

DATA SCIENCE MASTER

SEMANTIC KNOWLEDGE REPRESENTATION

Ontology Building

Mounira Harzallah

mounira.harzallah@univ-nantes.fr

Tél: 02 28 09 21 27

Plan

Ontology Building Steps

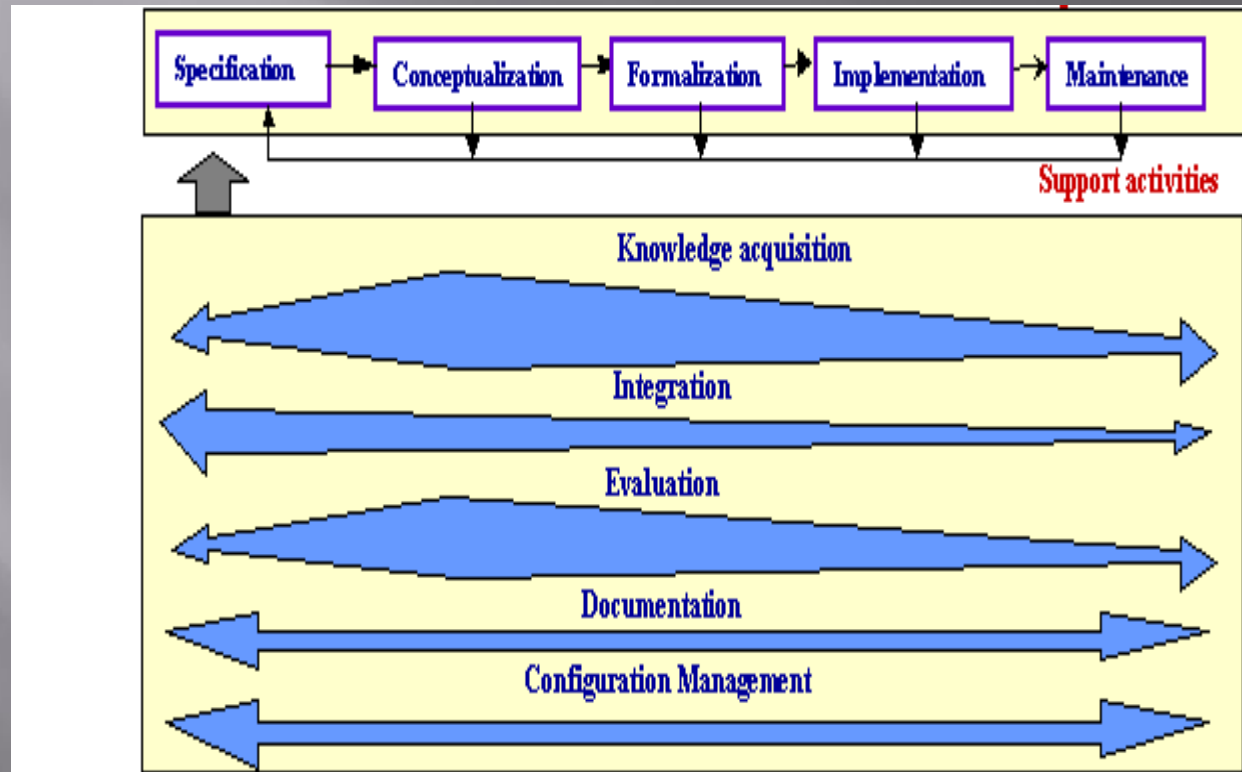
Ontology Conceptualization : Manual techniques

Ontology Building

General method

- ▣ Ontology Specification
- ▣ Conceptualisation
 - Knowledge acquisition
 - Informal specification of ontology components
- ▣ Formalisation
- ▣ Implementation
- ▣ Evaluation

METHONTOLOGY



Activities in the ontology building process of METHONTOLOGY.

Ontology Building

Step 1: Specification

1. Definition of the domain of the ontology to build
2. Definition of the ontology competences
e.g. Knowledge elicitation, Document annotation,
Information retrieval, Object comparison
3. Definition of user scenario
4. Concepts/Relations that should be in the ontology to build

Step Results.

ORSD : Ontology Required Specification Document

Ontology Building

Step 1: Specification

Examples of user scenario

- ▣ To compare the similarity of papers. Papers should be annotated by the ontology. Then, for two given papers, the sets of concepts that annotate them are compared using semantic measures.

Ontology Building

Step 2: Conceptualisation

- ▣ Knowledge Acquisition:

Identification of concepts that exist in the considered domain and that are **relevant** to ontology competences.

