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?
- Terminology in scientific domains
- Object of study of terminology
- Concepts and their relations
- Terms and their formation
- Linguistic resources and ontologies
- Reusing non ontological resources
- Final remarks

What is terminology?

The term terminology is polysemic:

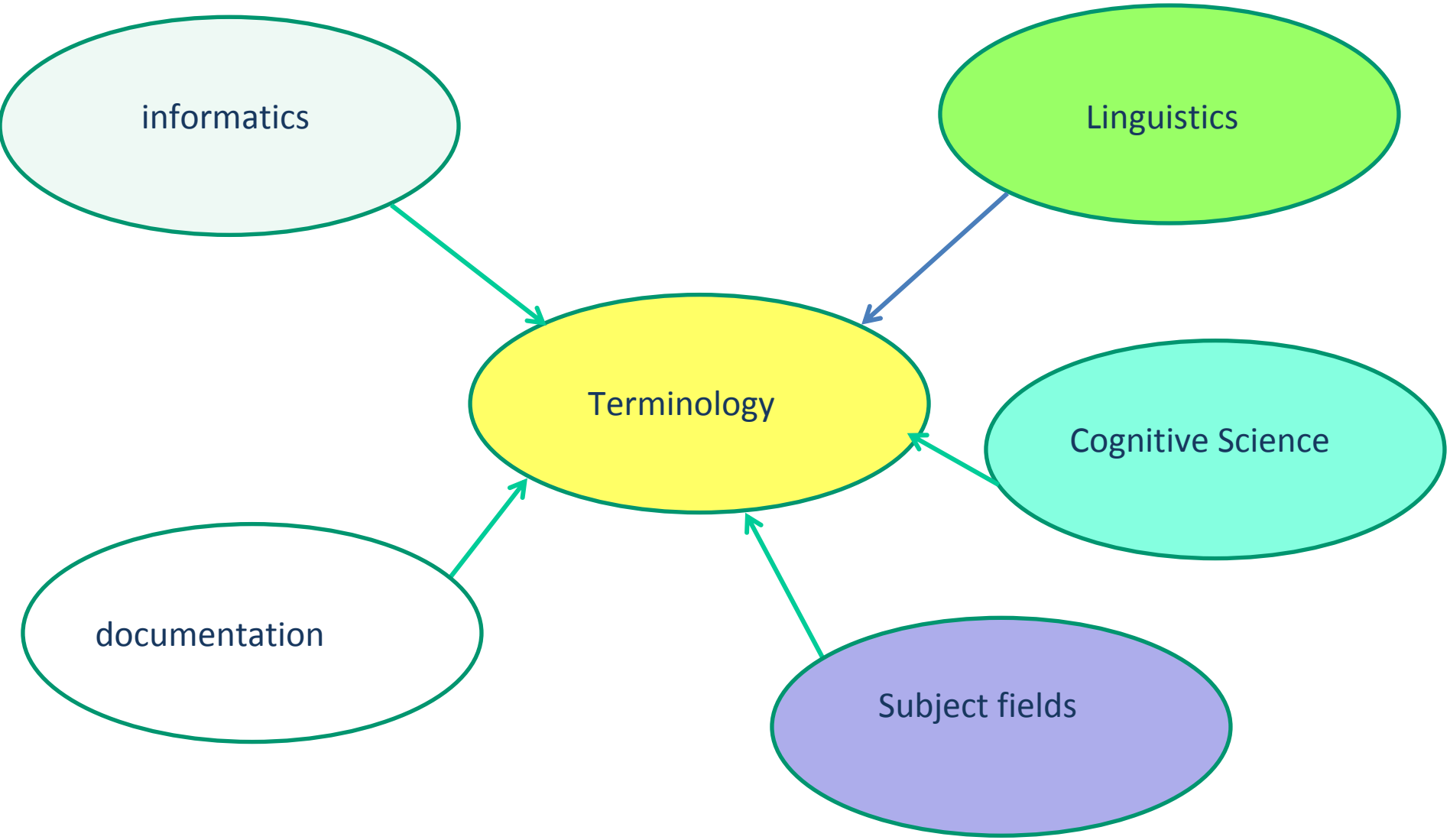
- **As a product: set of terms from a given subject field.**
- **As a discipline: set of fundamental principles and conceptual bases that govern the study of specialized terms, their description, analysis and relations.**
- **As a practice: set of principles oriented towards term compilation**

Terminology in technical & scientific domains

- No professional communication can exist without terminology
- No knowledge transfer can exist without terminology
- Without terminology there is not
  - –Intelectual 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)

Terminology: interdisciplinary subject field



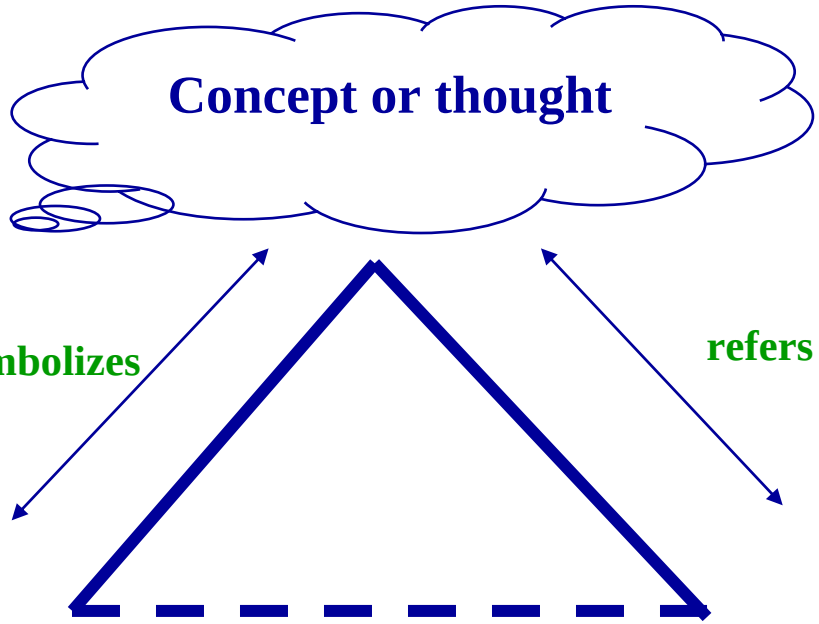
Object of study in terminology

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

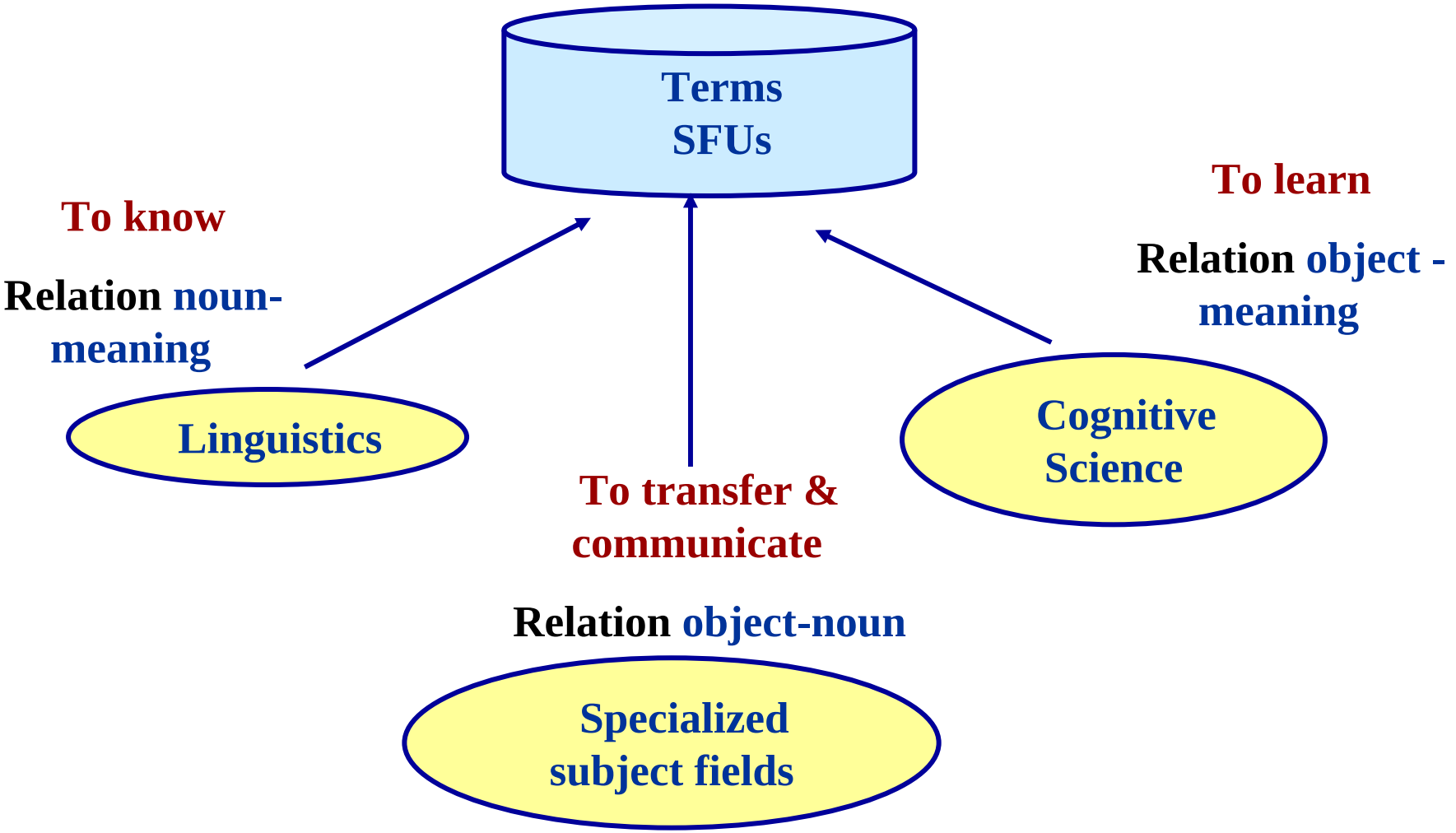
NOWADAYS

- Emphasis on terminology management:
  - products, tools and applications

Semantic triangle by Ogden & Richards



The object of study of terminology (1)



The object of study of terminology (2)

- identifying concepts and concept relations
- analysing and modelling concept systems on the basis of identified concepts and concept relations
- establishing representations of concept systems through concept diagrams
- defining concepts
- attributing designations (predominantly terms) to each concept in one or more languages
- recording and presenting terminological data, principally in print and electronic media (terminography)

ISO FDIS 704 :2009

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 (2009)
- BUT concepts are influenced by social and cultural circumstances given at a certain moment
  - SO this can lead to different classifications in the conceptual system
- 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

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.*

Characteristics 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, metalic, …
    - **Colour**: red, blue, green, orange…
    - **Position:** vertical, hanging, slanting
  - **extrinsic,** relation of the object with others
    - **Mode of employement** or application: *analogic*, *digital*, *hybrid.*
    - **Origin** or how an object comes into existence: producer, inventor, provider, the place of its production, (town, country), …

Concept: abstraction based on the set of all the characteristics of mechanical mice

Term: mechanical mouse

- a device;
- ivory-coloured;
- hand-manoeuvred along a firm, flat surface;
- has a ball on its underside;
- has three buttons;
- has a wire for connecting to a computer;
- rollers detect the movement of the ball;
- the ball controls the movement of a cursor on a computer display screen.

