

Terminology in ontologies and other lexicographic resources

Dra. Guadalupe Aguado de Cea

lupe@fi.upm.es

<http://www.oeg-upm.net>

Ontological Engineering Group

Facultad de Informática

Universidad Politécnica de Madrid

Campus de Montegancedo sn,

28660 Boadilla del Monte, Madrid, Spain

Outline

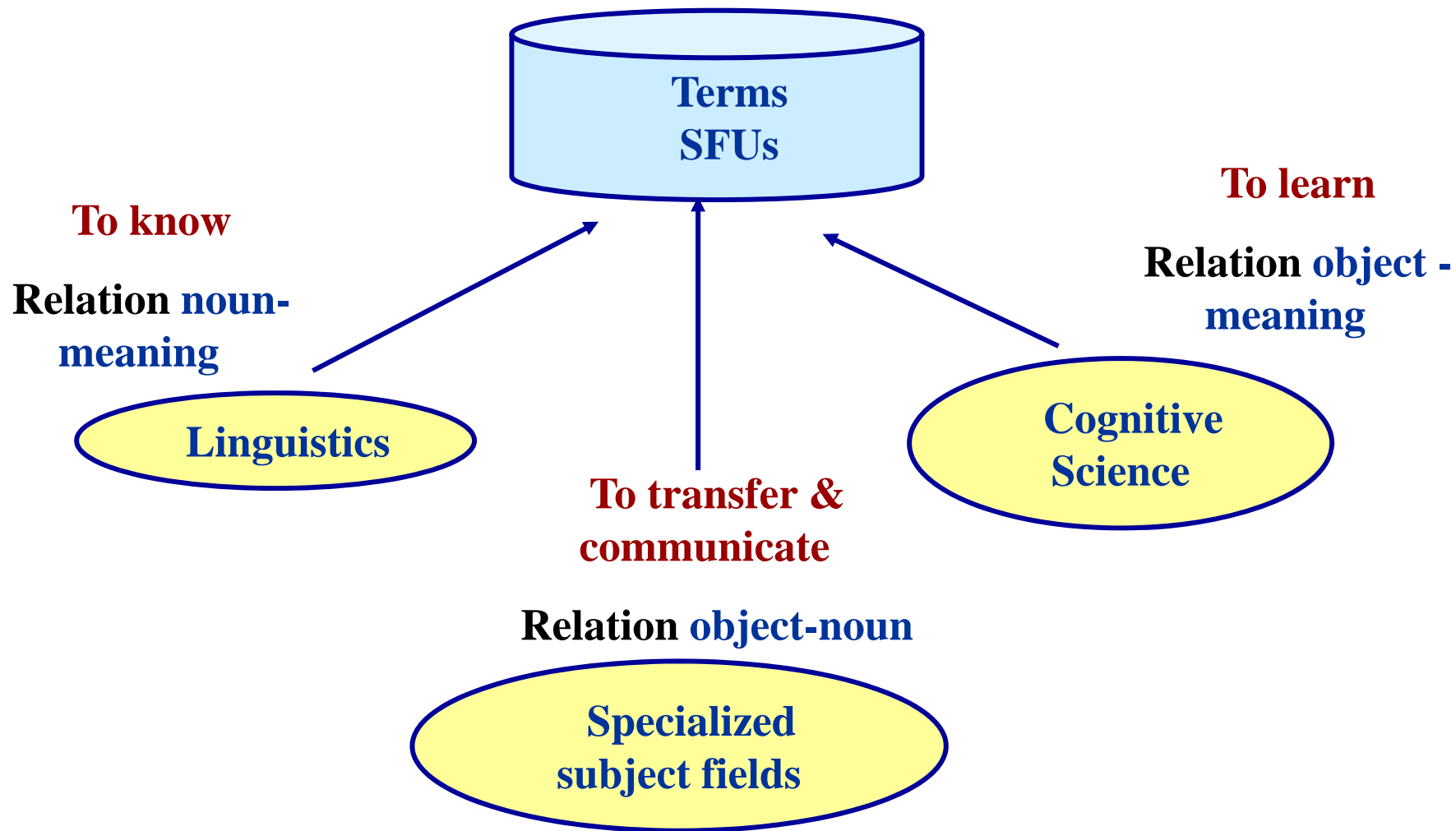
- What is terminology?
- Object of study and working methods in terminology
- The insights of terminology
- Concepts and their relations
- Terms and their formation
- Linguistic resources and ontologies
- Reusing non ontological resources
- Comparative analysis of resources

What is terminology?

The term terminology is polysemic:

- As a **product**: set of terms from a given subject field.
- As a **discipline**: set of fundamental principles and conceptual bases that govern the study of specialized terms, their description, analysis and relations.
- As a **practice**: set of principles oriented towards term compilation

The object of study of terminology



Working methods of Terminology

- **Identifying** terminological units (TUs)
 - **Conceptual perspective:** their relations with the other terms of the conceptual system and their distinctive features.
 - **Functional perspective:** specialized phraseological units (SPUs):
 - *to set up a loom*, (montar un telar),
 - *to produce cloth*, (construir un tejido),
 - *to download a file* (bajarse un fichero, descargar un fichero),
 - *to run a program* (ejecutar un programa).
- **Contextual Analysis:** Analysing terms in context and the terminological variants according to degrees of specialization: *ordenador, computador, ordenata, máquina, PC*,
- **Creation of new terms:** (a) new terms are needed to describe new process, devices, phenomena, (b) lack of suitable terms when translating neologisms from another language: *folksonomies, servlet, applet, widget, snippet*
- **Normalization**

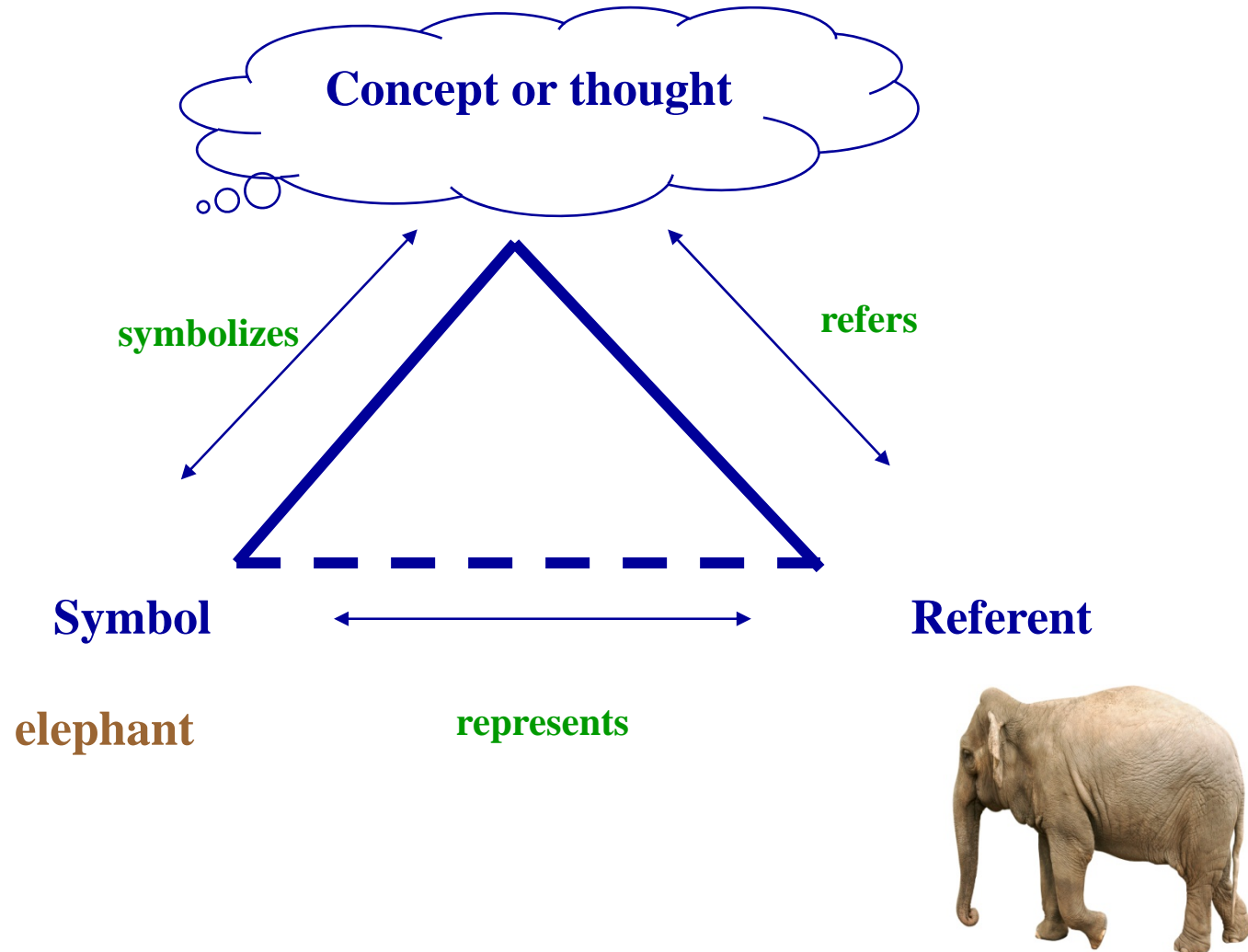
Theoretical and descriptive goals of terminology

- Concepts
- Terms
- Relation between terms and concepts
- Definitions (not included in this lesson)

Definition of a concept

- *In general: Unit of knowledge created by a unique combination of characteristics. ISO 1087- 1 (2000)*
- *In terminology: Concepts shall be considered mental representation of objects within a specialized context or field*
ISO/DIS 704 (2006)
- **BUT** concepts are influenced by social and cultural circumstances given at a certain moment, and this can lead to different classifications in the conceptual system
- From a **socio-communicative** perspective: **Concepts** can be seen as:
 - Units of **thought**: represent and recognize the object mentally
 - Units of **knowledge**: represent knowledge in each subject field
 - Units of **communication**: transmit knowledge by means of linguistic symbols

Semantic triangle by Ogden & Richards



Description of a concept

- Concepts are described according to their common **features**, **properties** or **characteristics**, either by intension or extension
- **Intension**
 - Set of characteristics which makes up the concept (ISO 1087-1: 2000)
 - The bigger the number of common characteristics, the more restricted is the intension.
 - The intension of the concept **winter** in polar countries includes: low temperatures, ice, wind, snow, etc.
- **Extension**
 - Totality of objects to which a concept corresponds (ISO 1087-1: 2000)
 - A general concept has a wide extension as it includes two or more objects by reason of common properties.
 - The extension of the concept **planet** includes: *Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune and Pluto.*

Characteristic of a concept

- “Abstraction of a property of an object or of a set of objects” (ISO 1087-1:2000)
- According to the **importance** in forming a concept
 - **essential**: indispensable to understand and distinguish a concept
 - *The back of a seat distinguishes a stool and a chair.*
 - **complementary**: *colour, material, shape, ...*
- According to the **relation** with the object represented
 - **intrinsic**, which are observable properties:
 - **Shape**: oval, round, narrow, wide, ...
 - **Material**: wooden, stone, metallic, ...
 - **Colour**: red, blue, green, orange...
 - **Position**: vertical, hanging, slanting
 - **extrinsic**, relation of the object with others
 - **Mode of employment** or application: *analogic, digital, hybrid.*
 - **Origin** or how an object comes into existence: producer, inventor, provider, the place of its production, (town, country), ...