- ▣ Informal Specification

Identification of relationships, properties, instances and rules (axioms) related to concepts identified.

Ontology Building

Step 3: Formalisation

Formalise

- Concepts
- Properties
- Relations
- Axioms

Using a language +/- formal:

- Description logic

Ontology Building

Step 4: Implementation

- ❑ Choose a language to implement the ontology
 - ❑ RDF
 - ❑ OWL
 - ❑ ...
- ❑ Using an ontology editor
 - ❑ Protégé2000
 - ❑ KAON
 - ❑ ...
- ❑ Implementation

Ontology Building

Step 5: Evaluation

- **Evaluation method**
 - By experts
 - By its comparison with a Gold standard
 - With an application that uses it
- **Kinds of errors within ontology**
 - **Inconsistency**
 - **Logical inconsistency** when some statements contradict one another
Ex. c is an instance of A and B and $A \sqcap B = \perp$
 - **Unadapted ontology** when something is not correct according to an intended conceptualization
 - **Incompleteness** when something is missing according to an intended conceptualisation. Ex A missing of a known disjunction
 - **Inconciseness**
Ex. A is B , B is C and A is C (Redundancy)

Step 2: Conceptualisation

Ontology reuse

The **reuse of ontology** (domain ontology or upper ontology) allows to build a larger ontology with a better quality, while reducing the development cost.

To locate candidate ontologies for reuse

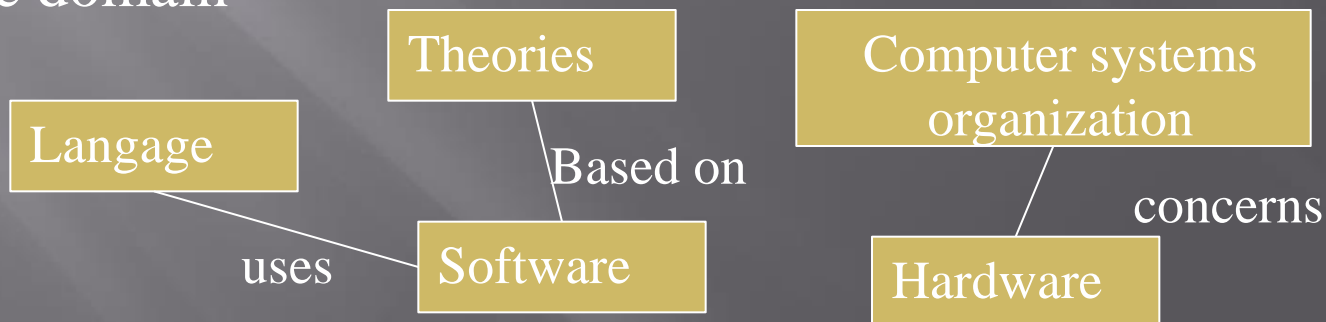
Step 2: Conceptualisation

Use of a core ontology of the domain

The core ontology is a model that represents the core categories (core concepts) of the ontology to build and the relationships between them.