- a device;
- blue and grey;
- hand-manoeuvred along a firm, flat surface;
- has a ball on its underside;
- has two buttons;
- has a wire for connecting to a computer;
- without rollers;
- the ball controls the movement of a cursor on a computer display screen.

- a device;
- black-grey;
- hand-manoeuvred along a firm, flat surface;
- has a ball on its underside;
- has two buttons;
- has a wire for connecting to a computer;
- rollers detect the movement of the ball;
- the ball controls the movement of a cursor on a computer display screen.

ISO FDIS 704:2009

Relations between concepts: hierarchical relations

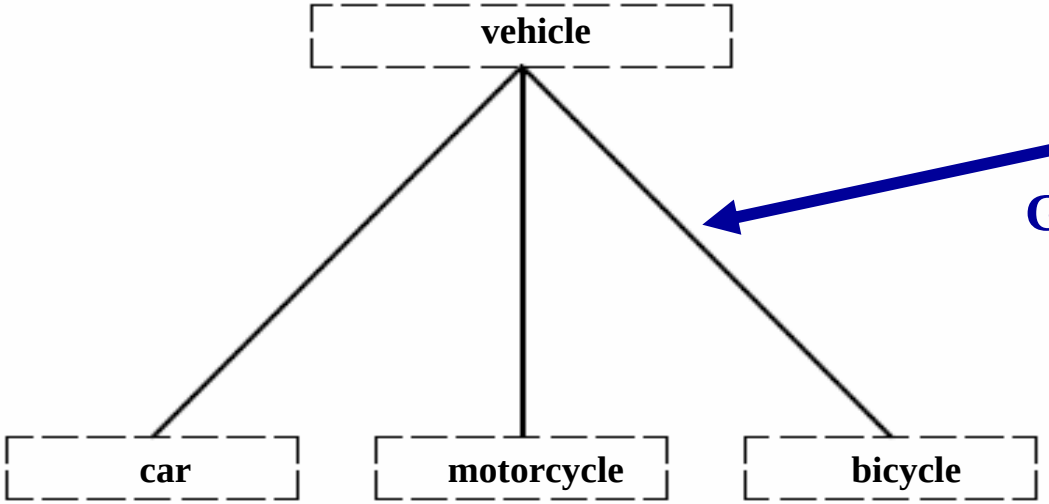
Close relation between a concept and its characteristics

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)

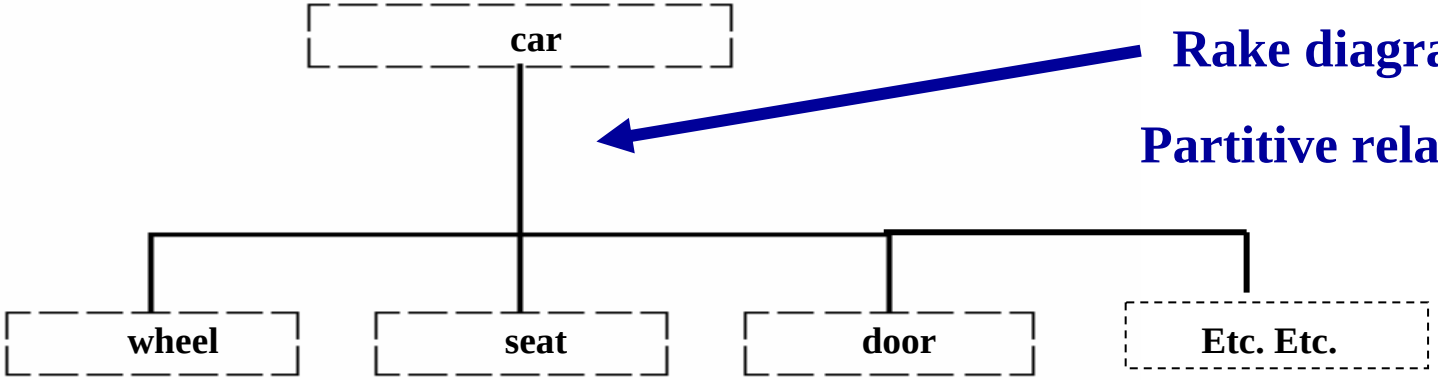
B. PARTITIVE RELATIONS (part-whole relation) PART_OF

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



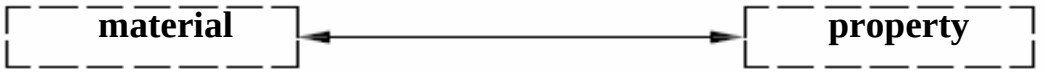
Tree diagram

Generic concept relations



Rake diagrams

Partitive relations



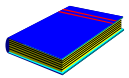
Line with arrowheads

Associative relations

Other Meronimic Relations

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

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 (associative 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)**

An *associative relation* exists when a thematic connection can be established between *concepts* by virtue of experience.

How do we express concepts?

- In natural language :
  - Terms (one-word or multi-word terms) 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 I

- 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, kindergarten= jardín de infancia*
  - – Semantic loans: *aplicación, utilidades, editar, icono, ratón, menú, semantic **grid***
  - –Transliteration: Pekin- Beijin

- According to its **formation:**
  - – One-word terminological units : *Programa, aplicación, icono, menú, ratón*
  - – Multiword terminological units : *programming language, computer assisted design/learning, high level language, object-oriented programming*

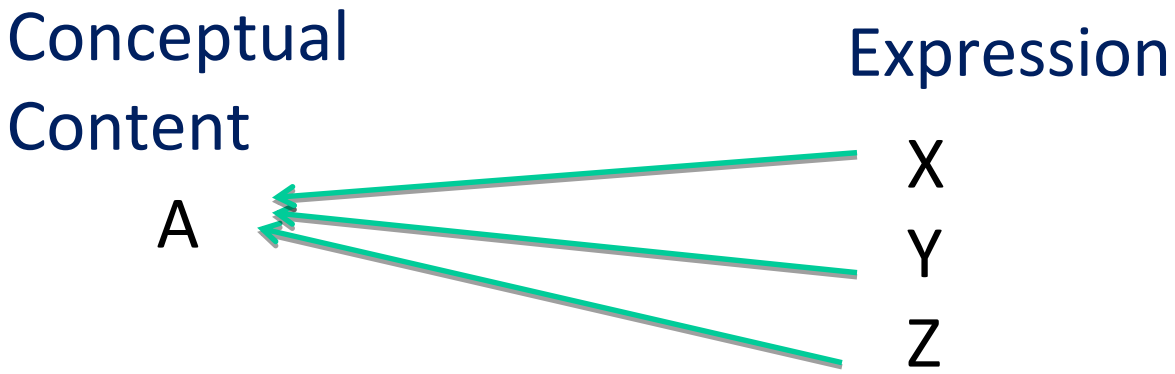
Term formation II

- According to **its components**:
  - Suffixation:
    - *teca/tica*: *animática, burótica, indumática, ofimática, robótica, telemática, turismática*
    - *ware: hard-, soft-, middle-,*
    - *itis: a)* inflamación*: bronquitis, faringitis, amigdalitis, otitis*
      - *b)* obsesión*: madriditis, mamitis, futbolitis*
  - Prefixation: ciber: *cibercoffee, cibermedicine, cibercrime, etc*
  - Composition: *screensaver = salvapantallas, reposapiés = footrest,*
  - Abbreviation (acronyms): *PC, PDF, TCP/IP, blog, MP3, wysiwyg,*
  - Conversion: *download, input, output, fax-to fax*
  - Neologization: *to twitter, to google, autoedición, "gustomizar",*
  - Metaphorization: *cloud computing, folksonomies, social tagging, surf the net, tag cloud, paquete de mejoras salariales, autopistas de la información, papelera, escritorio, bajar de la red, machacar un fichero, caerse el sistema,* etc

Relations between denomination and concept

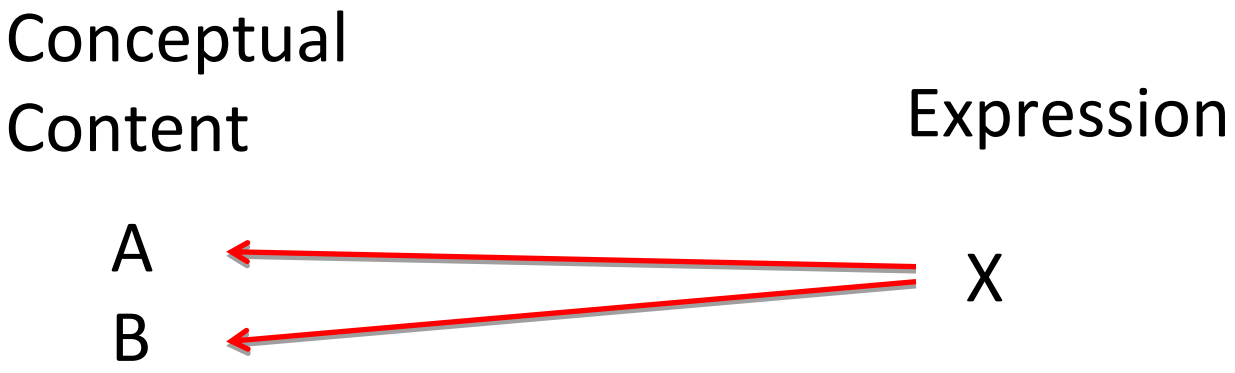
- Synonymy: the quality of two or more words with the same or similar meaning:
  - ***contaminación, polución; store, save a file***
- Polisemy: the capacity for a word(s) or a sign to have multiple meanings.
  - ***Cluster***
- Homonymy: The quality of a word or group of words that share the same speling but have different meanings, usually because they have different origins. :
  - ***Tarifa (ciudad), tarifa de precios***
  - ***Vino (bebida), vino (del verbo "venir")***

