

Ontologier — visioner och verktyg

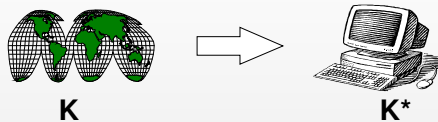
Henrik Eriksson
Institutionen för Datavetenskap
Linköpings universitet

Ontologies

- Provide a domain of discourse for characterizing some application area
- Enumerate concepts, attributes of concepts, and relationships among concepts, thus defining a structure for the application area
- Define constraints on relationships among concepts
- Can be use to
 - Describe aspects of the world
 - Define object-oriented models



Acquisition of ontology knowledge

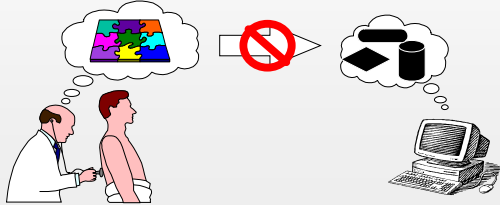


The *transformation* of knowledge from the forms in which it is available in the world into forms that can be used by computers



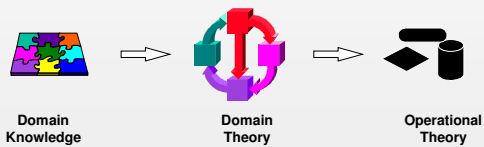
Acquiring Expert Knowledge: Transferring versus Modeling

It is not possible to transfer directly a domain's expert knowledge to a machine because the respective representations are too dissimilar



Acquiring Expert Knowledge: Transferring versus Modeling (2)

Knowledge acquisition is a *modeling* process. First build a *theory* of a domain and then make that theory *operational*



Ontologies vs. Object-Oriented Models

An ontology

- reflects the structure of the world
- is often about structure of concepts
- actual physical representation is not an issue

An OO class structure

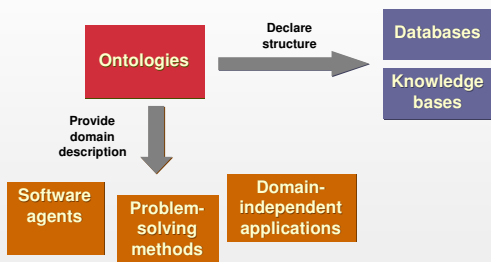
- reflects the structure of the data and code
- is usually about behavior (methods)
- describes the physical representation of data (long int, char, etc.)

The ontology-modeling process

- Determine scope
- Consider reuse
- Enumerate terms
- Define classes
- Define properties
- Define constraints
- Create instances



Ontologies are often just the beginning

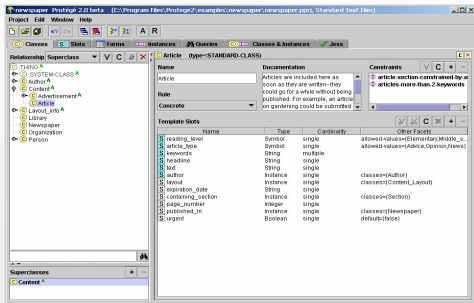


The Protégé tool

- An ontology editor and development system
 - Twenty thousand users world-wide
 - Developed at Stanford University
- Represents the latest in a series of interactive tools for knowledge-system development
- Allows a variety of "plug ins" to facilitate customization in various dimensions
- Outstanding features
 - Automatic generation of graphical-user interfaces, based on user-defined models, for acquiring domain instances
 - Extensible knowledge model and architecture
 - Scalability to very large knowledge bases
 - Facilitates construction of ontologies and knowledge bases in a principled fashion



Protégé screenshot



Ontology Web Language — OWL

- Ontology language for the semantic web
- Builds on RDF and RDF Schema
- Adds richer descriptions of properties and classes
 - relations between classes (e.g. disjointness)
 - cardinality (e.g. "exactly one")
 - equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes
- OWL plug-in available for Protégé



OWL Sublanguages

- OWL Lite
 - Supports a classification hierarchy and simple constraints
- OWL DL
 - Supports maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time)
- OWL Full
 - Supports maximum expressiveness and the syntactic freedom of RDF with no computational guarantees

The Semantic Web

- Annotation of web pages is great for
 - Improved search
 - Services
 - Inference
- ...but much knowledge is found elsewhere in documents
 - PDF files
 - MS Word files



12

A vision for the semantic web



13

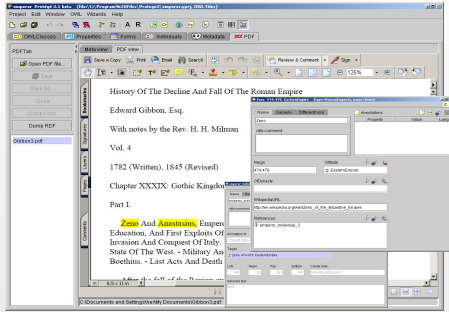
Semantic documents

- Knowledge representation
 - Semantic web: OWL
 - Ontologies
- Document models
 - Adobe's Portable Document Format (PDF)
 - Extensible Metadata Platform (XMP)
 - Microsoft Word (?)
- Functionality
 - Semantic search based on metadata
 - Reasoning, inferences