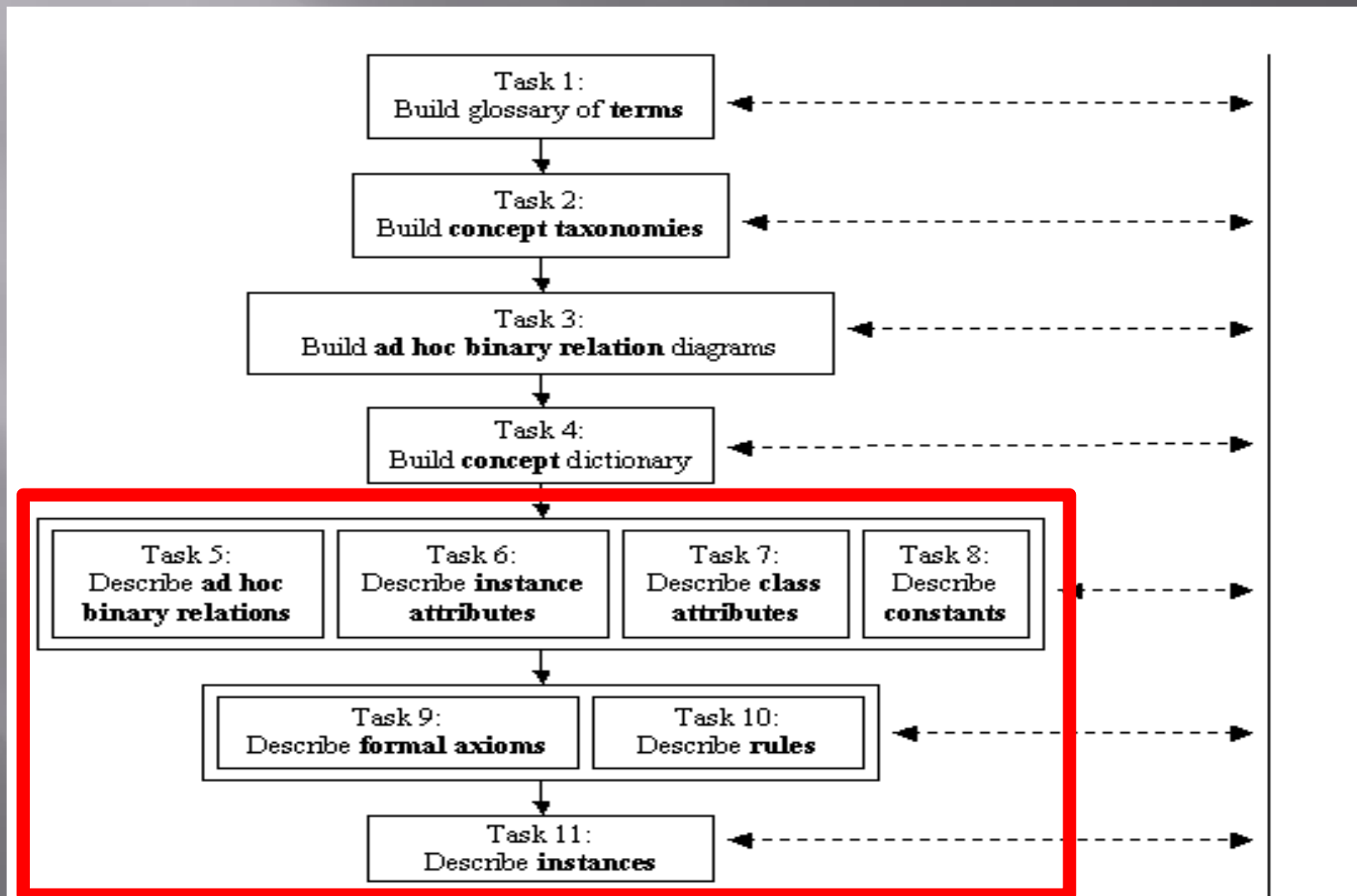
The core concepts should allow the definition of the other domain concepts by specialisation : *It is about a top down approach for ontology building.*

An excerpt of an example of a core ontology of the computer science domain



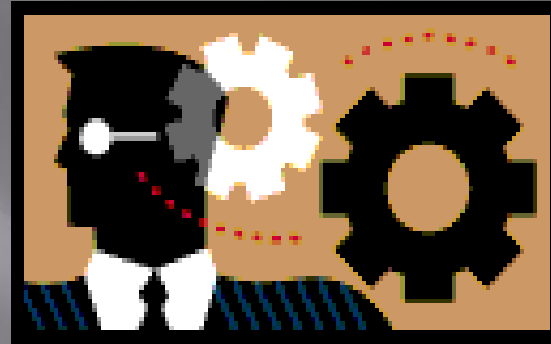
Step 2: Conceptualisation

Conceptualization tasks according to METHONTOLOGY



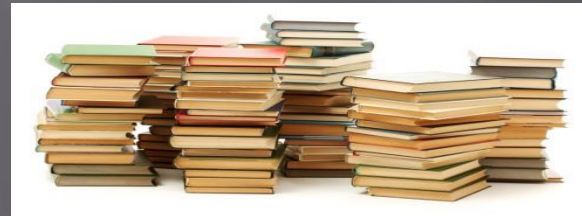
Step 2: Conceptualisation

With experts



AND/OR

From texts



Step 2: Conceptualisation

Ontology Conceptualisation Tasks :