Relation between denomination and concept: Synonymy in terminology



1. an acronym and the complete terminological unit:
UCP: Unidad central de proceso
2. An acronym thar represents the English term and the complete term in Spanish:
CPU: unidad central de proceso
LAN : red de área local
3. An abbreviated form and the complete term:
un mini: un miniordenador
una macro: una macroinstrucción
4. A scientific denomination and the popular one:
chip: circuito integrado
5. A standardised term and the dialectal variant
hormigón in Spain and *concreto* in South America
array, *matriz* in Spain y *arreglo* in South America
6. Symbols and their terms
Ca = Calcio
7. Variants of a term:
tecla de borrar = tecla de suprimir
menú de persiana = menú desplegable
Collaborative tagging, social classification,social indexing, social tagging

Relation between denomination and concept: Polisemy in terminology



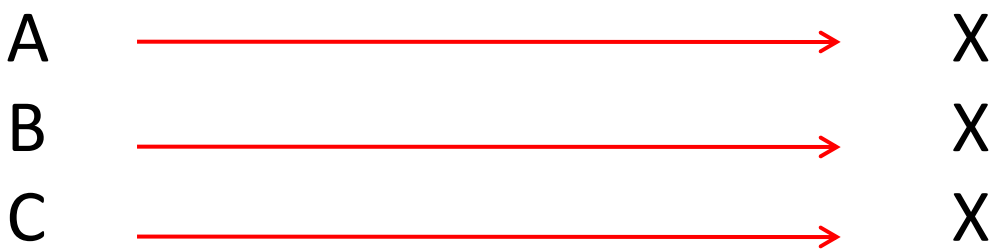
1. A group of loosely coupled computers that work together closely (HW)
2. A group of disk sectors used in a File Allocation Table (SW)

Cluster (Comp.)

Relation between denomination and concept

¿¿ Homonymy in terminology??

Conceptual Content → Expression



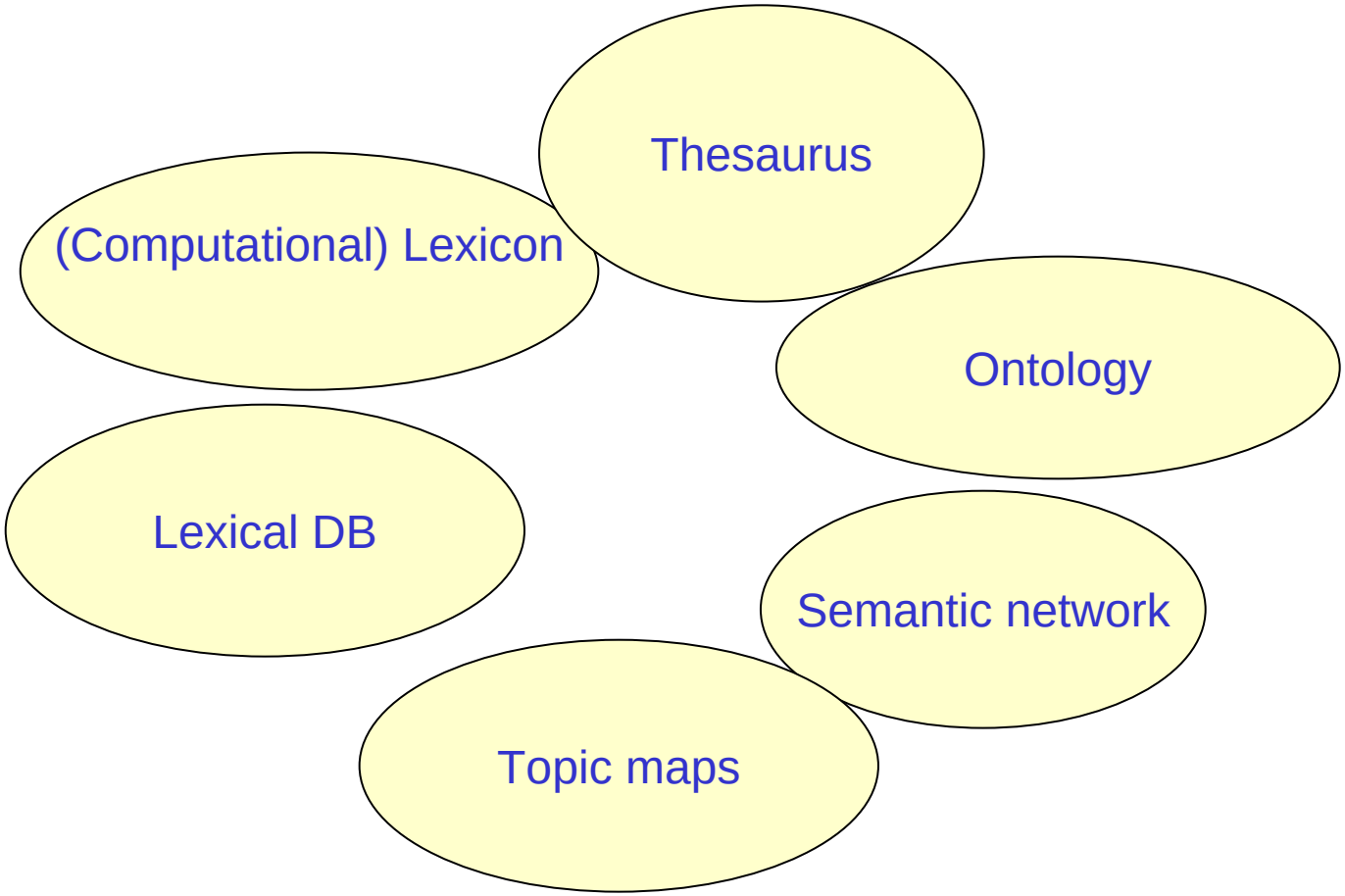
Sp. - Vino (verb: venir)
- *vino* (drink)
- *Concreto* (adjective)
- *concreto* (hormigón)

En: Fluke
A fish, and a flatworm
The end parts of an anchor
The fins on a whale's tail.
A stroke of luck

Water in different concept systems

Chemistry	• molecule-composed compound of two atoms of hydrogen and one atom of oxygen (H2O), considered the universal solvent
Physics	• fluid which is colourless, odourless and tasteless used as the standard of specific gravity and of specific heat which freezes at 0 °C and boils at 100 °C
Physics	• chemical compound which is colourless, odourless, and tasteless and whose formula is H2O and which is naturally found in solid state at temperatures at and below 0 °C, in liquid state at temperatures between 0 °C and 100 °C, and as vapour at temperatures above 100 °C
Biology	• chemical substance that is essential to all known forms of life
Metrology	• chemical compound whose freezing and boiling points are the basis for the Celsius temperature scale, where freezing point equals 0 °C and boiling point equals 100 °C at standard atmospheric pressure
Astrology	• one of the four elements of life that is associated with the emotional and intuitive processes

Linguistic and conceptual resources: terminological chaos?



Some definitions

- Thesaurus: Organized controlled vocabulary of terms and their relations (hierarchical, ec … a subject domain.
- C … nformation developed by several ap
- L… data model that allows the storage , re
- S… emantic relations between concepts
- T… and scopes thet may exist in some in
- C…

word-sense-entry →
[- ORTHOGRAPHY : string
- WORD-MEANING : word-meaning-id+
SYNONYMS : word-meaning-id*
NEAR-SYNONYMS : word-meaning-id*
HYPONYMS : hyponym*
HYPERONYMS : hyperonym*
ANTONYMS : antonym*
MERONYMS : meronym*
HOLONYMS : holonym*
QUANTIFICATION : quantification*
COLLOCATIONS : collocation*
SEMANTIC-FRAME : sem-frame
ACTIONALITY : actionality
ENTRY-CREATOR: (HUMAN | MACHINE)
IS_VALIDATED: Boolean]

the semantics of the grammatical units
other kinds of information

(phonologic, morphologic, syntactic, semantic and p… gmatic)

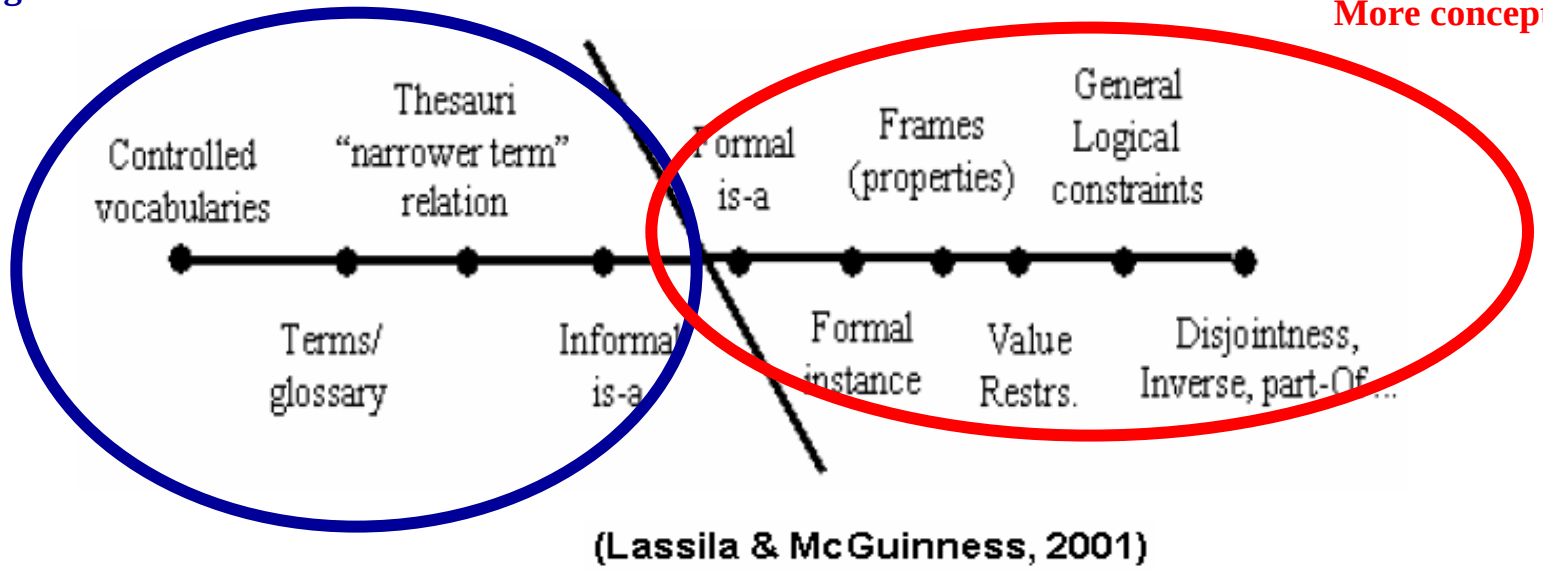
What is an ontology?