Subject field Computer hardware	Object (visual representation)	Concept Abstraction based on the set of all mechanical mice	Term: mechanical mouse	ISO /DIS 704, 2006
Category	Property	Characteristic	Degree of essentiality	
Level of abstraction	concreteness	1. concreteness	essential	
Type of object	Input device	2. input device	essential	
colour	Ivory-coloured	3. Any colour	Non-essential	
usage	Hand-manoeuvred along a hard, flat-surface	4. Has a ball on its underside on which the device rolls	essential	
composition	Has a metal ball on its underside on which the device rolls	5. Has a ball on its underside on which the device rolls	essential	
composition	Has three buttons	6. Has at least one button	essential	
connection	Connected to a computer by a connecting wire	7. May be connected to a computer by a wire	Non-essential	
function	Rollers detect the movement of the ball	8. Rollers (mechanical sensors) detect the movement of the ball	essential	
function	The ball controls the movement of a cursor on a computer display screen	9. Ball movement controls the movement of a pointer on a computer display screen	essential	

Relations between concepts: hierarchical relations

Close relation between a concept and its characteristics

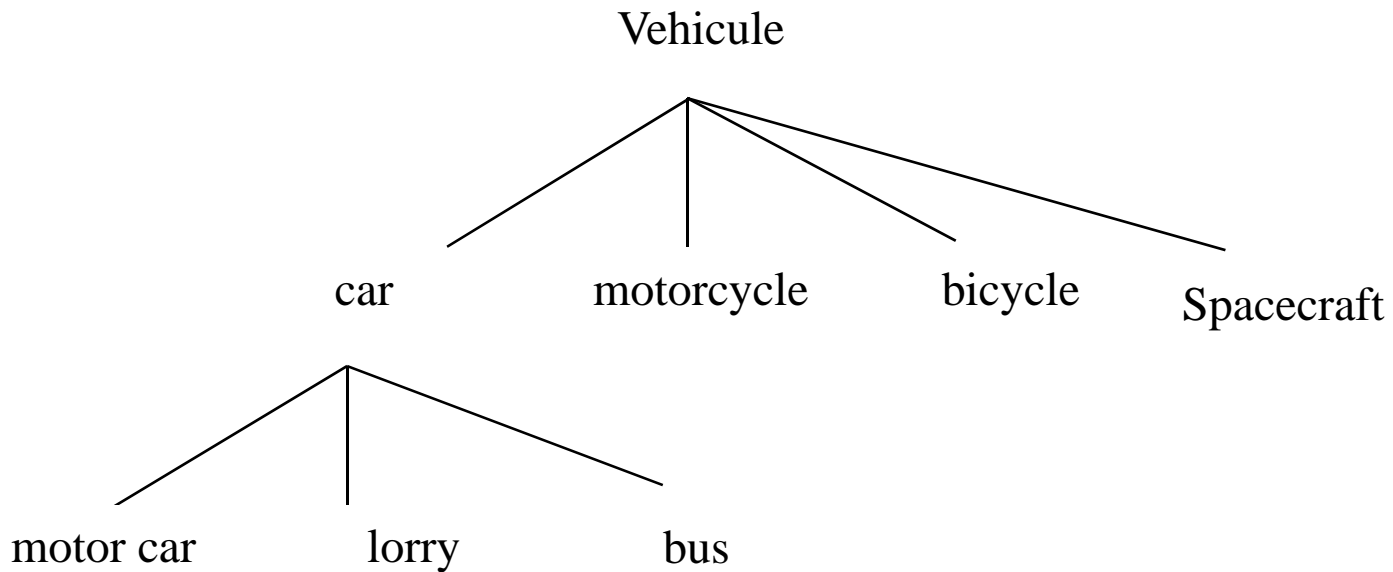
A. **GENERIC RELATIONS** (**genus-species** relation) IS_A

- One of the concepts includes another concept
 - **vertical:** hypernym- hyponym; superordinate –subordinate
 - **horizontal:** two specific ideas of the same generic concept with some distinguishing characteristics
- In thesaurus
 - *Broader than* (BT)
 - *Narrower than* (NT)
 - *Associated to* (AT)

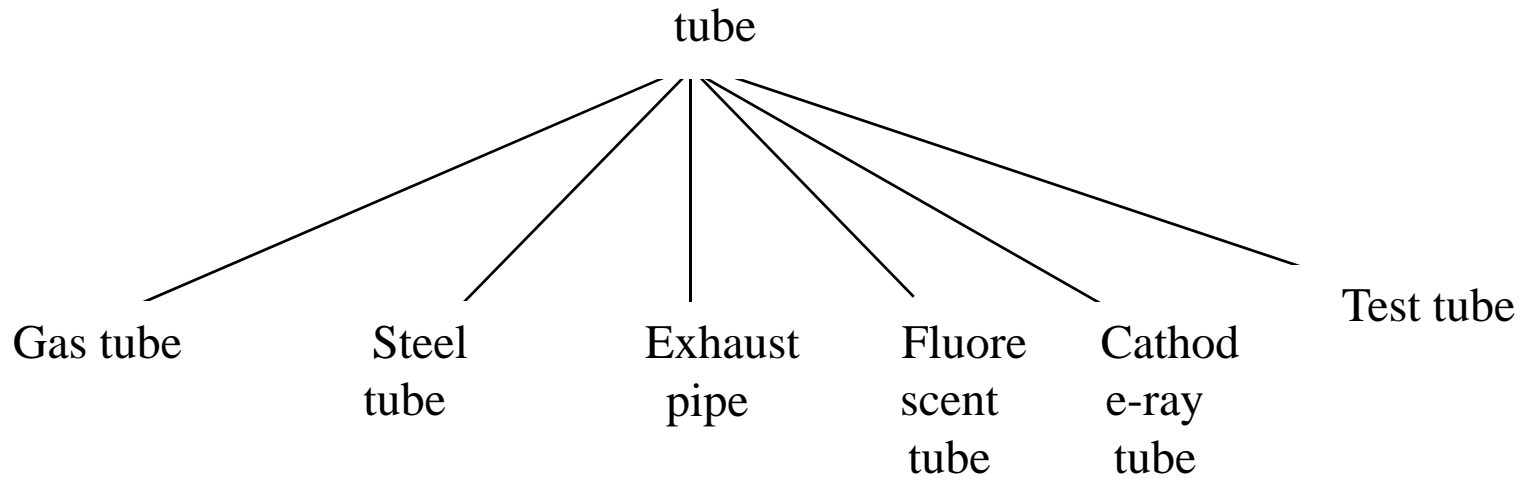
B. **PARTITIVE RELATIONS** (**part-whole** relation) PART_OF

- These relations are also called **meronimic** (HAS_PART)
 - *Car: wheels, seats, doors, boot, steering wheel, gearbox...*
- Different types of meronimic relations

Vertical Relation



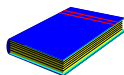
Horizontal Relation : co-hyponyms



Other Meronymic Relations

Relación	Ejemplo
componente - objeto	<i>pedal - bicicleta</i>
miembro - colección	<i>barco - flota</i>
porción - masa	<i>rebanada - pan</i>
material - objeto	<i>acero - coche</i>
fase - actividad	<i>pagar - comprar</i>
lugar - área	<i>oasis - desierto</i>

Tabla II.2: Modelo de Winston *et al.* (1987)



Climent, S. 1999 *Individuación e información parte-todo. Representación para el procesamiento computacional del lenguaje*

Non-hierarchical relations (*ad-hoc* relations)

- *Caused by* : (acid rain- nuclear explosion)
- *Product of*: (paper- wood pulp)
- *Property of* (compressibility -gas)
- *Quantitative measure* (temperature-heat)
- *Instrument for* (computer- data processing)
- *Counter-agent for* (insecticide- insects)
- *Container of* (toolbox- tools)
- *Method of* (diamond drilling- drilling)
- *Material for* (iron-bridge building)
- *Place for* (coal mine- coal exploitation)
- *Associated with* (production-consumption)

How do we express concepts?

- In **natural language** :
 - Terms (one-word or multi-word units) that denote or refer to a concept in a subject field
 - Definitions
 - Glosses, etc.
- In **artificial language**
 - Codes
 - Formulas
- In a **multimedia resource**:
 - Icons
 - Photos
 - Diagrams
 - Graphs
 - Video-clips
 - Audio-clips
 - Other multimedia representations

Term formation

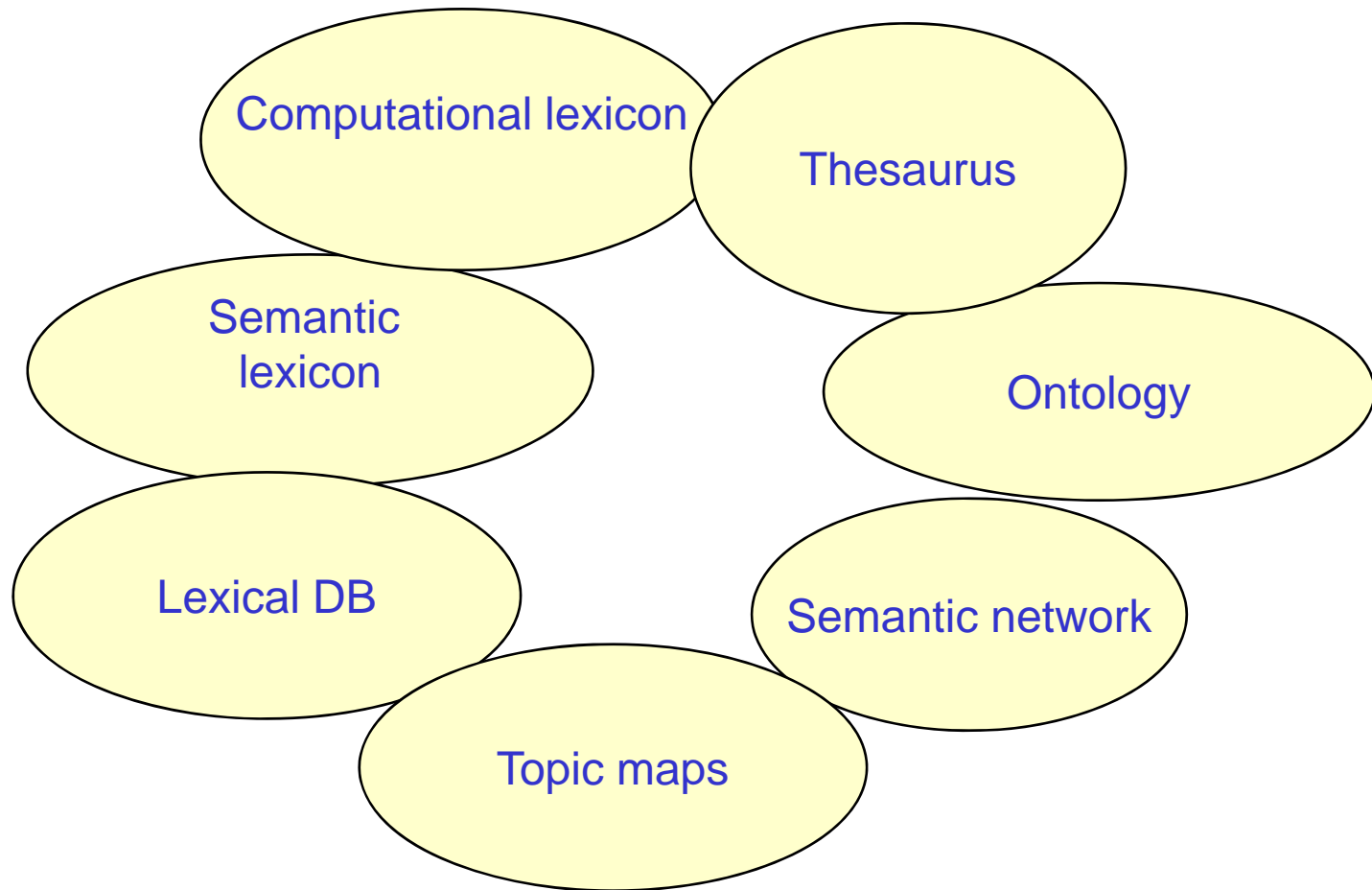
- According to its **origin**:
 - **Borrowings from other languages**: *hardware, software, football, cookies, folksonomies*
 - **Adapted borrowings**: *formatear, inicializar, fútbol, etc.*
 - **Loans of structure**: *inteligencia artificial, lógica difusa, programación orientada a objetos, anotación social*
 - **Semantic loans**: *aplicación, utilidades, editar, icono, ratón, menú,*
- According to its **formation**:
 - **One-word lexical units** or **terminological units**: *Programa, aplicación, icono, menú, ratón*
 - **Multiword lexical units**: *lenguaje de programación, lenguaje de alto nivel, programación orientada a objetos*
- According to its **components**:
 - **Suffixation**:
 - *teca/tica*: *animática, burórica, indumática, ofimática, robótica, telemática, turismática*
 - *itis*: a) *inflamación: bronquitis, faringitis, amigdalitis, otitis*
b) *obsesión: madriditis, mamitis, futbolitis*
 - **Prefixation**: *ciber: cibercafé, cibernauta, cibermedicina, ciberdelito, etc*
 - **Composition**: *salvapantallas, reposapiés, sujetamanos, radiotelevisión,*
 - **Abbreviation (acronyms)**: *PC, PDF, TCP/IP, MS-DOS, MP3, wysiwyg,*
 - **Conversion**: *download, input, output,*
 - **Neologization**: *autoedición, “gustomizar”,*
 - **Metaphorization**: *paquete de mejoras salariales, navegar, autopistas de la información, papelera, escritorio, bajar de la red, machacar un fichero, caerse el sistema, etc*

Terminology in technical & scientific domains

- No **professional communication** can exist without terminology
- No **knowledge transfer** can exist without terminology
- Without terminology there is not
 - Intellectual and material **development**
 - Professional **research and training**
- As a **consequence**,
 - no further development would take place
 - A country would isolate from the rest of developed countries

(Picht, 1979)

Linguistic and ontological resources: terminological chaos?