Bottom-up approach

OntoLearn \in Tool ^I

Tool \sqcap hypernym_relation $\equiv \perp$

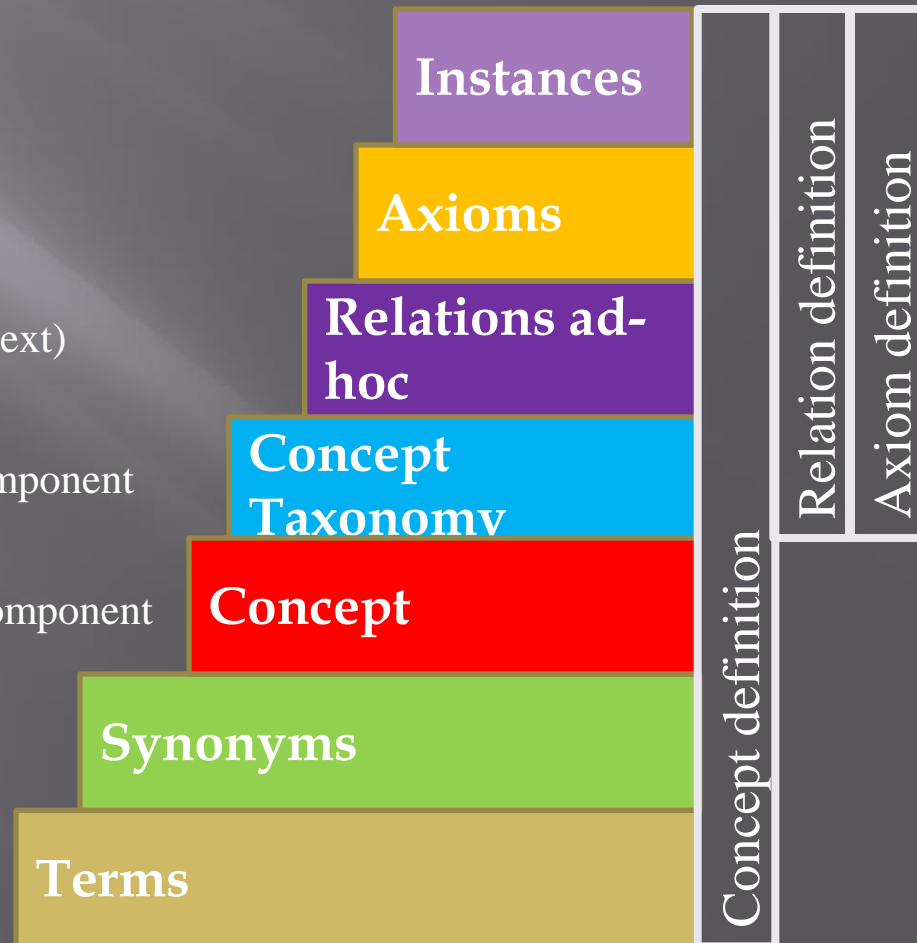
Extract (tool, hypernym_relation), Extract from(tool, text)

Hypernym_relation \sqsubseteq relation, concept \sqsubseteq ontology component

Hypernym_relation, tool, relation, concept, ontology_component

{hypernym_relation, is-a_relation, taxonomy_relation},
{System, Tool}

ontology, system, outil, step, ontology_component,
hypernym_relation, is-a_relation, taxonomy, hierachiy,
validation,...



Step 2: Conceptualisation

Techniques used

- Manual – discussions with experts, analyses of texts
- Automatic – text mining and data mining
- Hybrid – filtration of data extracted automatically, by experts

Step 2: Conceptualisation

Glossary definition

1. **Identification of relevant terms** by experts based on their knowledge. Texts, models and database relevant to ontology domain can be used
2. Use of **a core ontology to classify** extracted terms
3. Identification and clustering of **synonym terms in synset**
4. Natural Language definition of terms.
5. **From terms to concepts** : Choice of a label for a synset that will be the noun of the concept that represents this synset