- "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/

Classification from an ontological perspective (Lassila & McGuinness)

Lightweight Ontologies **Heavyweight Ontologies**

More linguistic **More conceptual**

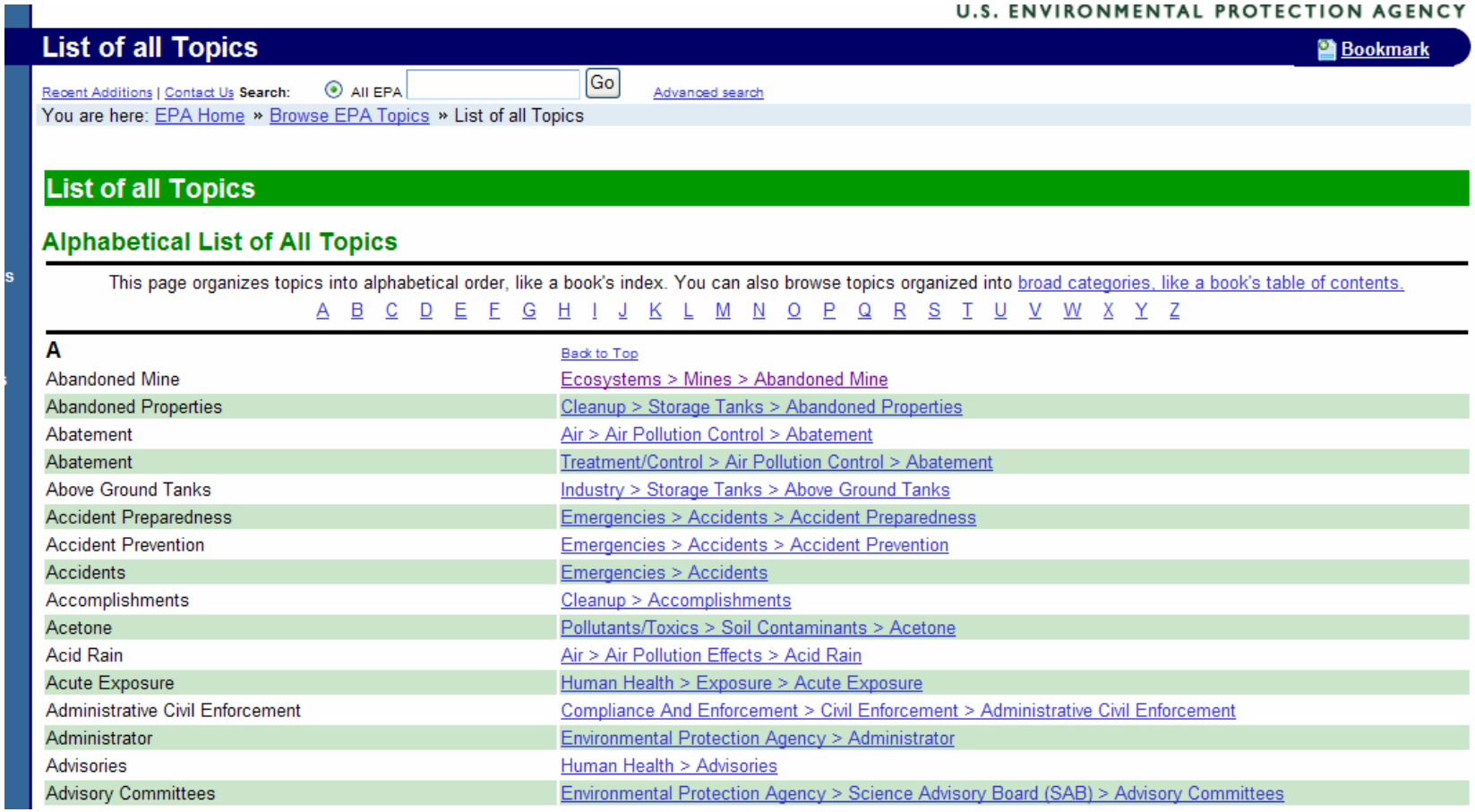


(Lassila & McGuinness, 2001)

- **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, not always with definitions.
- Terms usually belong to a subject field
- Terms are defined according to the meaning in that field only



U.S. ENVIRONMENTAL PROTECTION AGENCY

List of all Topics Bookmark

Recent Additions | Contact Us **Search:** All EPA Go 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

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

Listado alfabetido 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

INSPEC Thesaurus

Type 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)

Equivalence relation

Generic relation

Hierarchical relation

Specific relation

Associative relation

Other data

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

ASFA Thesaurus

P | alphabetic | go to term | next page

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

P-waves SELECT WORD BLOCK

Pack ice SELECT WORD BLOCK

Packages USE: Containers SELECT WORD BLOCK

Packaging fishery products USE: Packing fishery products SELECT WORD BLOCK

Packaging materials USE: Packing materials SELECT WORD BLOCK

Packing fishery products SELECT WORD BLOCK

Packing materials SELECT WORD BLOCK

Paddy fields USE: Rice fields SELECT WORD BLOCK

ASFA Thesaurus

PACK ICE SELECT

uf Ice floes

BT SELECT Floating ice

rt SELECT Ice barriers

rt SELECT Ice canopy

rt SELECT Ice drift

rt SELECT Ice fields

ASFA Thesaurus

PACKING FISHERY PRODUCTS SELECT

(Referring to methods, techniques and material for packing industrial fishery products)

uf Packaging fishery products

rt SELECT Fishery industry

rt SELECT Fishery products

rt SELECT Packing materials

rt SELECT Processed fishery products

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 o meaning, each representing a certain synset of the search term. In order to find out which element of the web re to which meaning or synset, please click on it. Two thing 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 ...)

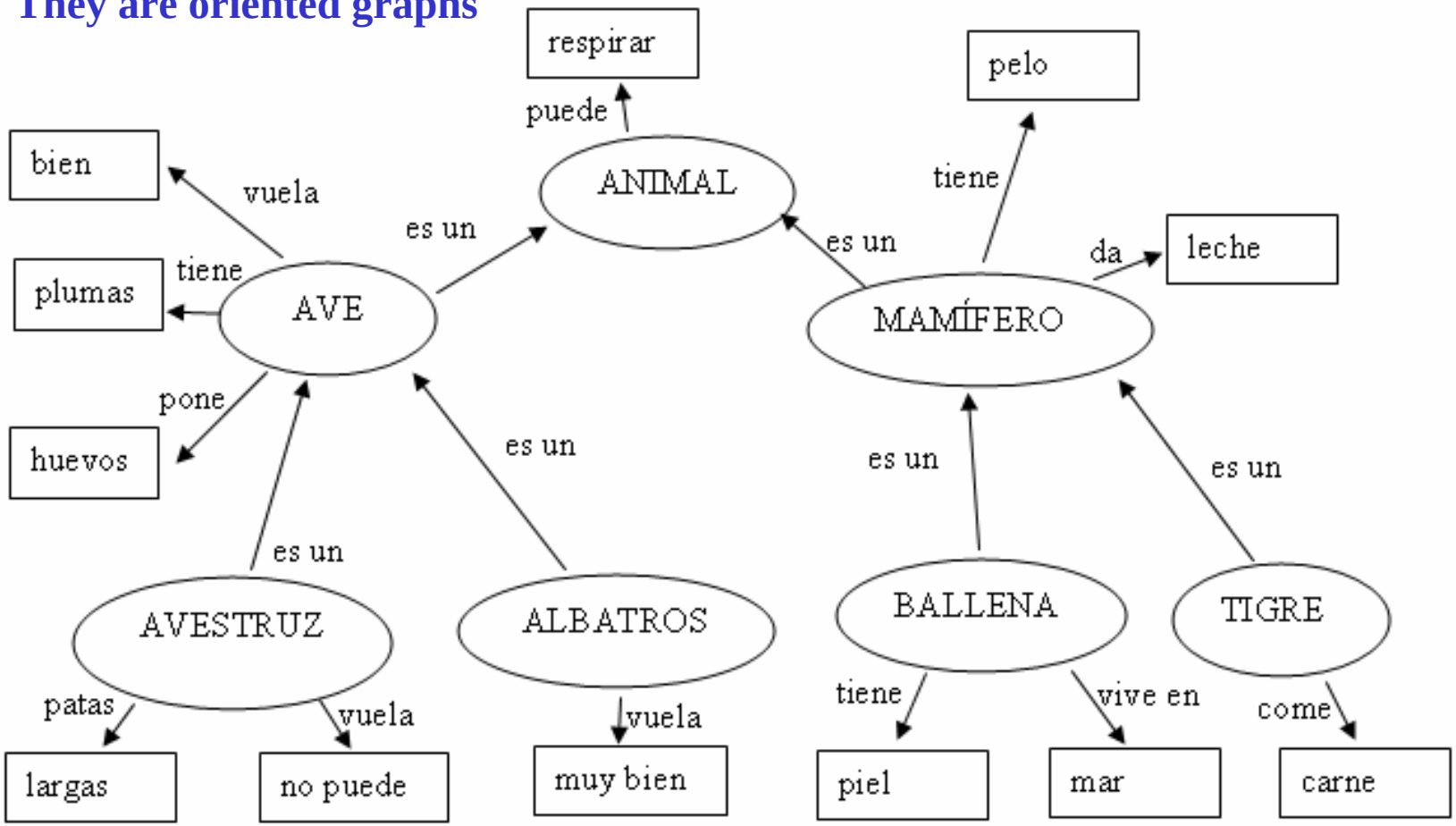
[3] Related verbs

[4] Related nouns

Types of relations

Semantic network

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



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

Lexicon

Account *n.* Domain [financial]