Some definitions

- **Thesaurus**: Organized controlled vocabulary of terms and their relations (hierarchical, etc.) in a subject domain.
 - **Conceptual Model**: Information developed by several people to represent a data model that allows the storage , retrieval and manipulation of semantic relations between concepts and scopes that may exist in some domain.
 - **Lexical Model**: the semantics of the grammatical units and other kinds of information
- word-sense-entry** →
- **CHARACTERISTICS** : [- ORTHOGRAPHY : string
 - **CHARACTERISTICS** : [- WORD-MEANING : word-meaning-id+
 - **CHARACTERISTICS** : [- SYNONYMS : word-meaning-id*
 - **CHARACTERISTICS** : [- NEAR-SYNONYMS : word-meaning-id*
 - **CHARACTERISTICS** : [- HYPONYMS : hyponym*
 - **CHARACTERISTICS** : [- HYPERONYMS : hyperonym*
 - **CHARACTERISTICS** : [- ANTONYMS : antonym*
 - **CHARACTERISTICS** : [- MERONYMS : meronym*
 - **CHARACTERISTICS** : [- HOLONYMS : holonym*
 - **CHARACTERISTICS** : [- QUANTIFICATION : quantification*
 - **CHARACTERISTICS** : [- COLLOCATIONS : collocation*
 - **CHARACTERISTICS** : [- SEMANTIC-FRAME : sem-frame
 - **CHARACTERISTICS** : [- ACTIONALITY : actionality
 - **CHARACTERISTICS** : [- ENTRY-CREATOR: (HUMAN | MACHINE)
 - **CHARACTERISTICS** : [- IS_VALIDATED: Boolean]

(phonologic, morphologic, syntactic, semantic and pragmatic) 

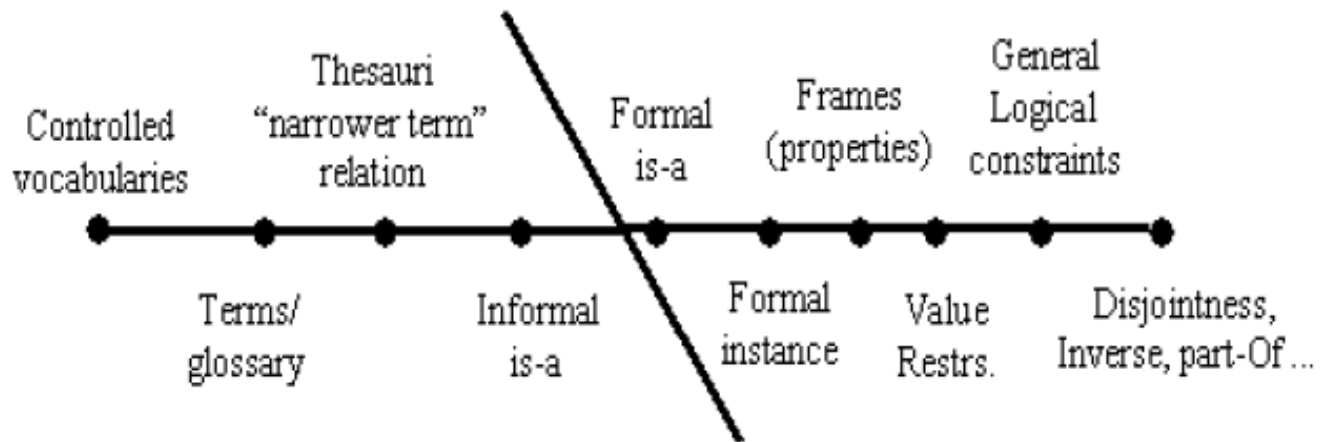
What is an ontology?

- “An ontology is similar to a dictionary or glossary, but with greater detail and structure that enables computers to process its content. (IEEE Standard Upper Ontology Working Group)
- “An ontology consists of a set of concepts, axioms, and relationships that describe a domain of interest.”
SUMO ontology <http://ontology.teknowledge.com/>

Ontology classification (Lassila & McGuinness)

Lightweight Ontologies

Heavyweight Ontologies



(Lassila & McGuinness, 2001)




- Gómez-Pérez, A., Fernandez-Lopez, M., Corcho, O. (2003) *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web*. Londres:Springer Verlag London Ltd.

Glossaries

- List of terms, generally with definitions.
- Terms usually belong to a concrete field
- Terms are defined according to the meaning in that domain
- Goal: to avoid polisemy.

U.S. ENVIRONMENTAL PROTECTION AGENCY

List of all Topics

 [Bookmark](#)

[Recent Additions](#) | [Contact Us](#) Search: [Advanced search](#)

You are here: [EPA Home](#) » [Browse EPA Topics](#) » List of all Topics

List of all Topics

Alphabetical List of All Topics

This page organizes topics into alphabetical order, like a book's index. You can also browse topics organized into [broad categories, like a book's table of contents.](#)

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

A	Back to Top
Abandoned Mine	Ecosystems > Mines > Abandoned Mine
Abandoned Properties	Cleanup > Storage Tanks > Abandoned Properties
Abatement	Air > Air Pollution Control > Abatement
Abatement	Treatment/Control > Air Pollution Control > Abatement
Above Ground Tanks	Industry > Storage Tanks > Above Ground Tanks
Accident Preparedness	Emergencies > Accidents > Accident Preparedness
Accident Prevention	Emergencies > Accidents > Accident Prevention
Accidents	Emergencies > Accidents
Accomplishments	Cleanup > Accomplishments
Acetone	Pollutants/Toxics > Soil Contaminants > Acetone
Acid Rain	Air > Air Pollution Effects > Acid Rain
Acute Exposure	Human Health > Exposure > Acute Exposure
Administrative Civil Enforcement	Compliance And Enforcement > Civil Enforcement > Administrative Civil Enforcement
Administrator	Environmental Protection Agency > Administrator
Advisories	Human Health > Advisories
Advisory Committees	Environmental Protection Agency > Science Advisory Board (SAB) > Advisory Committees

CINDOC Glossary

[Tesauros](#)[Inicio](#)[Alfabético](#)[Búsquedas](#)

Glosario de Máquinas Herramienta

ABCDEFGHIJKLMNOPQRSTUVWXYZ

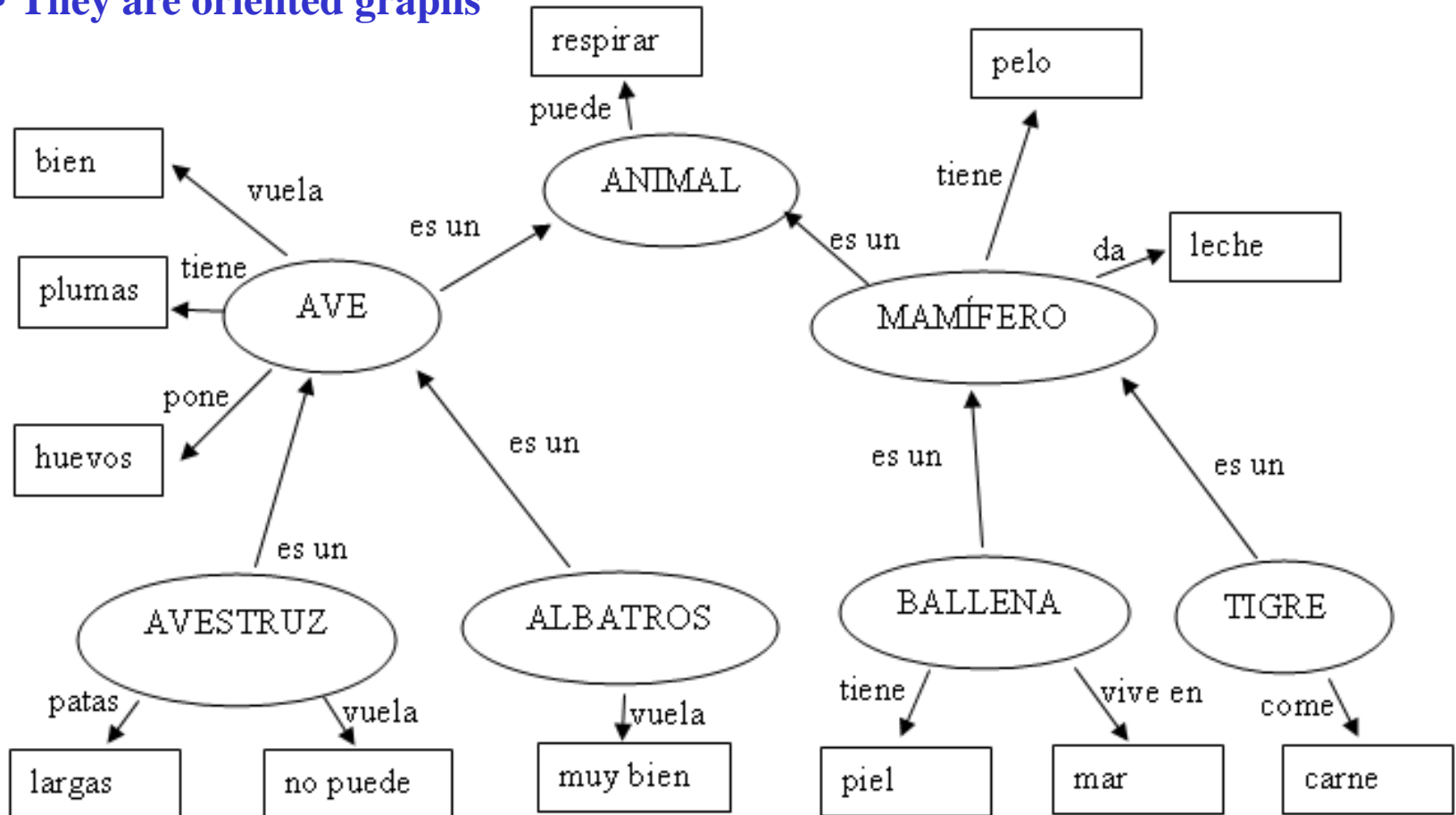
Listado alfabético de terminos [#1] *(no-descriptores en cursiva)*

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [Siguiente] [Fin]

a prueba de empleo incorrecto
abrazadera para tubos
accionamiento de la mesa
accionamiento del avance
accionamiento eléctrico
accionamiento forzado
accionamiento hidráulico
accionamiento individual
accionamiento neumático
accionamiento por cuerda
accionamiento por fricción
accionamiento por grupos
accionamiento por poleas escalonadas
accionamiento por trinquete
aceleración
acoplador roscado para tuberías
acoplamiento de desembrague
acoplamiento de ejes

Semantic network

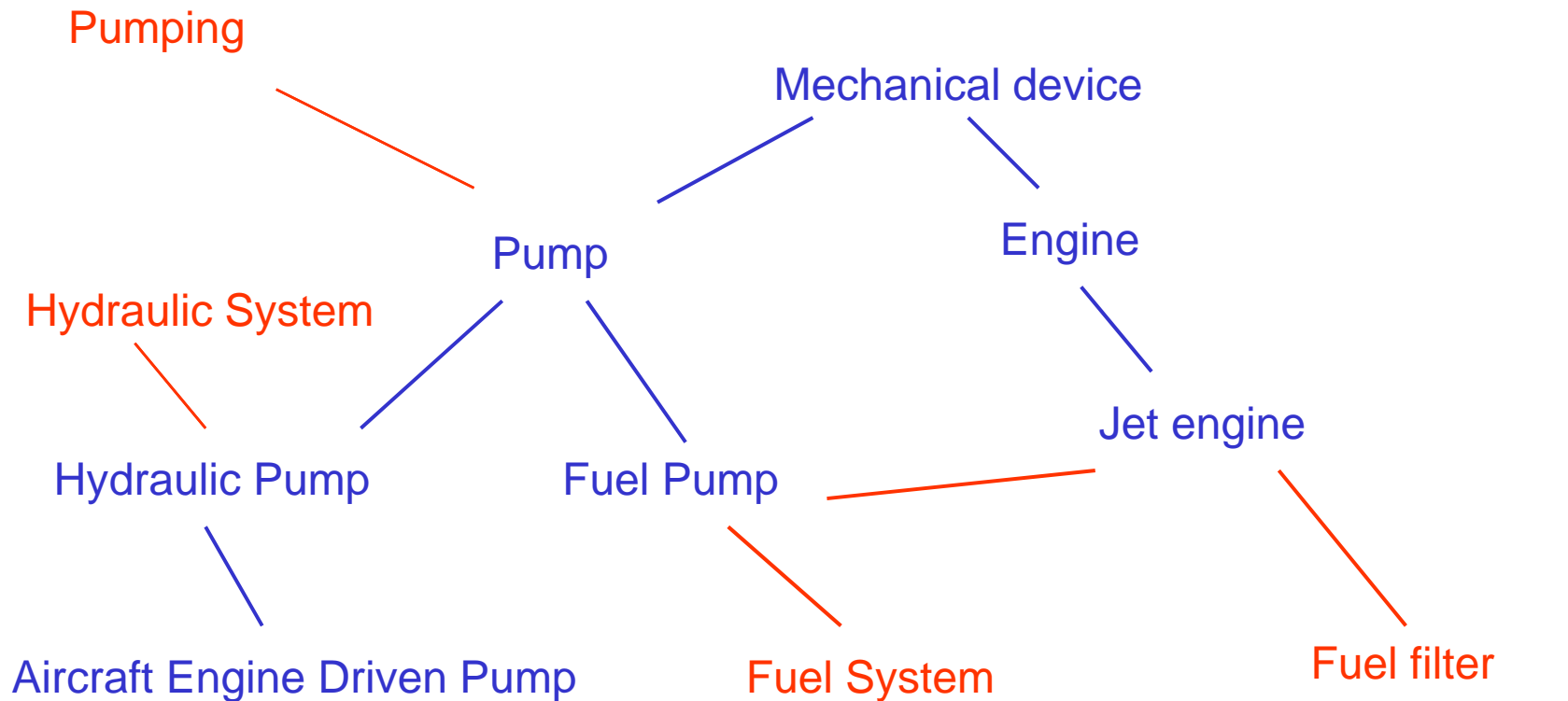
- Concepts structure with nodes and relations, not hierarchically organized
- Can include BT, NT, RT relations or other associative relations
- They are oriented graphs



A Thesaurus: Example

- Conceptual structure
- Semantically related terms by means of
 - **Explicitly defined** conceptual relations
 - **Restricted definition** of the terms