The synset {car, autocar, auto } are regrouped under the concept car

6. Identification and definition of concept properties (on natural language)

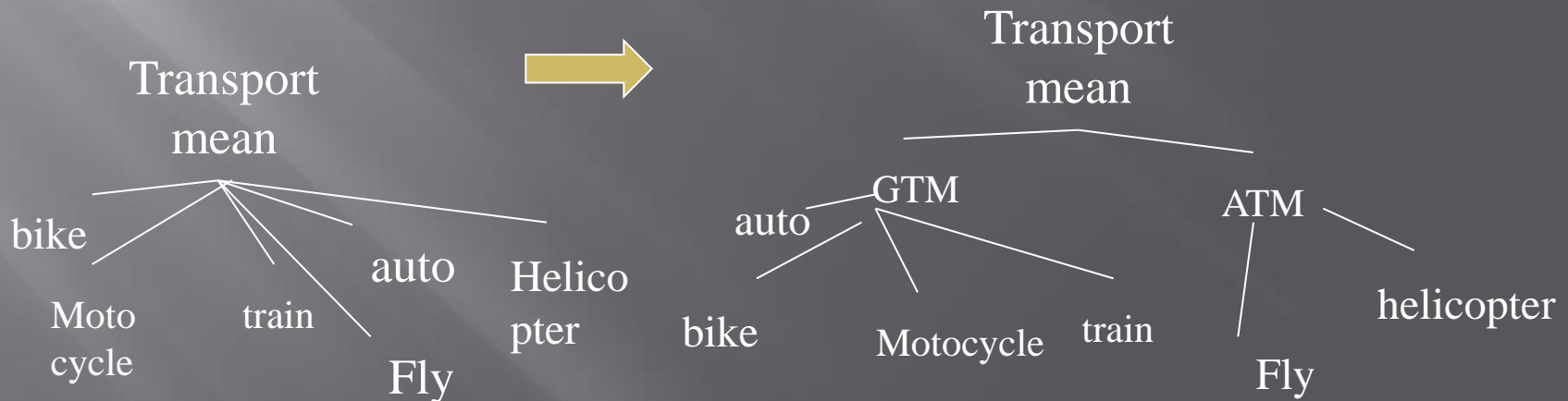
Step 2: Conceptualisation

Taxonomy building

Identification of subsumption relations between concepts that are classified under the same core concept (concepts that are subconcepts of the same core concept)

Use of concept properties to classify concepts

Addition of new concepts (if necessary) to improve the classification



Step 2: Conceptualisation

Ad-hoc relation identification

Identification of principal relations of the domain other than that already identified between the core concepts

Classic relations are usually used within an ontology

Structural relation: part of

Other classic relations: use, is related to, subtopic of

Identification of other relations relevant for the domain

Step 2: Conceptualisation

Manual Conceptualisation from texts

1 Build a core ontology and/or reuse of existing ontologies

2. Glossary Definition

- Extract and validate terms
- Structure the results regarding the core concepts (e.g. see table below)

Term	Tag	Frequency	Validated or not	Has synonyms	Concept label	Properties /Definition	Core concept

Manual conceptualisation from texts

Example

Learning domain ontologies from document warehouses and dedicated web sites.

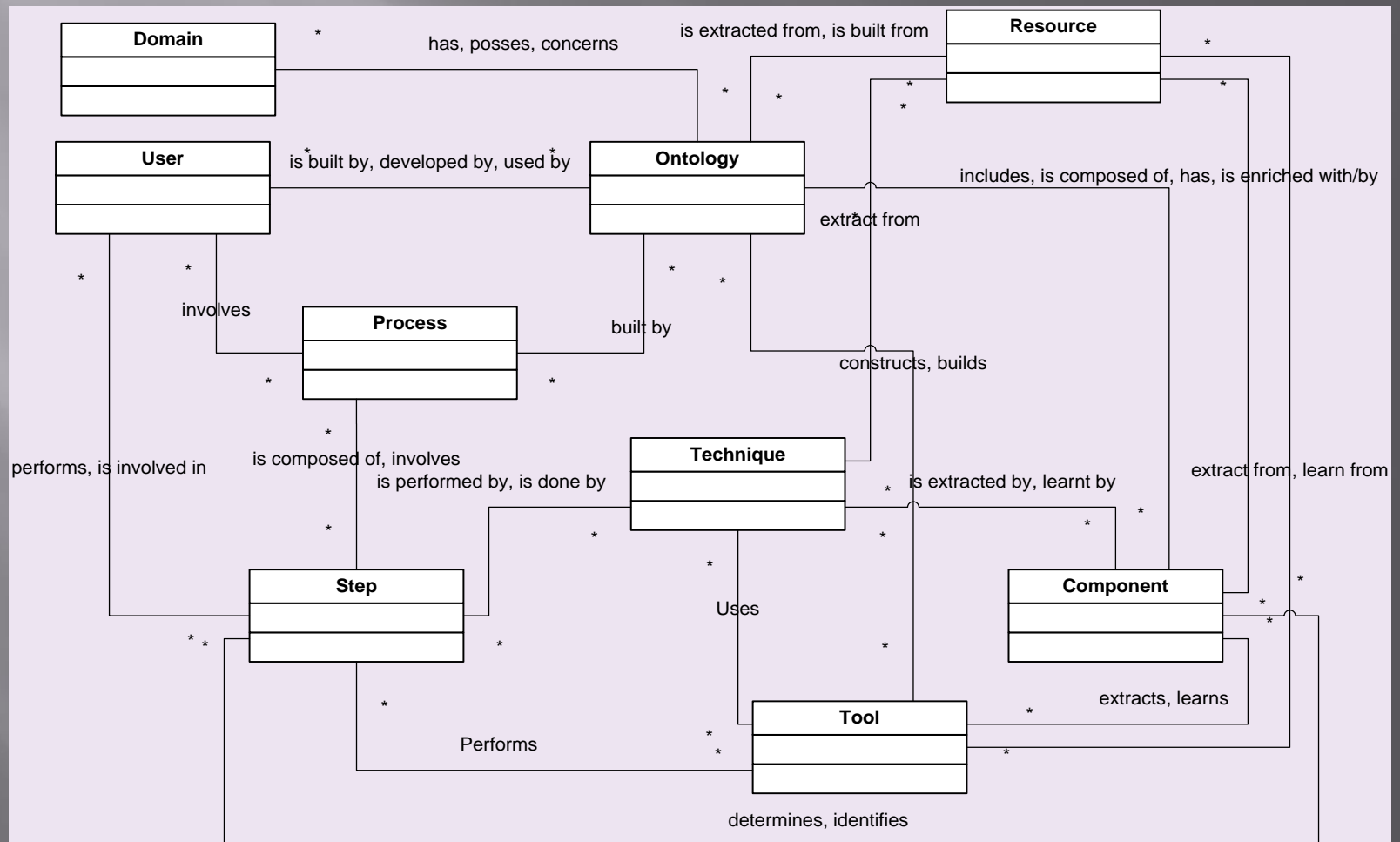
Ontology evaluation is a critical task, even more so when the ontology is the output of an automatic system, rather than the result of a conceptualization effort produced by a team of domain specialists and knowledge engineers. This paper provides an evaluation of the OntoLearn ontology learning system. The proposed evaluation strategy is twofold: first, we provide a detailed quantitative analysis of the ontology learning algorithms. Second, we automatically generate natural language descriptions of formal concept specifications in order to facilitate per-concept qualitative analysis by domain specialists.