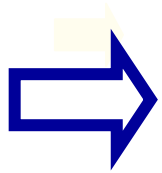
Account *v.* …

Bank_1 *n domain* [financial]

Bank_2 *n domain* [geography]

Money n. domain [financial]

bank……
……account
……
money……



Topic= financial field

Lexicon

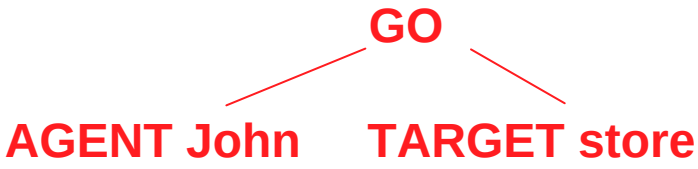
Went *v*past GO

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

John went to the store



Adapted from Nancy Ide

Types of lexicons

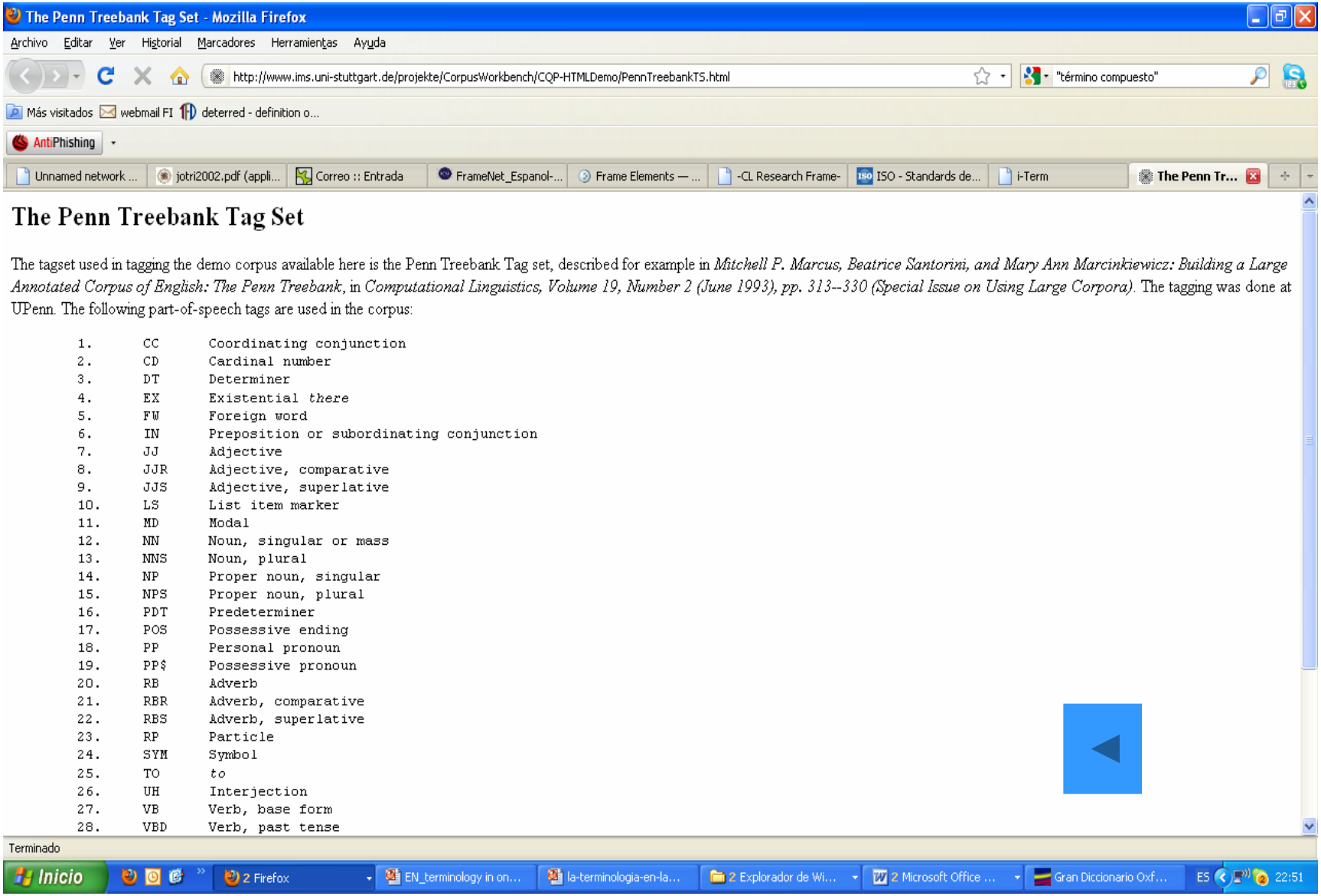
- Various types:
  - Morphosyntactic Information :

    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

PennTreeBank



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*]

WordNet 3.0 Vocabulary Helper

pollution Search

Help for Fwa Word Lookup Interfaces

Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun pollution

3 senses of **pollution**

Sense 1

pollution -- (undesirable state of the natural environment being contaminated with harmful substances as a consequence of human activities)

- environmental condition -- (the state of the environment)
- impurity, impureness -- (the condition of being impure)

Sense 2

befoulment, defilement, **pollution** -- (the state of being polluted)

- dirtiness, uncleanness -- (the state of being unsanitary)

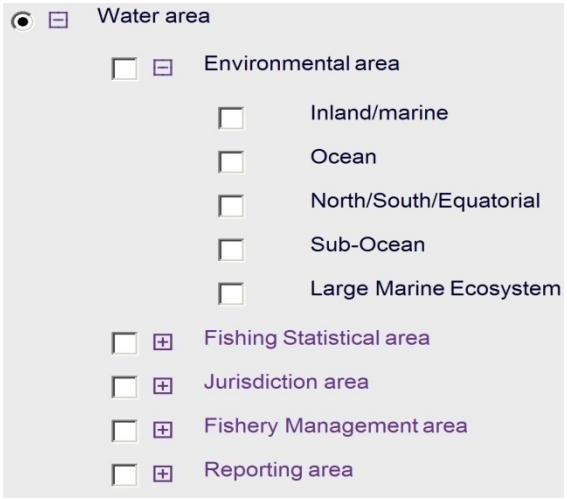
Sense 3

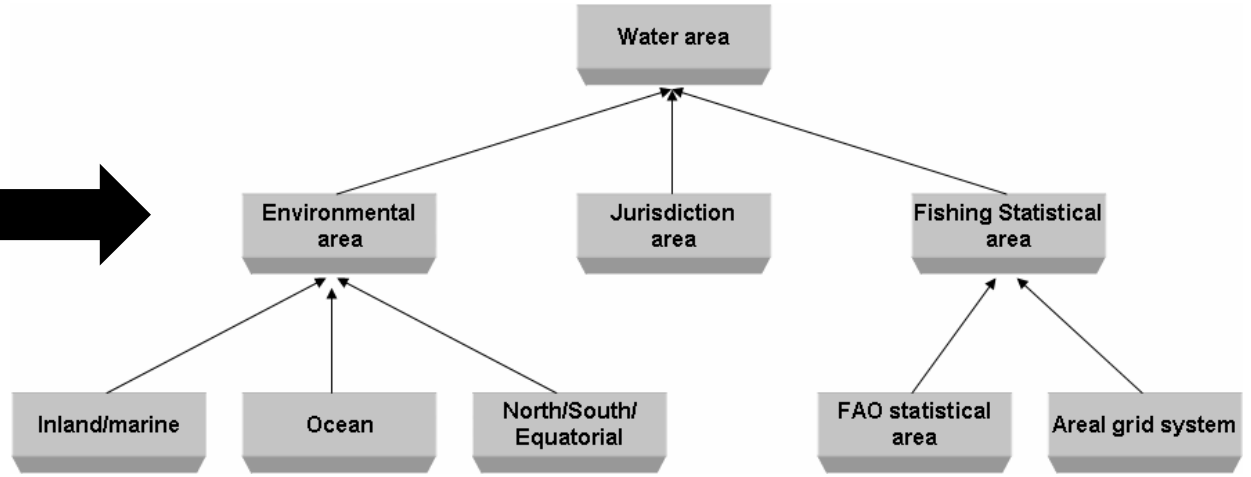
contamination, **pollution** -- (the act of contaminating or polluting; including (either intentionally or accidentally) unwanted substances or factors)

- soiling, soilure, dirtying -- (the act of soiling something)
- 1. (1) **pollution** -- (undesirable state of the natural environment being contaminated with harmful substances as a consequence of human activities)
- 2. befoulment, defilement, **pollution** -- (the state of being polluted)
- 3. contamination, **pollution** -- (the act of contaminating or polluting; including (either intentionally or accidentally) unwanted substances or factors)

Classification Scheme

- A classification scheme[1] 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 scheme[2].