Related term: —



(Adapted from Uschold)

INSPEC Thesaurus

Types of data included

THESAURUS search words: **natural languages**

UF natural language processing (UF=used for natural language processing)

BT languages (BT=broader term is languages)

TT languages (TT=top term in a hierarchy of terms)

RT artificial intelligence (RT=related term/s)

computational linguistic
formal languages
programming languages
query languages
specification languages
speech recognition
user interfaces

CC C4210L; C6140D; C6180N; C7820(CC=classification code)

DI January 1985(DI=date [1985])

PT high level languages (PT=prior term to natural languages)

An excerpt from INSPEC Thesaurus

Cellular radio	used for (UF): cellular communication cellular telephones Groupe Speciale Mobile (GSM) microcellular radio pan-european radio vodafone
Land mobile radio Radiotelephony	These twp terms are broader terms (BT) to "cellular radio". If you searched under these terms, you will retrieve a larger set of documents
Radio applications Telecommunication	These terms are top terms (TT) in the hierarchy
Channel allocation Land mobile radio Personal communication networks Radio access networking Space division multiple access	All these terms are related terms (RT) to "cellular radio"
DI January 1985	Date when "cellular radio" was added
mobile radio systems	previous term (PT) used before 1985
B6250F; D4045	class codes

Table 1: Inspec Thesaurus (1999) -- an excerpt on cellular phones

nouns

round object that is hit or thrown or kicked in games; "the ball travelled 90 mph on his serve"; "the mayor threw out the first ball"; "the ball rolled into the corner pocket"

- ☐ Hypernyms (... is kind of)
- ☐ Hyponyms (kinds of ...)
- ☐ Antonyms (opposites of ...)
- ☐ Meronyms (parts of ...)
- ☐ Holonyms (... is part of)
- ☐ Related Verbs
- ☐ Related Adjectives

a solid ball shot by a musket; "they had to carry a ramrod as well as powder and ball"

an object with a spherical shape; "a ball of fire"

verbs

adjectives

SMART THESAURUS MUSIC is organized with respect to the semantic part of speech, which is supported by the lexical reference system: nouns, verbs, and adjectives.

Thus, three different tabs are presented to you. A simple click opens a certain tab, and, offers its content: a list of meanings, each representing a certain **synset** of the search term. In order to find out which element of the web refers to which meaning or synset, please click on it. Two things happen:

The meaning gets marked (with red color) and so do the corresponding elements of the web. A certain circle or sphere, representing a specific synset, becomes marked red, and also all of the edges that point to the set of synonyms (representing the synset). In addition, the 'meaning' opens its content and presents a list of lexical pointers associated with the selected part of speech. A click on one of these pointers, e.g. hypernym, lets you explore the broader terms associated with the selected synset.

SMART THESAURUS MUSIC supports the following lexical relationships:

Noun

- [1] Hypernym or broader term (...is a kind of)
- [2] Hyponym or narrower term (kinds of ...)
- [3] Antonym (opposites of ...)
- [4] Meronym (parts of ...)
- [5] Holonym (... is a part of)
- [6] Related verbs
- [7] Related Adjectives

Verb

- [1] Hypernym or broader term (...is a kind of)
- [2] Hyponym or narrower term (kinds of ...)
- [3] Related verbs
- [4] Related nouns

Types of relations

Lexicons

Generally, of two types

- *general*
 - contain language used in all/general contexts
- *specific*
 - contain the language used in a specific domain of knowledge

Implications of both types of lexicons

- Quantity and quality of information (granularity)
- Complexity of the design
- Complexity in the development process

How are lexicons used in NLP?

- They contain the necessary linguistic information to construct meaning representations

Lexico

n Account *n.* Domain [**financial**]
 Account *v.* ...
 Bank_1 *n* domain [**financial**]
 Bank_2 *n* domain [geography]
 Money *n.* domain [**financial**]

bank.....
account

 money.....



Went *v*past GO
 Go *v.* (NP_SUNJ ((role AGENT) (sem + animate))
 (VP ((verb GO)
 (PP ((prep TO)
 (NP ((role TARGET) (sem +loc))))))
 John *n.* sem: human
 Store *n.* sem: loc

Topic=
 financial
 field

John went to the store
 GO
 AGENT John TARGET store

Lexico

Went *v*past GO *n*

Go *v.* (NP_SUNJ ((role AGENT) (sem + animate))

(VP ((verb GO)

(PP ((prep TO)

(NP ((role TARGET) (sem +loc))))

John *n.* sem: human

Store *n.* sem: loc

Adapted from Nancy Ide

Types of lexicons

- Various types:

- **Morphosyntactic** Information :

- <http://www.mat.upm.es/~aries/description.html>

- <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQP-HTMLDemo/PennTreebankTS.html>



- **Semantic** Information:

- Semantic features: human being, animate, human, physical object, mental object
 - Different entries for different senses
 - Semantic relations: synonyms, quasi-synonyms, antonyms, etc.
 - Hierarchical relations: part_of, kind_of, etc...

- **Syntactic-semantic** Information : colocations

- Information about a **domain**

- **Definitions**



Lexicon about “Existence” (Faber and Mairal, 1999)

- 1.1 General: To exist / to continue to exist.
- 1.2 To begin to exist [*be, live*]
 - 1.2.1 To cause something to exist [*create, make*]
- 1.3 To exist in the perception of others [*appear*]
 - 1.3.1 To cause something to exist in the perception of others [*show*]
- 1.4 To exist in time (becoming real) [*happen*]
 - 1.4.1 To cause something to exist in time [*induce, provoke*]
 - 1.4.1.1. To cause something to exist in time in a particular way [*precipitate, hasten*]
 - 1.4.1.2 . To cause something to happen, making it possible [*allow, permit*]
 - 1.4.1.3. To cause something not to happen [*prevent, avoid, stifle, smother*]
- 1.5 To exist as something
 - 1.5.1 To exist as the representation of something else [*represent, express*]
 - 1.5.1.1 To cause something to exist as a representation of something [*copy, reproduce*]
 - 1.5.2 To exist as a part of something [*comprise, constitute*]
- 1.6 To begin to exist [*start, commence, be born*]
 - 1.6.1 To cause to begin to exist [*start, commence*]
 - 1.6.1.1 To cause to be born [*abort*]
 - 1.6.2 To begin to exist in the perception of others [*arise, form*]
 - 1.6.3 To begin to exist in time (becoming real) [*start, originate*]
 - 1.6.3.1 To cause something to begin to exist in time [*start, initiate*]
- 1.7 To continue to exist [*last, endure*]
 - 1.7.1 To stop something from continuing [*interrupt*]
- 1.8 To stop existing [*die*]
 - 1.8.1 To cause somebody/something to stop existing [*kill, murder*]
 - 1.8.2 To stop existing in the perception of others [*disappear, vanish*]
 - 1.8.2.1 To cause something to stop existing in the perception of others [*erase, delete*]
 - 1.8.2.2 To stop existing in time [*end, finish, cease*]
 - 1.8.2.3 To cause something to stop existing in time [*end, finish, cease*]





ARIES Natural Language Tools

The *ARIES Natural Language Tools* make up a lexical platform for the Spanish language. These tools can be integrated into NLP applications. They include: a large Spanish lexicon, lexical maintenance and access tools and morphological analyser/generator.

Non-exclusive, non-transferable licenses are available for the following components:

[The Prolog GRAMPAL analyser/generator](#)

A public domain demonstration system written in Prolog of our morphological treatment and lexicon. It includes a small demo lexicon, a DCG grammar for word formation and some predicates to test both analysis and generation. It runs under Sicstus Prolog 2.1.9.

The Prolog GRAMPAL dictionary

A collection of Prolog predicates suitable for use with the public domain GRAMPAL DCG grammar. It is capable of generating/recognizing well formed inflected forms for verbs, nouns and adjectives. It has no adverbs, determiners, conjunction, prepositions, etc. It does not treat clitic pronoun attachment nor derivatives.

The expanded ARIES dictionary

A collection of expanded entries (allomorphs) with morphological information. It contains a full set of morphemes dealing with clitic pronoun attachment (but without verb marking for correct attachments). It includes information about some derivative morphological processes (inflected adjectives from past participles and adverbs ended in "-mente" from adjectives).

The source ARIES lexical base

A collection of inflectional models, rules for off-line computing of allomorphs, unexpanded lemma entries, lexicalized irregular words. It is the most complete source of information we have available and the most useful for dictionary maintenance. A tool for expanding the source dictionary to the expanded dictionary is also provided. The current size of this lexicon is 38,500 lemma entries (21,000 nouns, 10,000 adjectives, 7,500 verbs and 500 auxiliary words) plus more than 600 inflectional morphemes.

Access tools

The C/C++ programming interface for lexical access to the ARIES dictionary: It is a set of tools and libraries to build trie indexes to the allomorph dictionary and to retrieve them by an application.

Morphological analyser

The C/C++ morphological analyzer that makes use of the lexical interface mentioned above. This permits to improve efficiency by integrating word segmentation with lexical access also. By now, it is a (pseudo)-unification chart based parser for context-free morphological grammars.

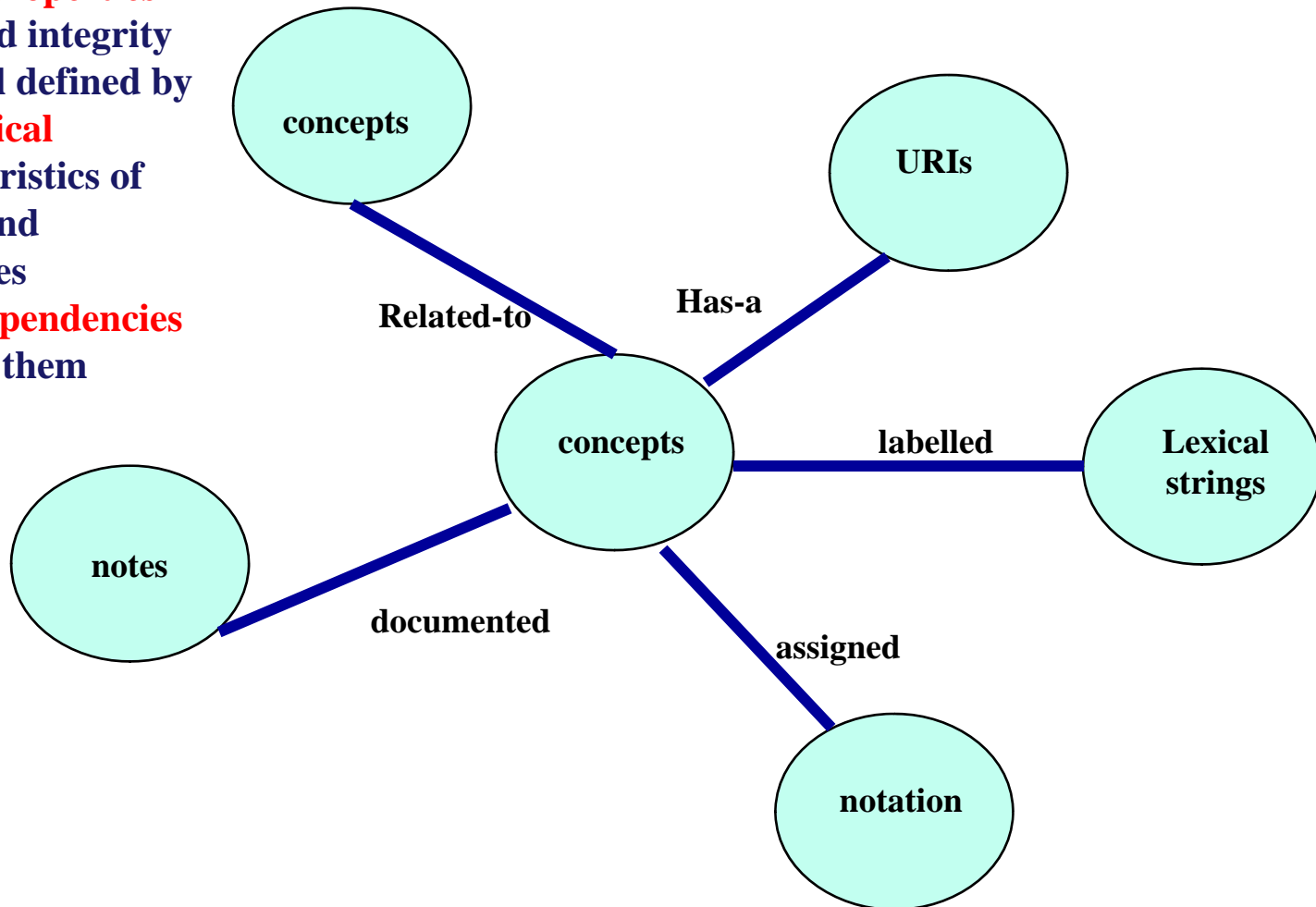
Transforming NOR into ontologies

SKOS

- Simple Knowledge Organization System
 - **Goal:** to enable easy publication of controlled structured vocabularies for the semantic web
 - Thesauri
 - Classification schemes
 - Subject heading systems
 - Taxonomies
 - Other ‘controlled language’
 - **How:** by using a common data model for sharing and linking knowledge organization systems
 - Simple, flexible, extensible, machine-understandable representation for sharing KOS
 - **BUT** SKOS is **not** a formal knowledge representation language.