Automatic generation of glosses in the ontolearn system.

OntoLearn is a system for automatic acquisition of specialized ontologies from domain corpora, based on a syntactic pattern matching technique for word sense disambiguation, called structural semantic interconnection (SSI). We use SSI to extract from corpora complex domain concepts and create a specialized version of WordNet. In order to facilitate the task of domain specialists who inspect and evaluate the newly acquired domain ontology, we defined a method to automatically generate glosses for the learned concepts. Glosses provide an informal description, in natural language, of the formal specifications of a concept, facilitating a per concept evaluation of the ontology by domain specialists, who are usually unfamiliar with the formal language used to describe a computational ontology. The proposed evaluation framework has been tested in a financial domain.

Manual conceptualisation from texts

Core-ontology-based approach



Manual conceptualisation from texts

Example. Term extraction and classification

Learning domain ontologies from document warehouses and dedicated web sites.



Ontology



Process



Step



Component



Tool



Resource



User



Technique



Is-a

Ontology evaluation is a critical task, even more so when the ontology is the output of an automatic system, rather than the result of a conceptualization effort produced by a team of domain specialists and knowledge engineers. This paper provides an evaluation of the OntoLearn ontology learning system. The proposed evaluation strategy is twofold: first, we provide a detailed quantitative analysis of the ontology learning algorithms. Second, we automatically generate natural language descriptions of formal concept specifications in order to facilitate per-concept qualitative analysis by domain specialists.

Automatic generation of glosses in the ontolearn system.

OntoLearn is a system for automatic acquisition of specialized ontologies from domain corpora, based on a syntactic pattern matching technique for word sense disambiguation, called structural semantic interconnection (SSI). We use SSI to extract from corpora complex domain concepts and create a specialized version of WordNet. In order to facilitate the task of domain specialists who inspect and evaluate the newly acquired domain ontology, we defined a method to automatically generate glosses for the learned concepts. Glosses provide an informal description, in natural language, of the formal specifications of a concept, facilitating a per concept evaluation of the ontology by domain specialists, who are usually unfamiliar with the formal language used to describe a computational ontology. The proposed evaluation framework has been tested in a financial domain.

Manual conceptualisation from texts






Example. Term extraction and classification

Learning domain ontologies from document warehouses and dedicated web sites.

Ontology	Ontology evaluation is a critical task, even more so when the ontology is the output of an automatic system, rather than the result of a conceptualization effort produced by a team of domain specialists and knowledge engineers. This paper provides an evaluation of the OntoLearn ontology learning system. The proposed evaluation strategy is twofold: first, we provide a detailed quantitative analysis of the ontology learning algorithms. Second, we automatically generate natural language descriptions of formal concept specifications in order to facilitate per-concept qualitative analysis by domain specialists.
Process	
Step	
Component	
Tool	Automatic generation of glosses in the ontolearn system.
Resource	OntoLearn is a system for automatic acquisition of specialized ontologies from domain corpora, based on a syntactic pattern matching technique for word sense disambiguation, called structural semantic interconnection (SSI). We use SSI to extract from corpora complex domain concepts and create a specialized version of WordNet. In order to facilitate the task of domain specialists who inspect and evaluate the newly acquired domain ontology, we defined a method to automatically generate glosses for the learned concepts. Glosses provide an informal description, in natural language, of the formal specifications of a concept, facilitating a per concept evaluation of the ontology by domain specialists, who are usually unfamiliar with the formal language used to describe a computational ontology. The proposed evaluation framework has been tested in a <u>financial domain</u> .
User	
Technique	