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

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)

Reusing and sharing KOS in the Semantic Web: SKOS

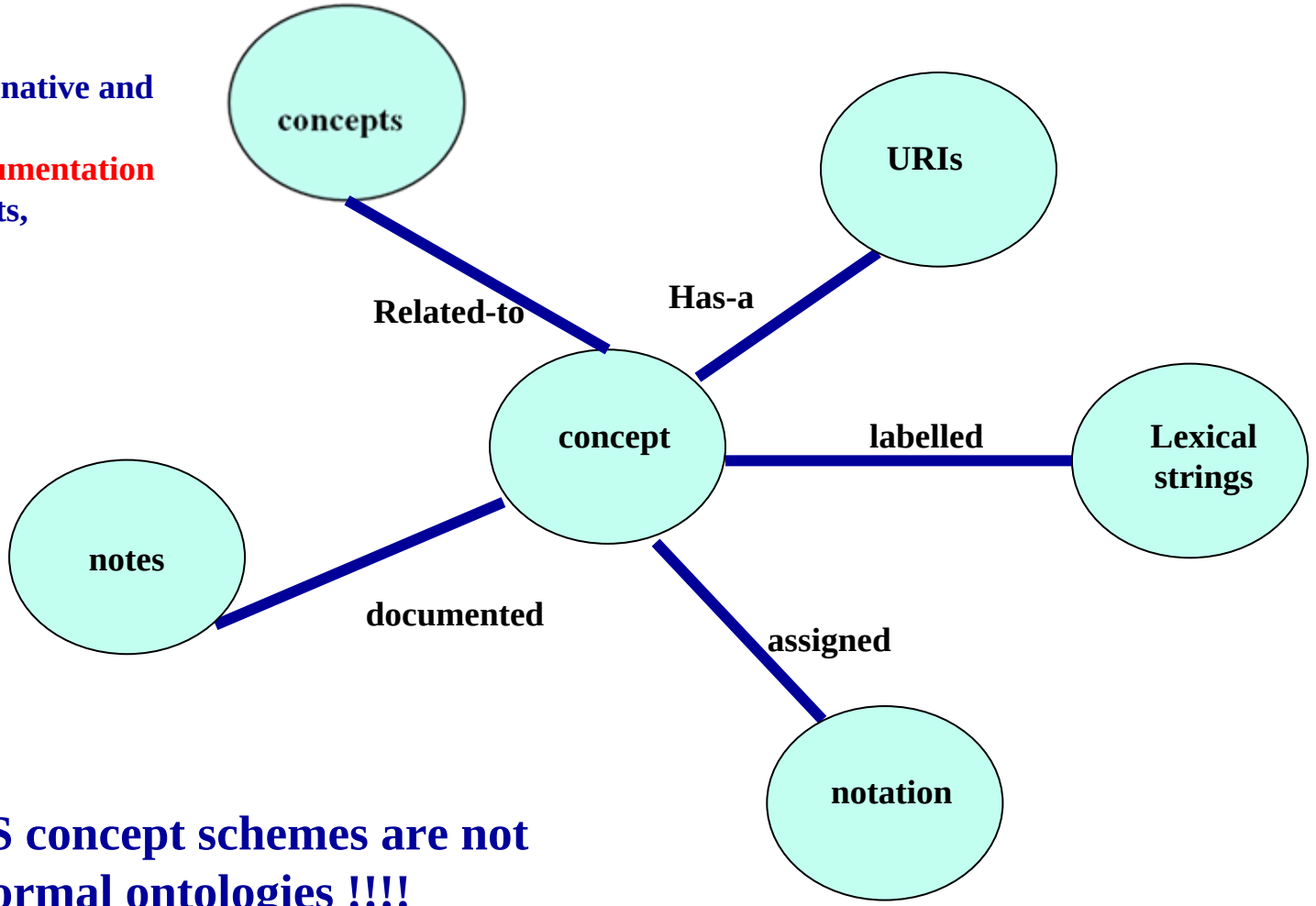
- Simple Knowledge Organization System
  - Simple, flexible, extensible, machine-understandable representation for sharing KOS
  - Goal: to enable easy publication of controlled structured vocabularies for the semantic web
    - Thesauri
    - Classification schemes
    - Subject heading systems
    - Taxonomies
    - Other 'controlled language'

Many exist and are in use in cultural heritage, medicine, libraries, …

  - How: by using a common data model for sharing and linking knowledge organization systems
  - BUT SKOS is **not** a formal knowledge representation language.

Elements in SKOS

- **Semantic Relationships**
  Broader/Narrower Terms
  Related Terms
- **Lexical Labels**
  Preferred, alternative and hidden labels
- **Additional documentation**
  Notes, comments, descriptions

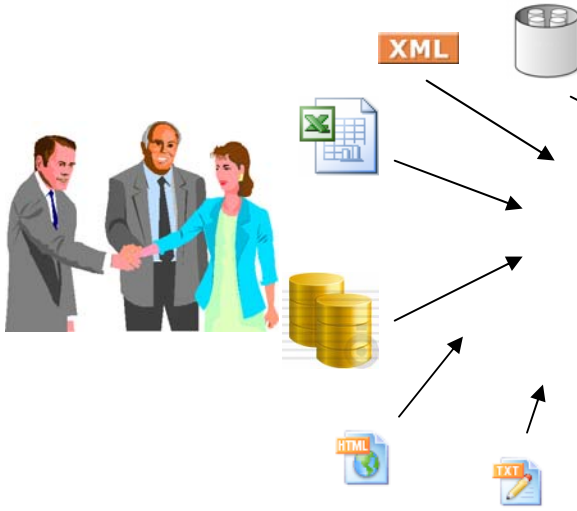


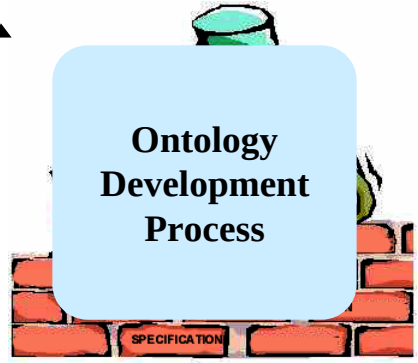
SKOS concept schemes are not formal ontologies !!!!

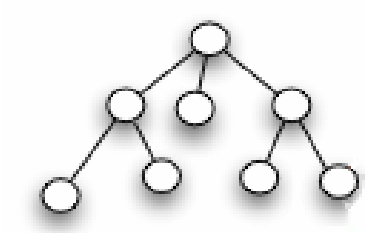
Transforming NOR into ontologies
García-Silva, Gómez-Pérez, Suárez-Figueroa, Villazón-Terrazas

From scratch

Re-engineering available resources







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

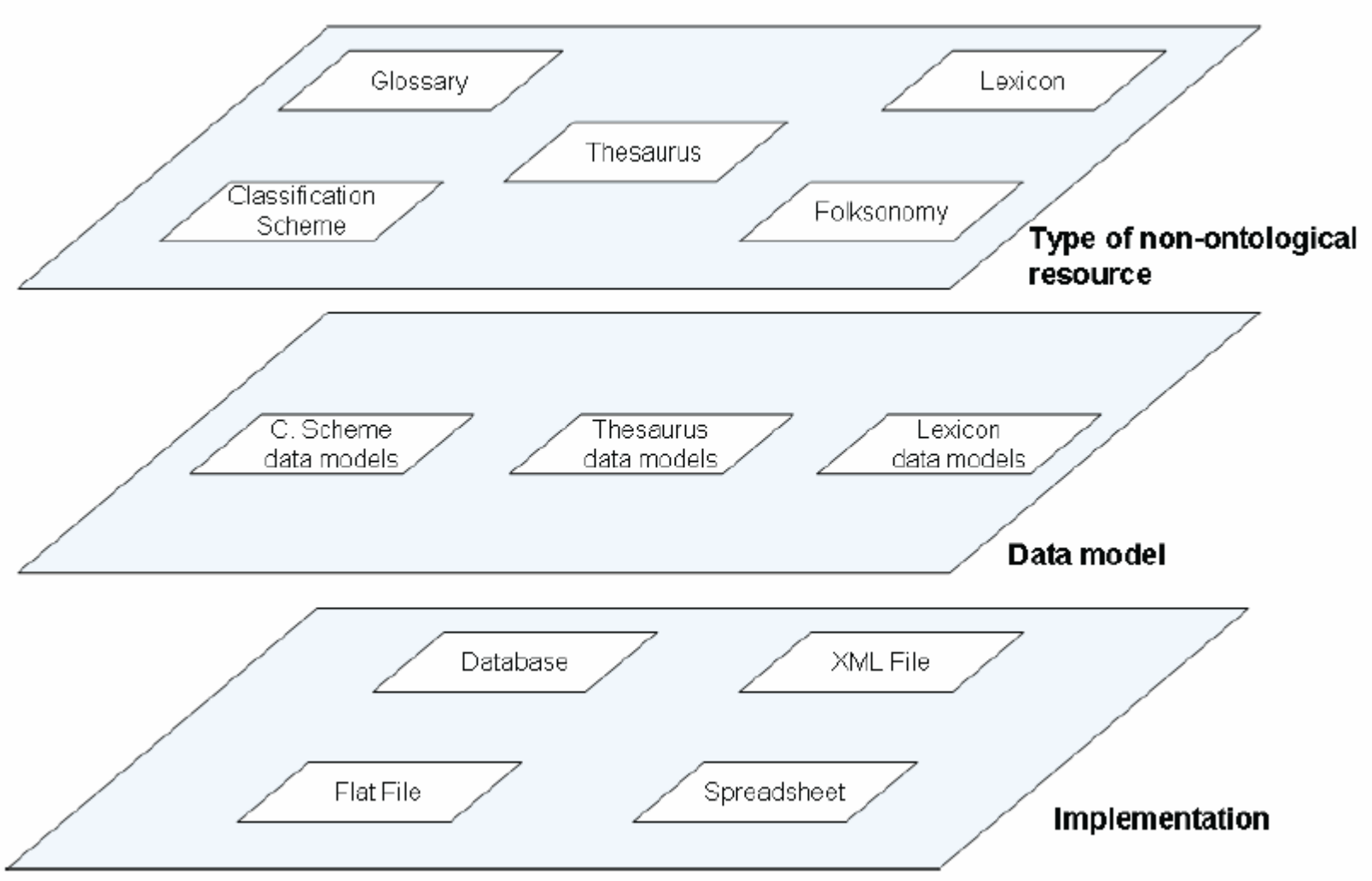
- *Catalogue of normalized terms*
- *Glossed Catalogue*
- *Taxonomy*

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.