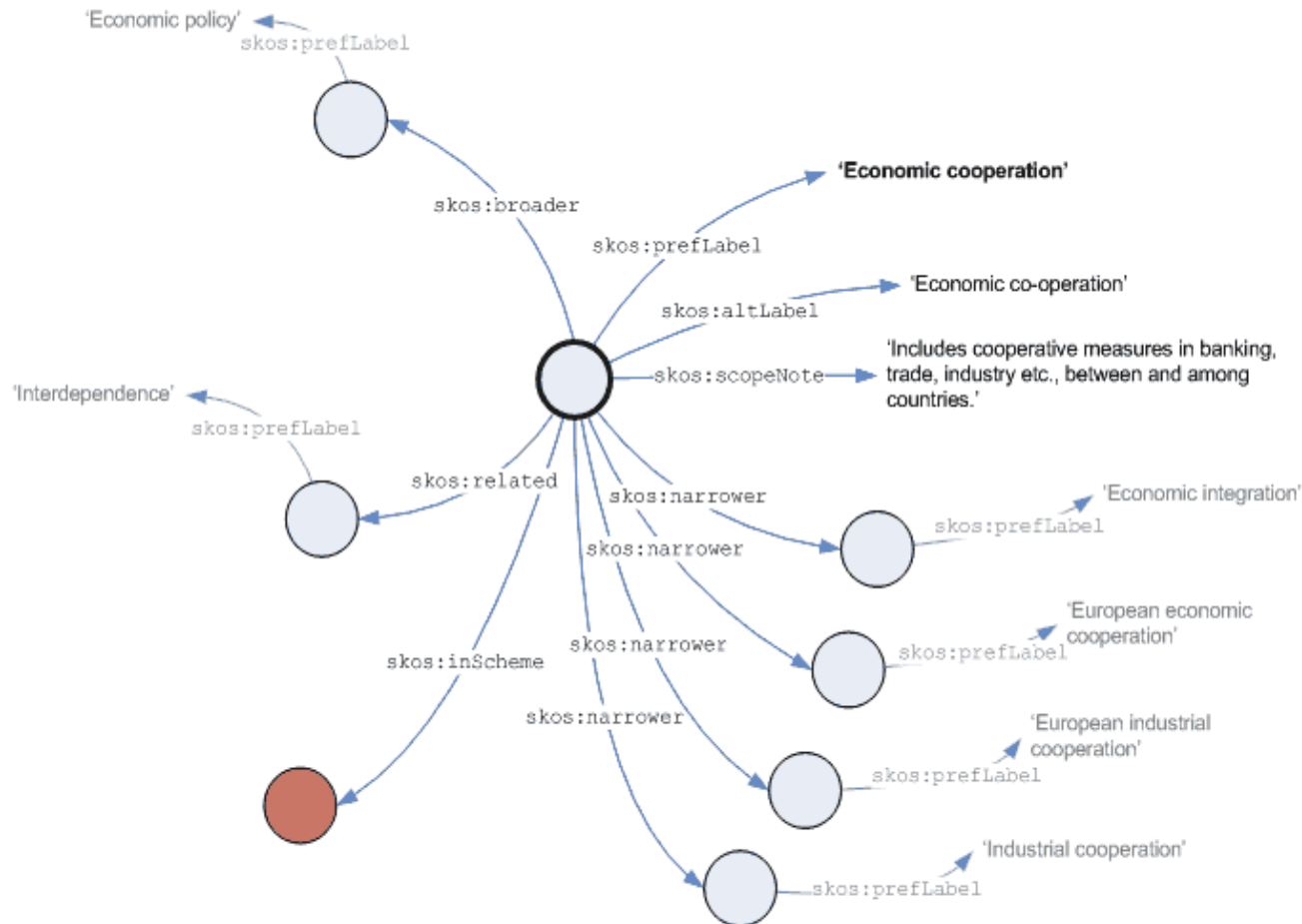
Elements in SKOS

- **classes** and **properties**
- structure and integrity of Data Model defined by
 - The **logical** characteristics of classes and properties
 - **interdependencies** between them



Graph from SKOS

Example in RDF



prefix skos: <http://www.w3.org/2004/02/skos/core#>

```
<A> rdf:type skos:Concept ;
      skos:prefLabel "love"@en ;
      skos:broader <B> .

<B> rdf:type skos:Concept ;
      skos:prefLabel "emotion"@en .

<C> rdf:type owl:Class ;
      rdfs:label "mammals"@en ;
      rdfs:subClassOf <D> .

<D> rdf:type owl:Class ;
      rdfs:label "animals"@en .
```

Applications and tools

- Applications

- AGROVOC and GEMET
- SWAD-Europe project framework. (Semantic web environmental directory)
- MeSH (converting thesauri to SKOS)
- used to represent geographical feature types in the GeoNames ontology.

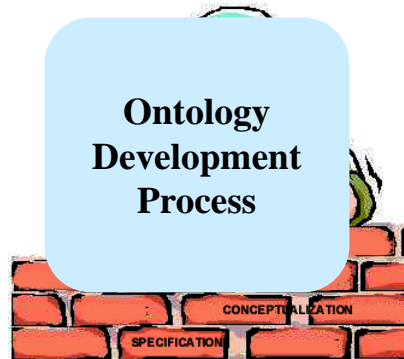
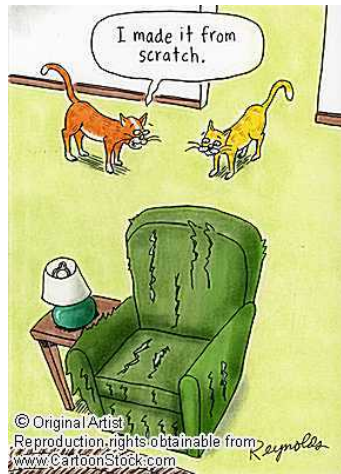
- Tools

- ThManager: Java open-source application for creating and visualizing SKOS vocabularies.
- W3C provides an experimental on-line validation service.
- Protege or Swoop

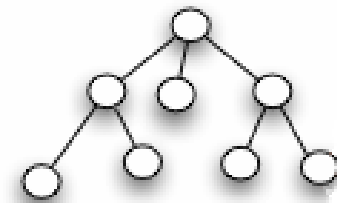
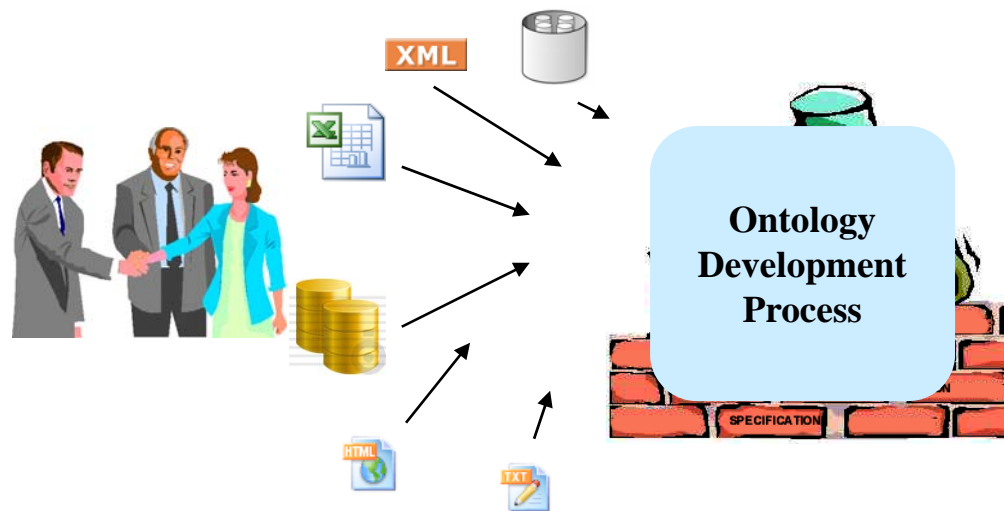
Transforming NOR into ontologies

