty
Verbose
Inconsistent
Redundant
Disconnected
Data is Dirty
Verbose
Inconsistent
Redundant
Disconnected
Stale
linked life data
Semantically-Interlinked Online Communities
Semantically-Interlinked Online Communities
40.745.554.078 Triples!
40 745.554.078 Triples!

(1.6 Billion)
Challenge 4
Increasing Importance
Semantic Web
Good News Quiz
http://slideshare.net/Frank.van.Harmelen/semantic-web-good-news
Semantic Web
Good News Quiz
http://slideshare.net/Frank.van.Harmelen/semantic-web-good-news
• New stakeholders
• No more fooling around
• Scary stuff...?
› Bridging the development gap
› Data publishing licenses
› Access policies
› Attribution
› “Data Hoarding”
The lack of adequate and appropriate hardware

Lack of cumulation of AI methods and techniques

Shortage of trained knowledge engineers

The problem of knowledge acquisition

The development gap

The Knowledge Reengineering Bottleneck
OWL: Experiences and Directions 2012
9th OWL: Experiences and Directions Workshop (OWLED 2012)
Heraklion, Crete, 27th-28th May, 2012

http://www.webont.org/owled/2012
@OWLED2012Worksh