Manuel conceptualisation from texts

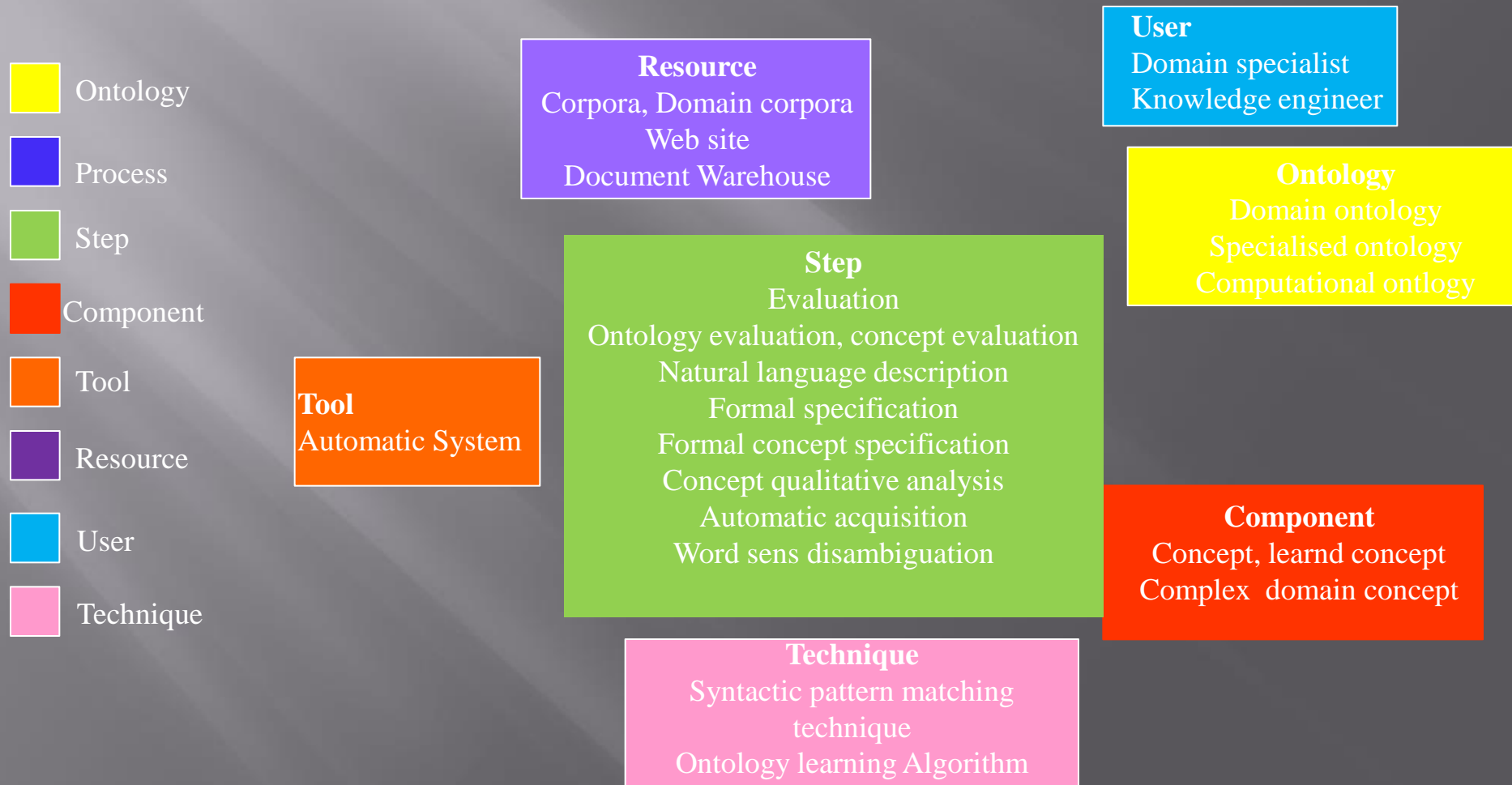
Example. Term extraction and classification