García-Silva, Gómez-Pérez, Suárez-Figueroa, Villazón-Terrazas

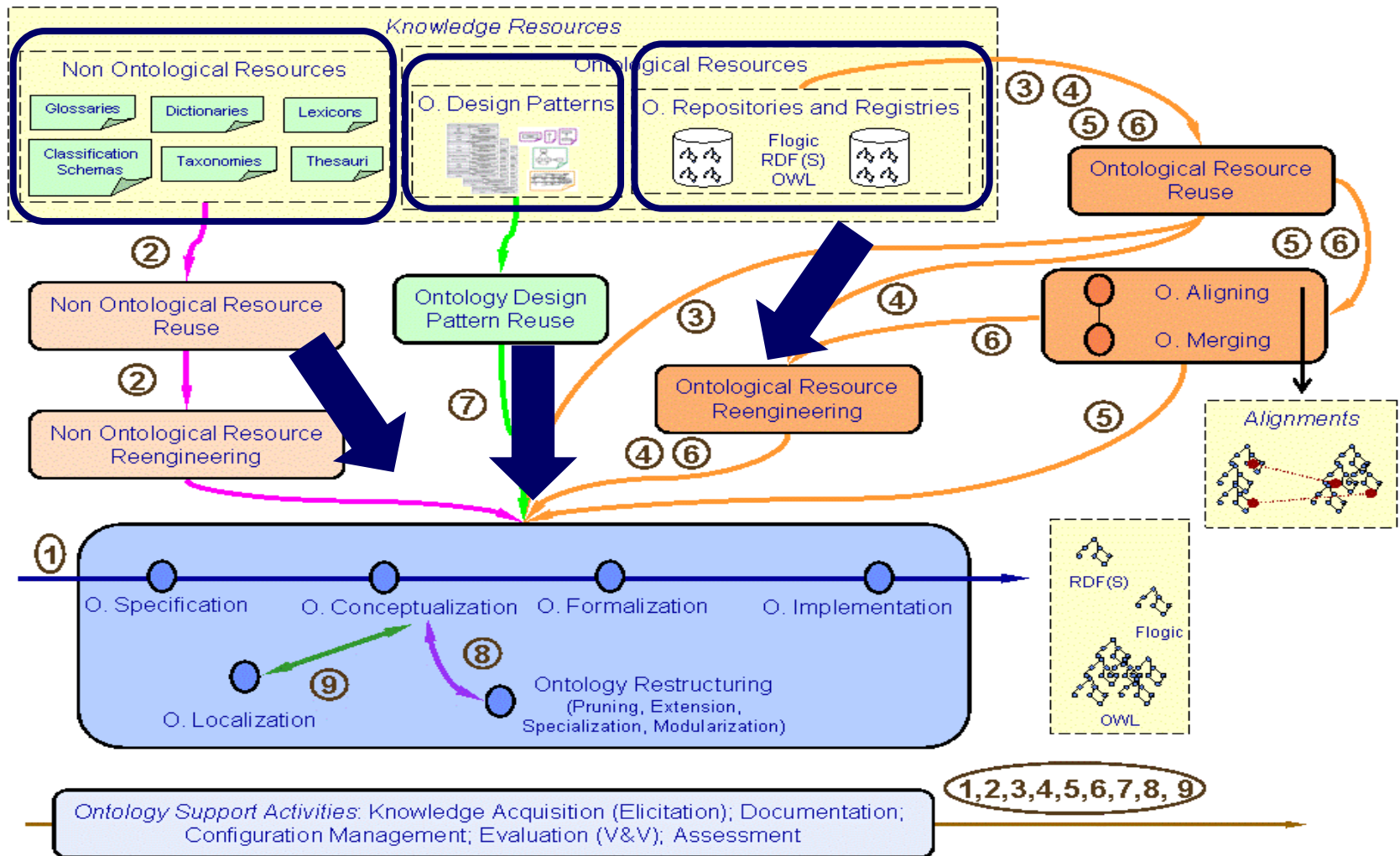
From scratch



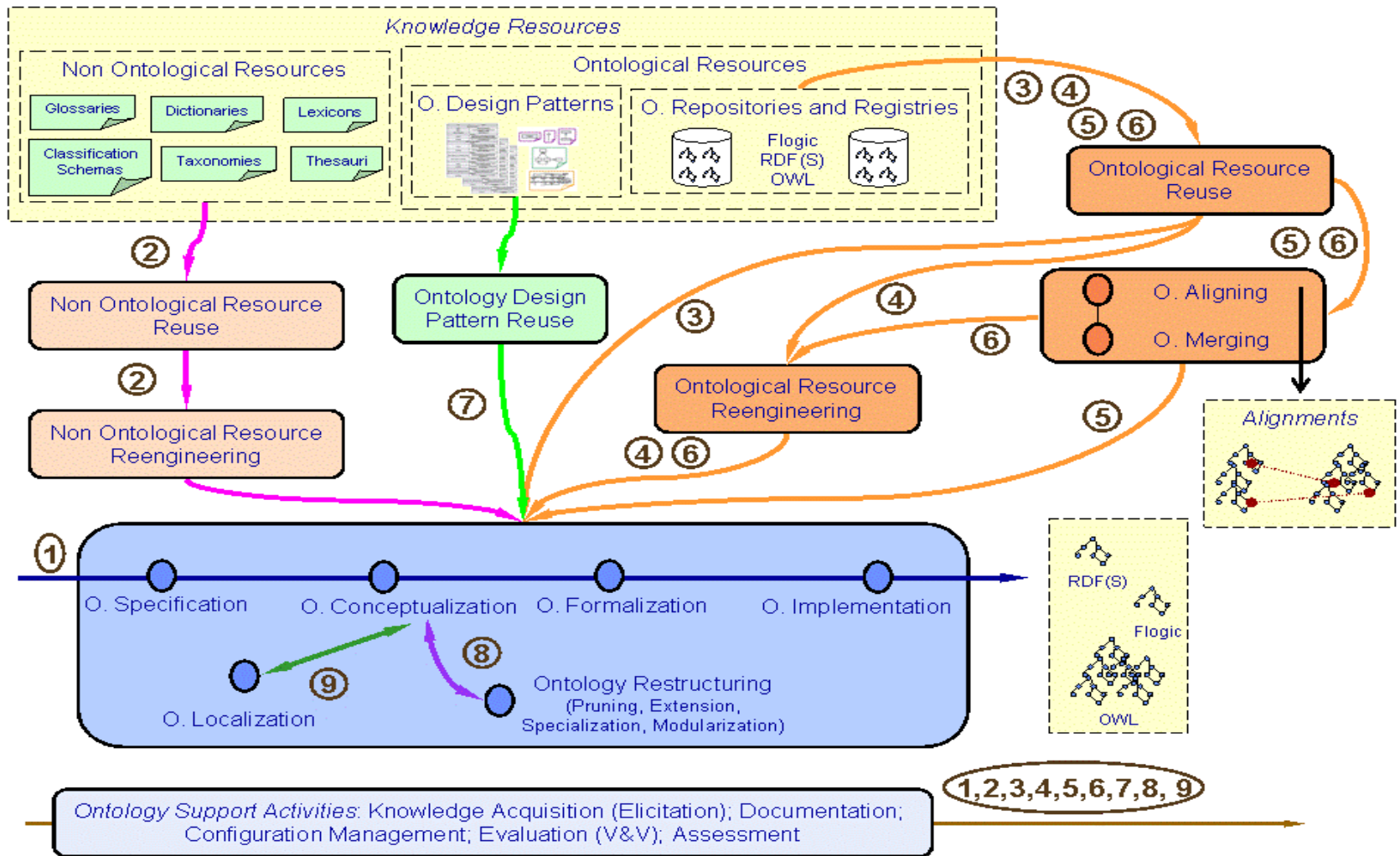
Re-engineering available resources




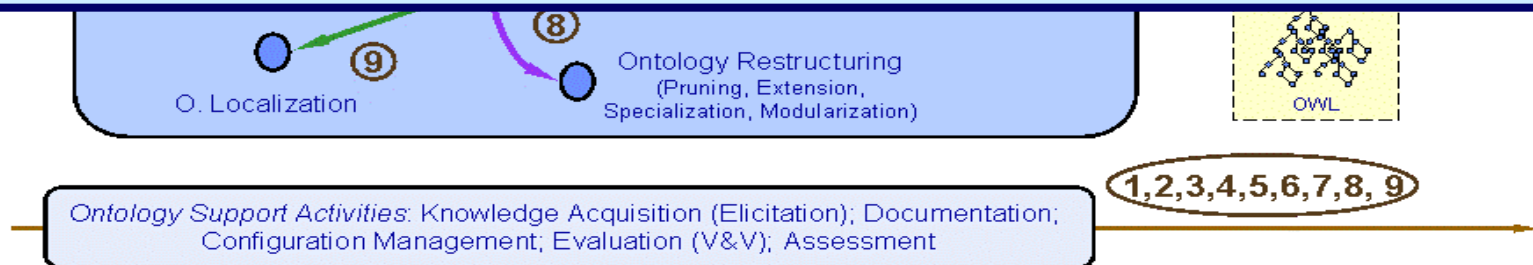
Introduction (II)



Introduction (II)



- 
- Knowledge Resources
- Non-Ontological Resources Ontological Resources
- **Non-Ontological Resource (NOR)** is an existing knowledge resource whose semantics has not yet been formalized by means of an ontology.
 - **Non-Ontological Resource Re-engineering** refers to the process of taking an existing non-ontological resource and transforming it into an ontology.



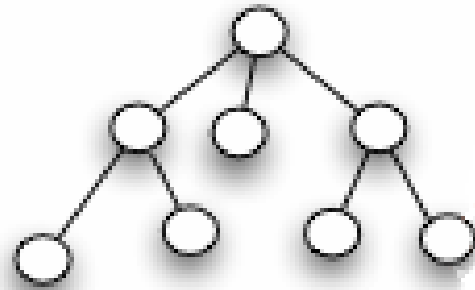
Motivation

resource



How?

I want to transform my non-ontological resource into an ontology



ontology

Types of Non-Ontological Resources

Maedche et al. 2001

- *Text*
- *Dictionary*
- *Knowledge base*
- *Relational schemata*

Sabou et al. 2007

- *Unstructured*
- *Semi-structured*
- *Structured*

Gangemi et al. 1998

- *Catalogue of normalized terms*
- *Glossed Catalogue*
- *Taxonomy*

Hodge 2000

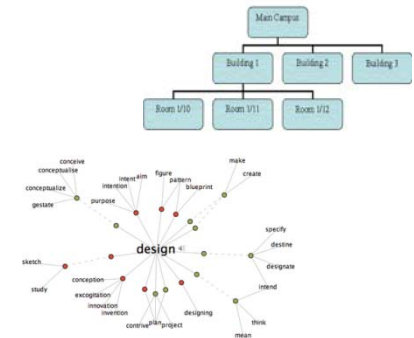
- *Terms Lists*
 - *Authority files*
 - *Glossaries*
 - *Dictionaries*
 - *Gazetteers*
 - *Lexicons*
- *Classification and categories*
 - *Subject headings*
 - *Classification schemes,*
 - *taxonomies and*
 - *categorization schemes*
- *Relationship lists*
 - *Thesauri*
 - *Semantic Newtorks*

They do not take into account the Non-Ontological Resource data model, an important artifact in the re-engineering process.

Non-Ontological Resource Re-engineering Approaches

Approaches focused on the NOR type

Approach	NOR Type	Data Model	Re-engineering Patterns	Tool
Heep et al. 2007	Classification schemes, thesauri	No	No	SKOS2GenTax
Mochol et al. 2006	Classification schemes	No	No	-
Sabou et al. 2007	Folksonomies	No	No	-
Sabou et al. 2007	Lexica	No	No	-
van Assem et al. 2004	Thesauri	No	No	-



Approaches focused on the NOR Implementation

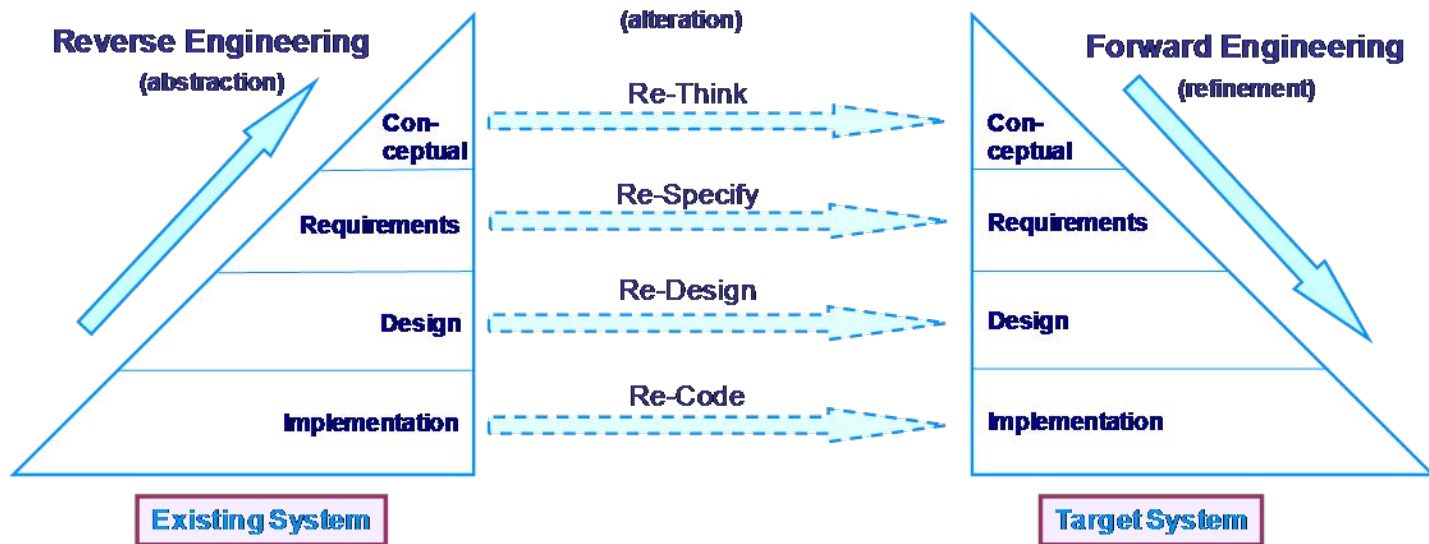
Approach	NOR Implementation	Data Model	Re-engineering Patterns	Tool
Stojanovic et al. 2002	Relational Database	No	No	KAON REVERSE
Barrasa et al. 2004	Relational Database	No	No	R2O, ODEMapster
García et al. 2005	XML Files	No	No	XSD2OWL, XML2RDF
Han et al. 2006	SpreadSheet	No	No	RDF123



They do not take into account the Non-Ontological Resource data model, an important artifact in the re-engineering process.

None of the approaches propose a set of re-engineering patterns to guide the re-engineering process.

Software Re-engineering



General Model for Software Reengineering

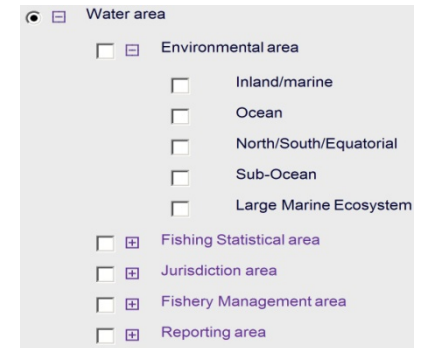
- 1) The conceptual level describes in general terms the functional characteristics of the system;
- 2) The requirement level is the specification of the problem being solved;
- 3) The design level is the specification of the solution; and
- 4) The implementation level refers to the coding, testing and delivery of the operational system



E.J. Byrne. A conceptual foundation for software re-engineering.
In Proceedings of the International Conference on Software Maintenance and Reengineering, pages 226–235.
IEEE Computer Society, 1992.

Example - FAO Water Areas Classification Scheme

Classification Scheme



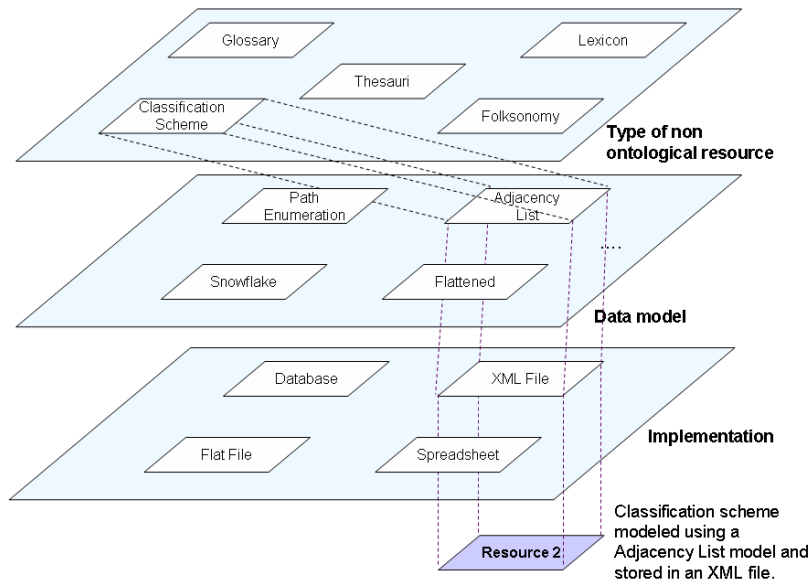
Adjacency List

ID	CSI_Name	Parent
20000	Water area	
21000	Environmental area	20000
24020	Jurisdiction area	20000
22000	Fishing Statistical area	20000
21001	Inland/marine	21000
21002	Ocean	21000
21003	North/South/Equatorial	21000
22001	FAO statistical area	22000
22002	Areal grid system	22000