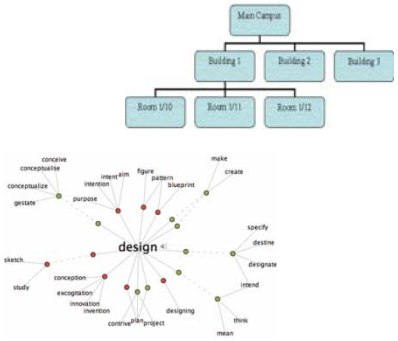
NOR categorization



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	-



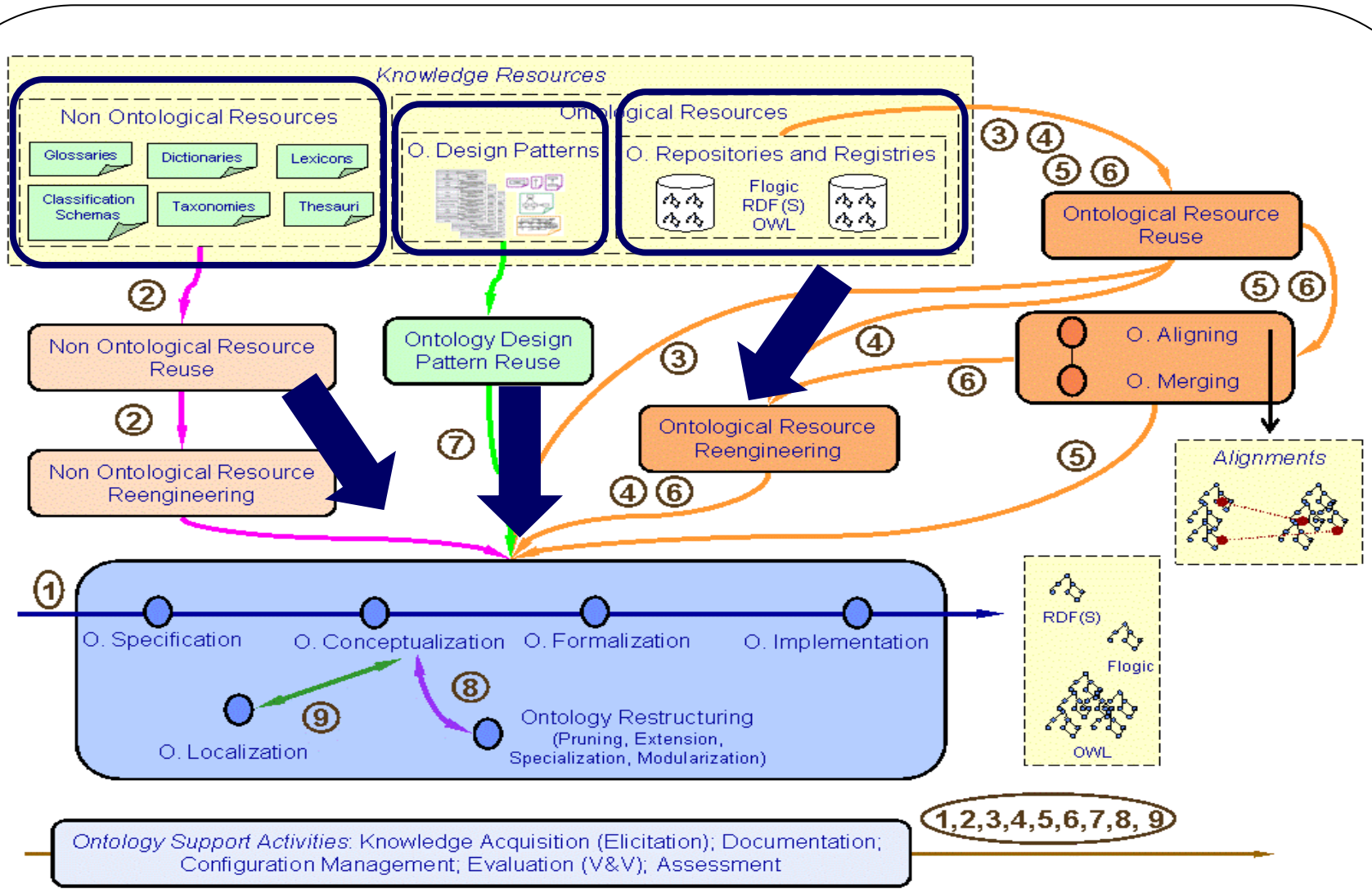
Approaches focused on the NOR Implementation

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

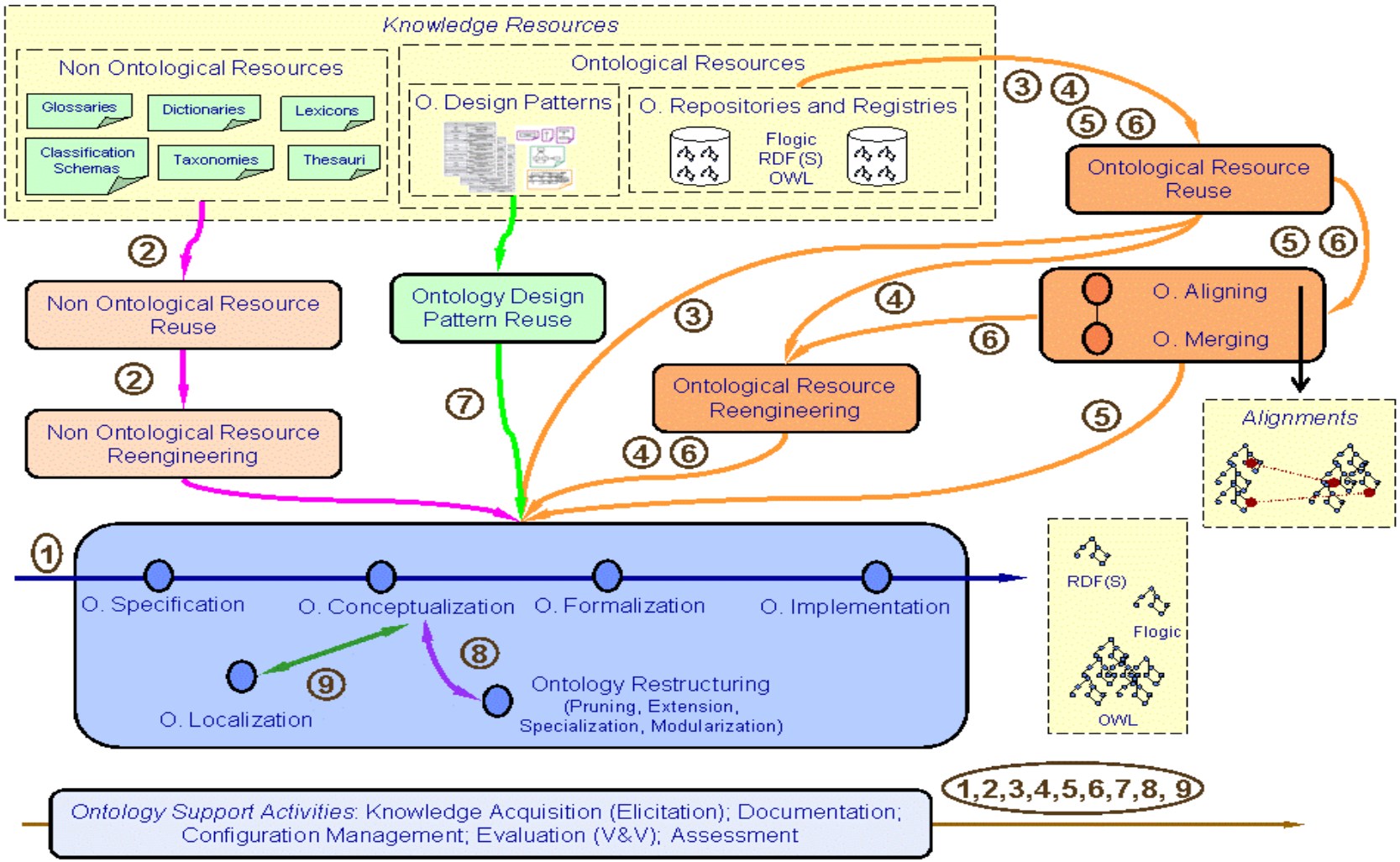
They do not take into account the Non-Ontological Resource data model, an important artifact in the re-engineering process.

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

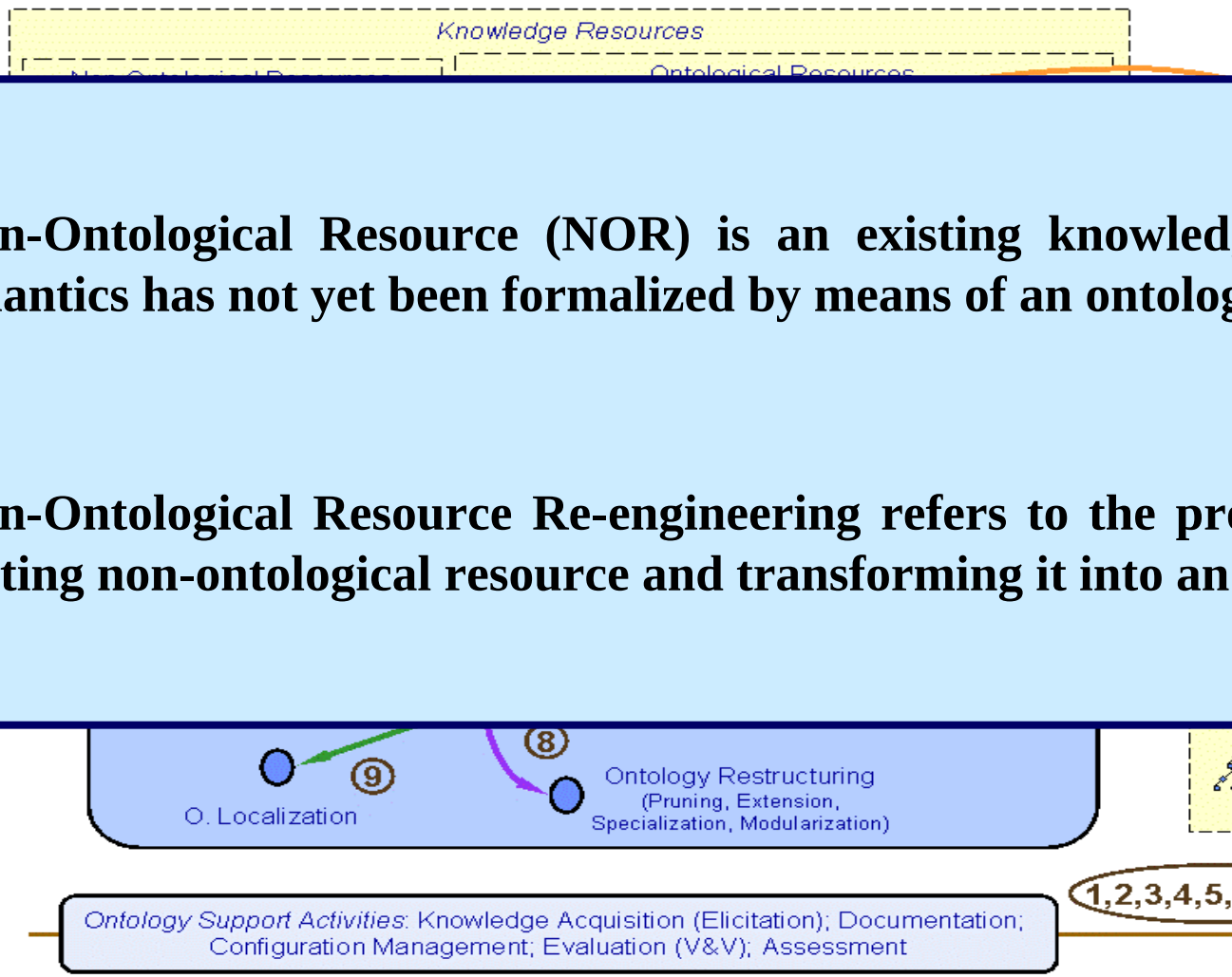
The scenario in NeOn



The scenario in NeOn



The scenario in NeOn



•Non-Ontological Resource (NOR) is an existing knowledge resource whose semantics has not yet been formalized by means of an ontology.