14

Semantic documents



Semantic document production at Statistics Sweden

- Repository of statistics reports
- Difficult to find relevant information
- Document model and knowledge representation
 - Adobe's Portable Document Format (PDF)
 - OWL (statistics ontology)
 - Protégé tab (PDF/OWL editor for tagging)
- Functions
 - Semantic search engine



The goal is to assist users in finding statistics-related information in a large repository of PDF documents

Example: Markup of PDFs

KPI, huvudgruppen, mån - Statistik från SCB

Statistiska centralbyrån
Statistik Sweden

Hitta rätt: A-O | Statistik efter ämne

Priser och konsumtion » Konsumtionsindex (KPI) »
Konsumtionsindex (1980=100), skuggindex samt index för huvudgrupper

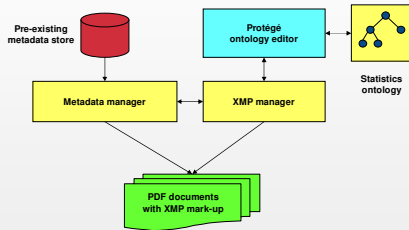
Produktindelningen ändrades (till COICOP) vid publiceringen avseende januari 2000.
 Vid indexreglering med konsumtionsindex används normalt fastställda priser. Indexstalen här är i viss utsträckning justerade i förhållande till de fastställda totalindexstalen.

	Alkohol- kvarter	Käslor	Bo- inven- turer	Hälsa- och sport	Tran- sport	Post o tele- komm- unik	Res- resa- kostn	Div- varor	KPI			
2004:5	241,1	349,6	165,7	325,3	229,6	732,6	352,9	214,7	194,3	387,1	395,0	380,1
2004:4	241,0	349,2	165,4	325,9	230,1	730,3	346,6	215,1	194,0	386,5	294,0	279,4
2004:3	241,7	349,1	162,2	327,7	229,8	729,6	344,7	215,7	194,0	385,9	293,4	279,4
2004:2	240,1	349,1	150,7	328,3	228,9	725,8	337,2	217,8	194,2	385,3	293,2	277,3

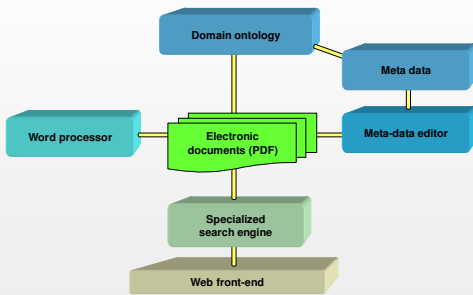
Table heading description in OWL

Table descriptions in OWL

Models and tools

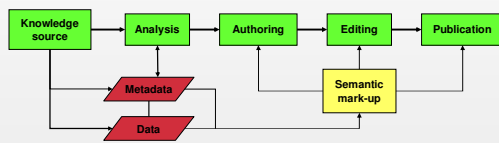


Architecture



Document production

- Idea: Tool support for the entire process
 - Metadata is kept throughout the process
 - Support for tagging based on data sources, including metadata



Research Goals

- Understand semantic document production from document creation to actual use of services
- Develop a semantic search engine and OWL-PDF markup system
- Investigate the advantage of semantic web technologies



Current questions and future research issues

- How should we describe documents?
 - OWL just the start
 - Standards?
- How should we describe the services?
- Interoperability of services and devices
 - Standards?
- Collaborative authoring of descriptions
 - Trust
- Tools for semantic web site management
 - OWL-editors in the commercial web production systems
- Translation among overlapping ontologies?
 - Semantic responsibility



Conclusions

- Great potential
 - Powerful description tools
 - Standards
 - Momentum
- ...but some problems
 - Entry threshold and learning curve
 - Fragmented languages/tools
 - Questionable web pages
- Semantic documents
- Applications emerging
 - Ontology applications exist
 - SW applications appearing



Links

- **Semantic web introduction**
 - <http://www.semanticweb.org/>
 - <http://infomesh.net/2001/swintro/>
 - <http://www.disobey.com/detergent/2002/sw123/>
 - <http://www.w3c.org/>
- **Ontology introduction**
 - <http://www-ksl.stanford.edu/ksl/what-is-an-ontology.html>
 - http://protege.stanford.edu/publications/ontology_development/ontology101.html
 - <http://www.w3.org/2004/OWL/>
- **Protégé**
 - <http://protege.stanford.edu/>