XML

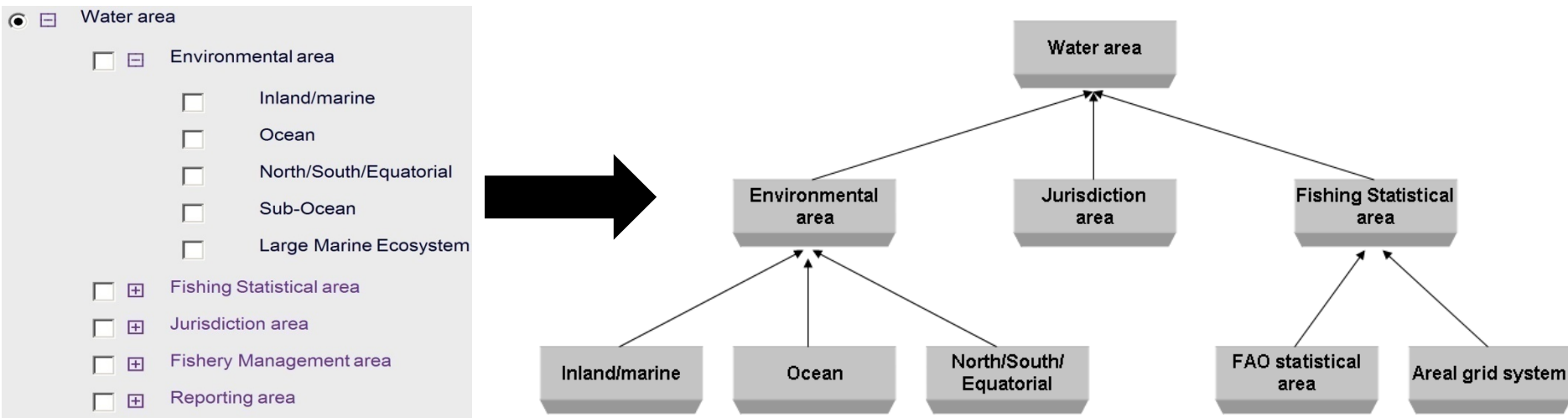
```

<Classification>
  <Category>
    <NodeId>20000</NodeId>
    <WaterCategory>Water Area</WaterCategory>
    <parentNodeId>1</parentNodeId>
  </Category>
  <Category>
    <NodeId>21000</NodeId>
    <WaterCategory>Environmental area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>22000</NodeId>
    <WaterCategory>Fishing statistical area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>24020</NodeId>
    <WaterCategory>Jurisdiction area</WaterCategory>
    <parentNodeId>20000</parentNodeId>
  </Category>
  <Category>
    <NodeId>21001</NodeId>
    <WaterCategory>inland/marine</WaterCategory>
    <parentNodeId>21000</parentNodeId>
  </Category>
  ...
</Classification>
  
```



Classification Scheme

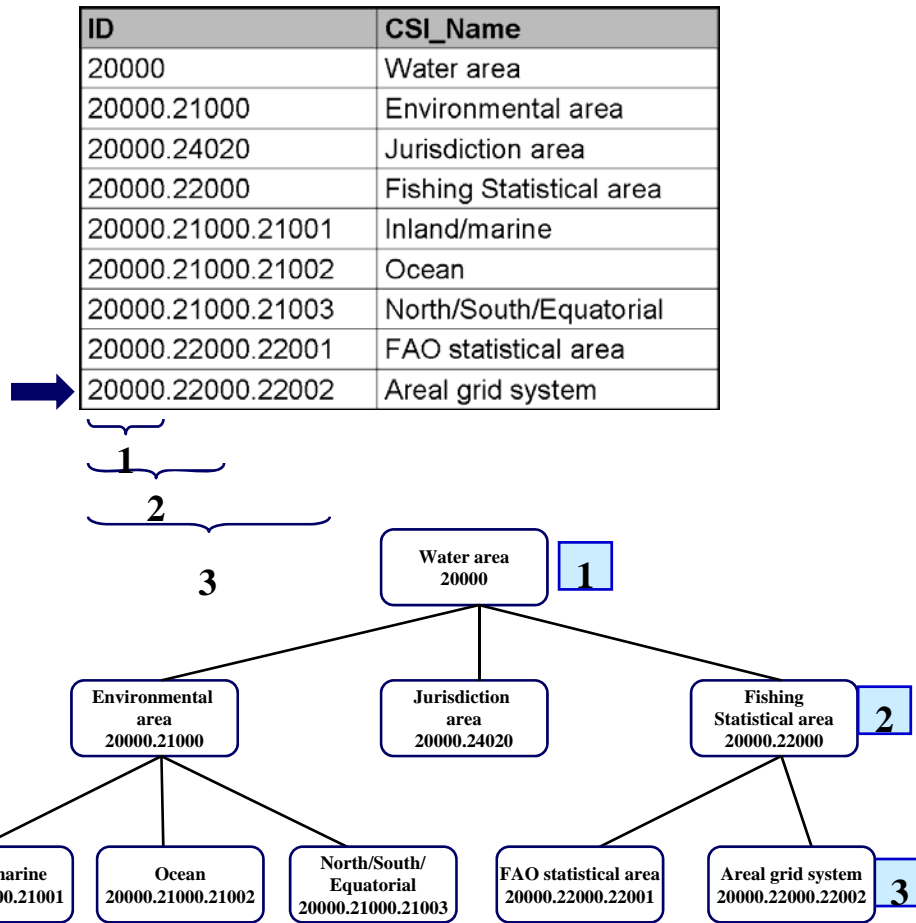
- A classification scheme¹ is the descriptive information for an arrangement or division of objects into groups based on characteristics, which the objects have in common. E.g. water area classification scheme².



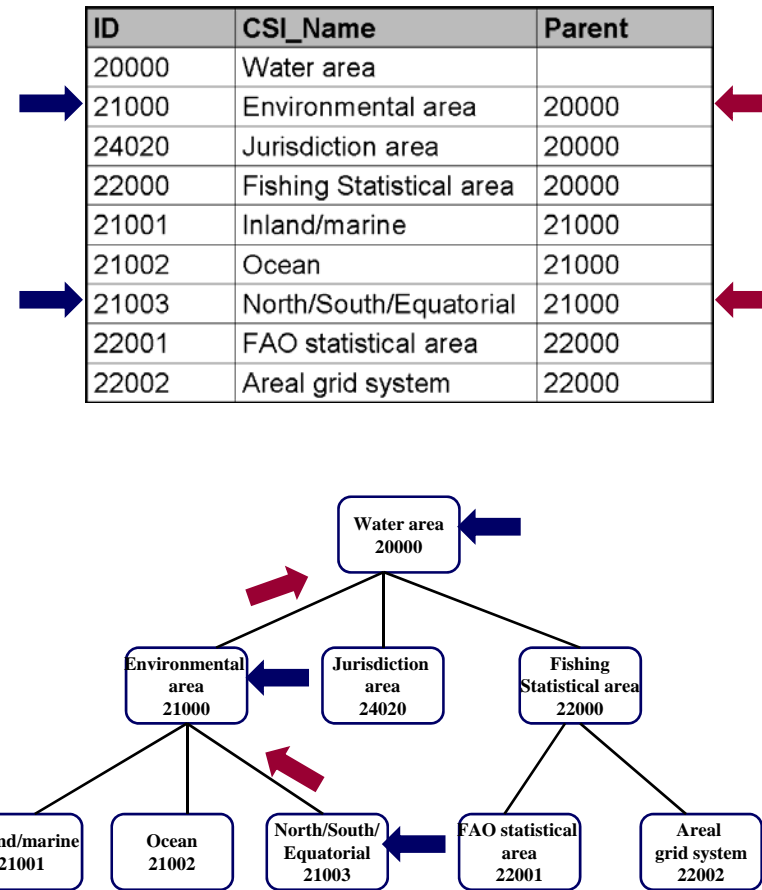
1. International Standard Organization (ISO). Information technology - Metadata registries – Part 1: Framework, 2004. Report ISO/IEC FDIS 11179-1.
2. <http://www.fao.org/figis/servlet/RefServlet>

Data Models (I)

- Path Enumeration Data Model** is defined as a model that stores for each node the path (as a string) from the root to the node.



- Adjacency List** is a recursive structure for hierarchy representations that comprises a list of nodes with a linking column to their parent nodes.



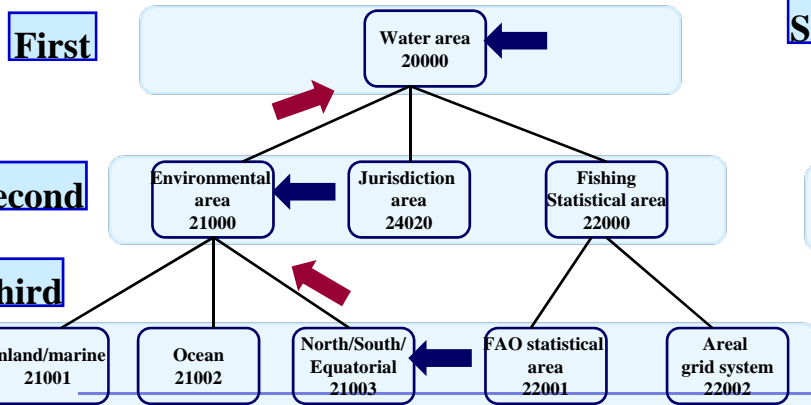
Data Models (II)

- Snowflake Data Model** is a normalized structure for hierarchy representations. For each hierarchy level a entity is created. In this model each hierarchy node has a column linked to its parent node.

First Level		
ID	CSI_Name	
20000	Water area	

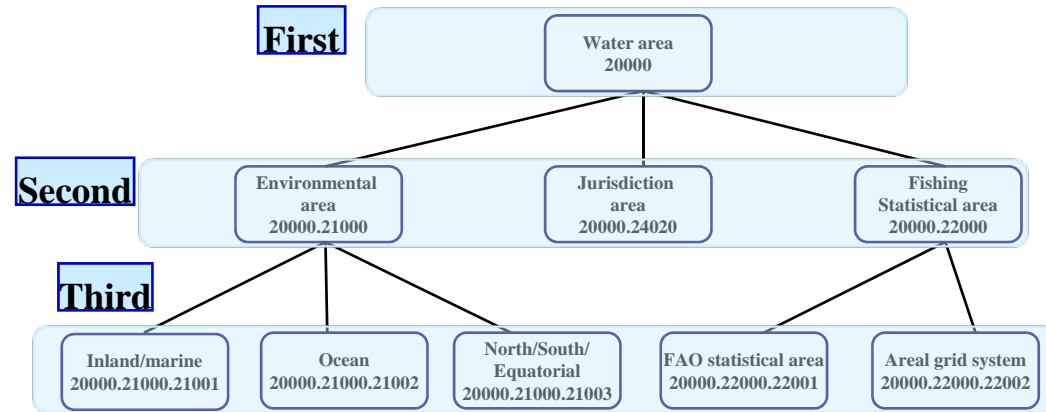
Second Level		
ID	First Level ID	CSI_Name
21000	20000	Environmental area
24020	20000	Jurisdiction area
22000	20000	Fishing Statistical area

Third Level		
ID	Second Level ID	CSI_Name
21001	21000	Inland/marine
21002	21000	Ocean
21003	21000	North/South/Equatorial
22001	22000	FAO statistical area
22002	22000	Areal grid system

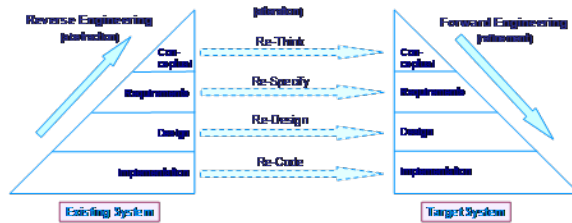
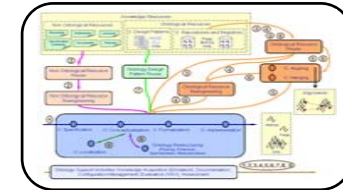


- Flattened Data Model**, is a denormalized structure. The hierarchy is represented with a table where each hierarchy level is stored on a different column.