	Ontology
	Process
	Step
	Component
	Tool
	Resource
	User
	Technique

Term	Tag	FR	Y: Y/N	Synonym	Conce	CoreCon
Domain Ontology	NN		Y			Ontology
document warehouse	NN		Y			Resource
web site	NN		Y			Resource
Ontology evaluation	NN		Y			STEP
Task	NN		Y	STEP	STEP	STEP
Ontology	NN		Y			Ontology
automatic system	NN		Y		?	TOOL
Conceptualisation	NN		Y			STEP
Domain specialist	NN		Y			USER
Knowledge engineer	NN		Y			USER
Evaluation	NN		Y			STEP
OntoLearn ontology learning system	PN		Y	Ontolearn	OntoLea	TOOL
Ontology learning algorithm	NN		Y			Technique
Natural language description	NN		Y			STEP
Formal concept specification	NN		Y			STEP
Concept qualitative analysis	NN		Y			STEP
STructural semantic interconnection	NN		Y			STEP
System	NN		Y	TOOL	TOOL	TOOL
Automatic acquisition	NN		Y			STEP
Specialized ontology	NN		Y			Ontology
domain corpora	NN		Y			Resource
syntactic pattern matching technique	NN		Y			Technique
word sens disambiguation	NN		Y			STEP
strcutural semantic interconnection	PN		Y	SSI		Technique
corpora	NN		Y			Resource
Complex domain concept	NN		Y			Component
Wordnet	PN		Y			Ontology
method	NN		Y	Technique	Techniqu	Technique
gloss	NN		Y			Component
learned concept	NN		Y			Component
Formal specification	NN		Y			STEP
Concept	NN		Y			Component
Concept Evaluation	NN		Y			STEP
Computational ontology	NN		Y			Ontology






Manual conceptualisation from texts

Example. Term extraction and classification



Manual conceptualisation from texts

Example. Hypernym relation extraction

-  Ontology
-  Process
-  Step
-  Component
-  Tool
-  Resource
-  User
-  Technique

Learning domain ontologies from document warehouses and dedicated web sites.

Ontology evaluation is a critical task, even more so when the ontology is the output of an automatic system, rather than the result of a conceptualization effort produced by a team of domain specialists and knowledge engineers. This paper provides an evaluation of the OntoLearn ontology learning system. The proposed evaluation strategy is twofold: first, we provide a detailed quantitative analysis of the ontology learning algorithms. Second, we automatically generate natural language descriptions of formal concept specifications in order to facilitate per-concept qualitative analysis by domain specialists.

Automatic generation of glosses in the ontolearn system.

OntoLearn is a system for automatic acquisition of specialized ontologies from domain corpora, based on a syntactic pattern matching technique for word sense disambiguation, called structural semantic interconnection (SSI). We use SSI to extract from corpora complex domain concepts and create a specialized version of WordNet. In order to facilitate the task of domain specialists who inspect and evaluate the newly acquired domain ontology, we defined a method to automatically generate glosses for the learned concepts. Glosses provide an informal description, in natural language, of the formal specifications of a concept, facilitating a per concept evaluation of the ontology by domain specialists, who are usually unfamiliar with the formal language used to describe a computational ontology. The proposed evaluation framework has been tested in a financial domain.

Manual conceptualisation from texts

Example. Ad-hoc relation extraction



Ontology	Ontology evaluation is a critical task, even more so when the ontology is the output of an automatic system, rather than the result of a conceptualization effort produced by a team of domain specialists and knowledge engineers. This paper provides an evaluation of the OntoLearn ontology learning system. The proposed evaluation strategy is twofold: first, we provide a detailed quantitative analysis of the ontology learning algorithms. Second, we automatically generate natural language descriptions of formal concept specifications in order to facilitate per-concept qualitative analysis by domain specialists.
Process	
Step	
Component	
Tool	Automatic generation of glosses in the ontolearn system.
Resource	OntoLearn is a system for automatic acquisition of specialized ontologies from domain corpora, based on a syntactic pattern matching technique for word sense disambiguation, called structural semantic interconnection (SSI). We use SSI to extract from corpora complex domain concepts and create a specialized version of WordNet. In order to facilitate the task of domain specialists who inspect and evaluate the newly acquired domain ontology, we defined a method to automatically generate glosses for the learned concepts. Glosses provide an informal description, in natural language, of the formal specifications of a concept, facilitating a per concept evaluation of the ontology by domain specialists, who are usually unfamiliar with the formal language used to describe a computational ontology. The proposed evaluation framework has been tested in a financial domain.
User	
Technique	