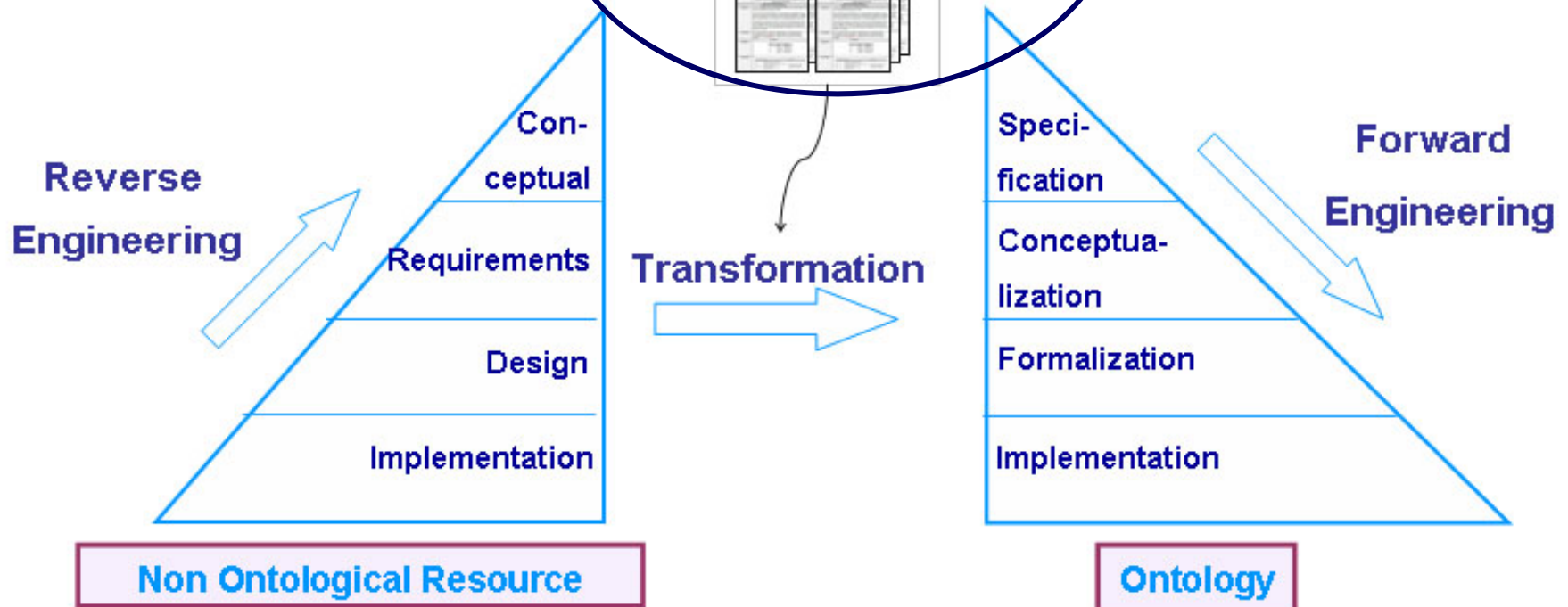
First Level		Second Level		Third Level	
ID	CSI_Name	ID	CSI_Name	ID	CSI_Name
20000	Water area	21000	Environmental area	21001	Inland/marine
20000	Water area	21000	Environmental area	21002	Ocean
20000	Water area	21000	Environmental area	21003	North/South/Equatorial
20000	Water area	22000	Fishing Statistical area	22001	FAO statistical area
20000	Water area	22000	Fishing Statistical area	22002	Areal grid system
20000	Water area	24020	Jurisdiction area		



NeOn Approach for Re-engineering NORs



General Model for Software Reengineering



PR-NOR Template

Slot	Value
General Information	
Name	Name of the component
Identifier	An acronym composed of: component type + abbreviated name of the component + number
Component Type	Pattern for Re-engineering Non-Ontological Resource (PR-NOR)
Use Case	
General	Description in natural language of the re-engineering problem addressed by the pattern for re-engineering non-ontological resources.
Example	Description in natural language of an example of the re-engineering problem.
Pattern for Re-engineering Non-Ontological Resource.	
Resource to be Re-engineered	
General	Description in natural language of the non-ontological resource.
Example	Description in natural language of an example of the non-ontological resource.
Graphical Representation	
General	Graphical representation of the non-ontological resource
Example	Graphical representation of the example of non-ontological resource.
Designed Ontology	
General	Description in natural language of the ontology created after applying the pattern for re-engineering the non-ontological resource.
Graphical Representation	
(UML)General Solution Ontology	Graphical representation, using the UML profile [BHO6] , of the ontology created for the non-ontological resource being re-engineered.
(UML)Example Solution Ontology	Example showing a graphical representation, using the UML profile [BHO6] , of the ontology created for the non-ontological resource being used.
How to Re-engineer	
General	Description in natural language of the general re-engineering process, using a sequence of activities.
Example	Description in natural language of the re-engineering process applied to the non-ontological resource example, using the above sequence of activities.
Implementation (Optional)	Link to a website which holds the code for an implementation, in a particular programming language, of the re-engineering process.
Relationships (Optional)	
Relations to other modelling components	Description of any relation to other PR-NOR patterns or other design patterns.

INPUT: Non-Ontological Resource

OUTPUT: Ontology

PROCESS: How

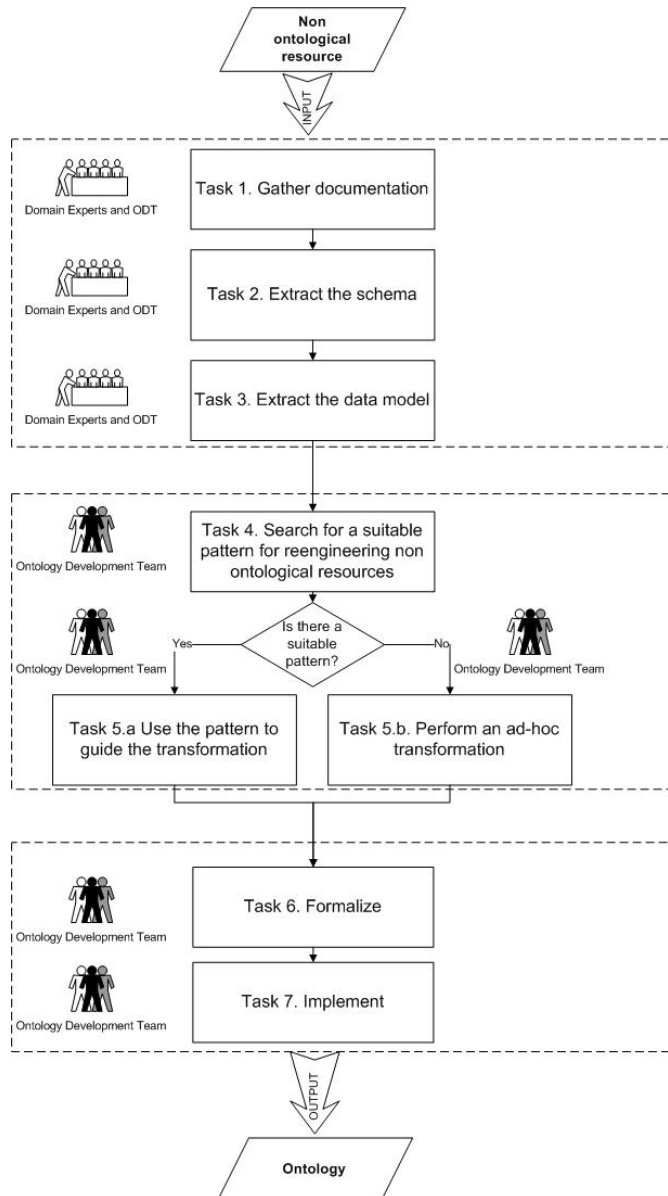
PR-NOR Template

INPUT: Resource to be Re-engineered	
General	Description in natural language of the non-ontological resource.
Example	Description in natural language of an example of the non-ontological resource.
Graphical Representation	
General	Graphical representation of the non-ontological resource.
Example	Graphical representation of the example of non-ontological resource.

OUTPUT: Designed Ontology	
General	Description in natural language of the ontology created after applying the pattern for re-engineering the non-ontological resource.
Graphical Representation	
(UML) General Solution Ontology	Graphical representation, using the UML profile [BH06], of the ontology created for the non-ontological resource being re-engineered.
(UML) Example Solution Ontology	Example showing a graphical representation, using the UML profile [BH06], of the ontology created for the non-ontological resource being used.

PROCESS: How to Re-engineer	
General	Description in natural language of the general re-engineering process, using a sequence of activities.
Example	Description in natural language of the re-engineering process applied to the non-ontological resource example, using the above sequence of activities.

Non-Ontological Resource Re-engineering Process (I)



Activity 1
Non-Ontological Resource
Reverse Engineering

Activity 2
Non-Ontological Resource
Transformation

Activity 3
Ontology
Forward Engineering

Classification Scheme Example

resource

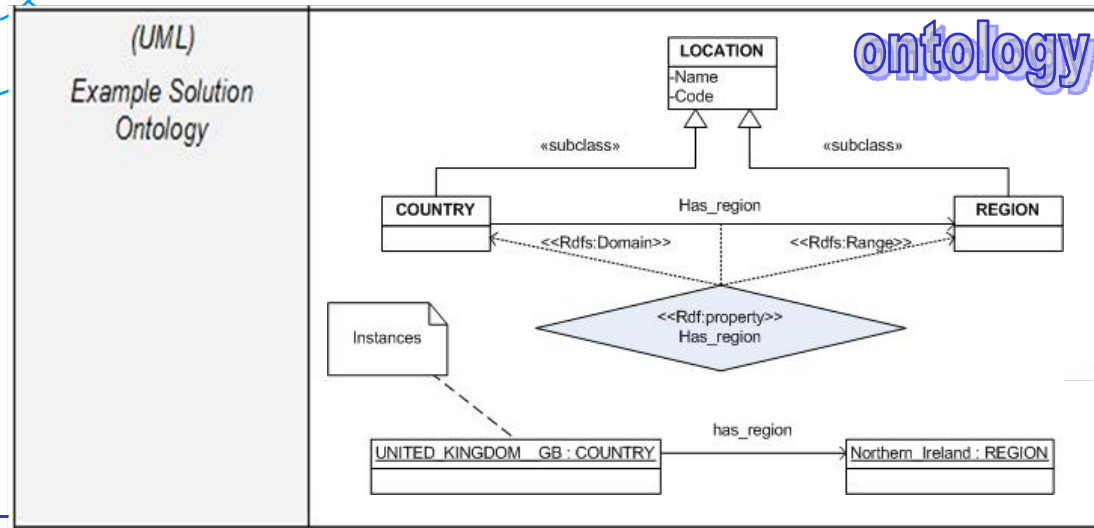
ISO 3166-1 Country	
Code	Name
GB	UNITED KINGDOM
ES	SPAIN
..	..

ISO 3166-2 Subdivision		
Code	Name	ISO 3166-1 Code
GB-NI	Northern Ireland	GB
GB-EA	East Anglia	GB
..

Algorithm

1. Create a COUNTRY class for the ISO 3166-1 Countries entity and a REGION class for the ISO 3166-2 Subdivisions entity.
2. Create the *Has_region* binary relation with COUNTRY as domain and REGION as range.
3. Create a LOCATION class and assert that COUNTRY and REGION are *subClassOf* LOCATION.
4. For each record on the ISO 3166-1 Countries entity create an instance of the COUNTRY class.
5. For each COUNTRY instance look for its REGION on the ISO 3166-2 Subdivisions entity and create an instance of REGION for each subdivision found. Also create an instance of the *Has_region* relation associated to the current country instance and related to the current region instance.

I want to transform my snowflake-based classification scheme into an ontology



Terminology and ontologies

- **Terminology** helps in the **knowledge organization** by establishing relations between terms and concepts
- An **ontology** is a conceptualization or representation of a domain, agreed by experts and readable by a machine
- Approach oriented to the **communication** among users of an organization
- **Terminological change**: from terminological data bases to terminological and knowledge/ontological data bases : terms in context and with relations among them
- **Terms retrieved from texts** can be the **starting point** in the development of ontologies

Ontologies in the terminological field

- Technologies that allow:
 - Acquiring and modelling knowledge
 - Browsing and recovering information
 - Representing and visualizing information
 - Interpreting and sharing knowledge
- Systems that include semantics and context by means of ontologies and textual corpora (Termontography)
- Current tools make interoperability and semantic web applications easier

Comparison factors

	Terminology	Ontologies
Level of formality in the definition	Text in NL	Formal language without ambiguities
Computer support	Terminological bases with few relations among concepts	Sound knowledge representation languages with relations among concepts
Users	Translators Domain experts Linguistic mediators Text editors	Information interchange between people and machines
Language	NL for expressing knowledge with precision	Labels for naming concepts have less importance

Collaboration between terminologists and ontologists

Ontology developers	Terminologists
Requirements specification	Requirements specification
Knowledge extraction	Knowledge extraction
Knowledge conceptualization	Knowledge conceptualization
Formalization with SRC based on frames or descriptive logic	Description in controlled NL to satisfy ontology engineering requirements
	Multilingual analysis

Comparison of lexical resources

CLASSIFICATION CRITERIA	GLOSSARY	DATABASE	DICTIONARY	THESAURUS	LEXICON	ONTOLOGY
Organization	alphabetical order	alphabetical order	alphabetical order	semantically + generically related lexical entries	semantically related lexical entries	semantically related lexical entries
Semantic information	definition in NL	definition + other kinds of info. in NL	definition + pos + etymologies + derivation + usage examples in NL	hierarchical, associative, equivalent relationships	explicit hierarchy (synonymy, antonymy, meronymy...) + grammatical + contextual information	explicitly defined hierarchy relationships around a unique concept
Physical format	paper + electronic format	electronic format	paper + electronic format	paper + electronic format	electronic format	electronic format (readable also by machines)
Domain of knowledge	general + specific	general + specific	general + specific	specific	general + specific	general + specific (agreed by domain experts)

Recommended papers

Terminology and ontologies, March 2009

- Van Assem, Malaisé, Miles & Schreiber: “A method to convert a thesaurus to SKOS”.
<http://www.cs.vu.nl/~guus/papers/Assem06b.pdf>
- Van Assem, Menken, Schreiber, Wielemaker & Wielinga: “A method for converting thesauri to RDF/OWL
<http://www.cs.vu.nl/~guus/papers/Assem04a.pdf>
- Lauser, Sini, Lian, Keizer and Katz
<ftp://ftp.fao.org/docrep/fao/009/ah801e/ah801e00.pdf>
- Campbell, Oliver, Packman & Shortliffe “Representing thoughts, words and things in UMLS.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61323>

Terminology in ontologies and other lexicographic resources

Dra. Guadalupe Aguado de Cea

lupe@fi.upm.es

<http://www.oeg-upm.net>

Ontological Engineering Group

Facultad de Informática

Universidad Politécnica de Madrid

Campus de Montegancedo sn,

28660 Boadilla del Monte, Madrid, Spain