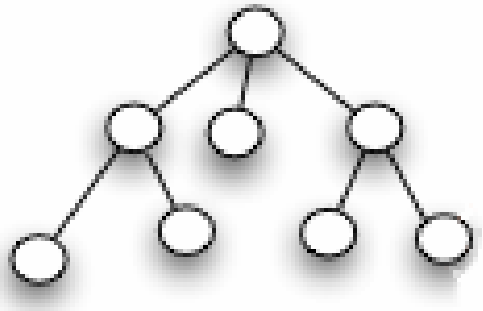
•Non-Ontological Resource Re-engineering refers to the process of taking an existing non-ontological resource and transforming it into an ontology.

Motivation

resource

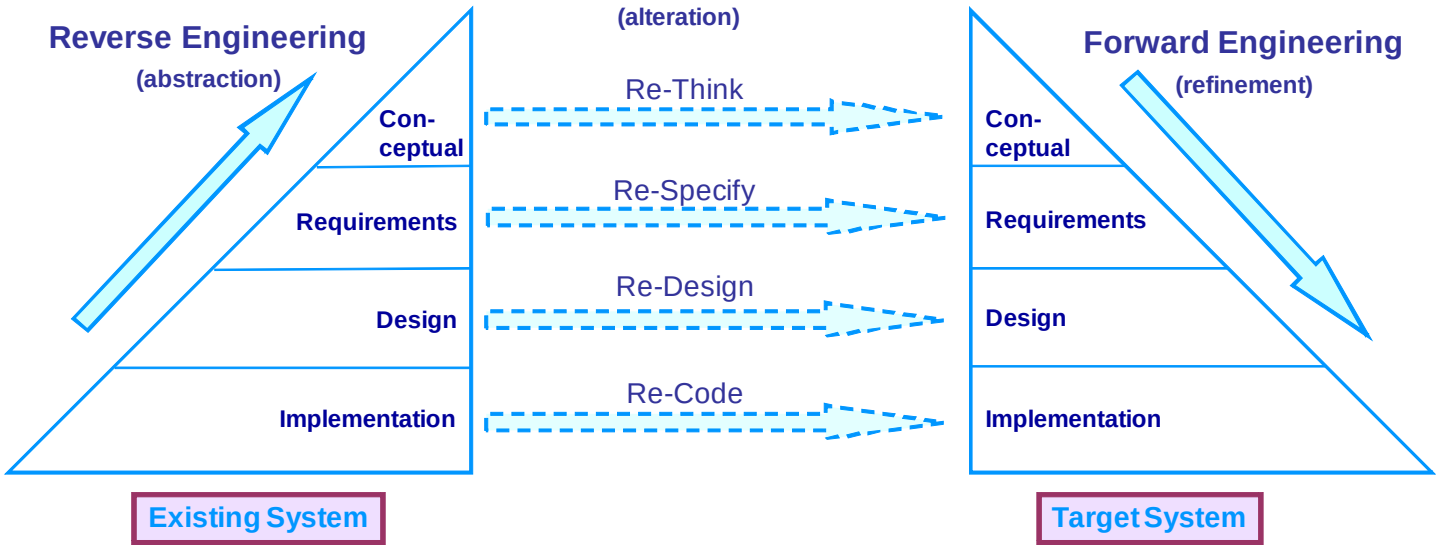
How?

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



ontology

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

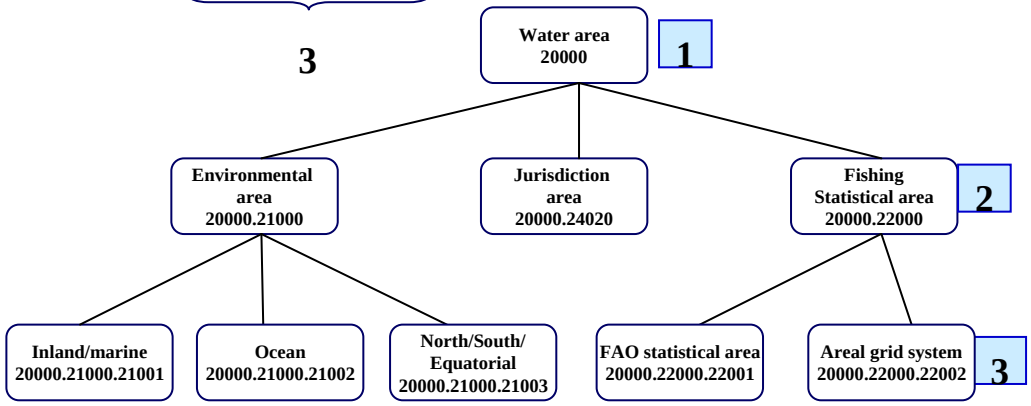
E.J. Byrne. A conceptual foundation for software re-engineering.
In Proceedings of the International Conference on Software Maintenance and Reengineering, pages 226–235.
IEEE Computer Society, 1992.

Data Models (I)

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

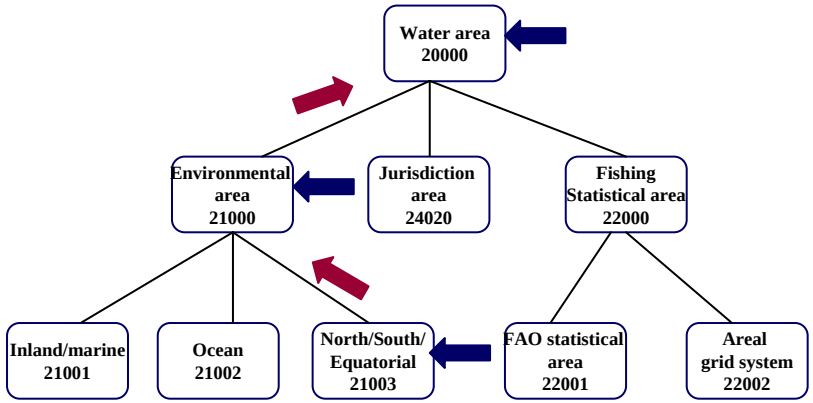
ID	CSI_Name
20000	Water area
20000.21000	Environmental area
20000.24020	Jurisdiction area
20000.22000	Fishing Statistical area
20000.21000.21001	Inland/marine
20000.21000.21002	Ocean
20000.21000.21003	North/South/Equatorial
20000.22000.22001	FAO statistical area
20000.22000.22002	Areal grid system

1
2
3



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

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



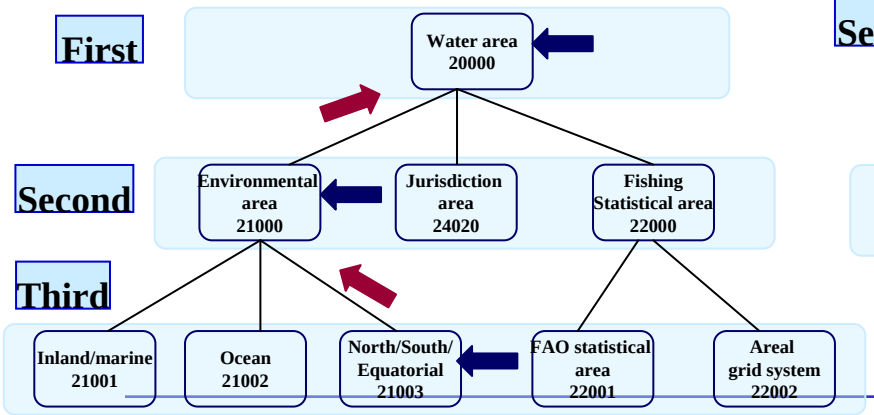
Data Models (II)

- **Snowflake Data Model** is a normalized structure for hierarchical representations. For each hierarchical level an 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

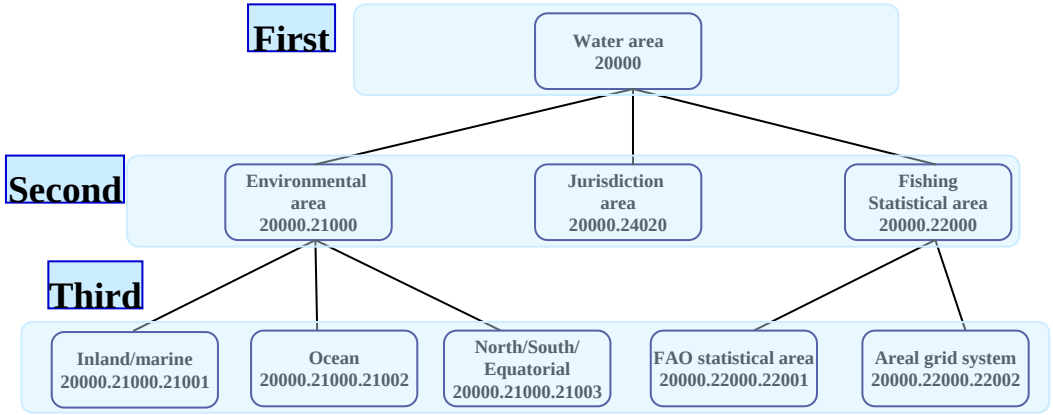
Second Level		
ID	First Level ID	CSI_Name
21000	20000	Environmental area
24020	20000	Jurisdiction area
22000	20000	Fishing Statistical area

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

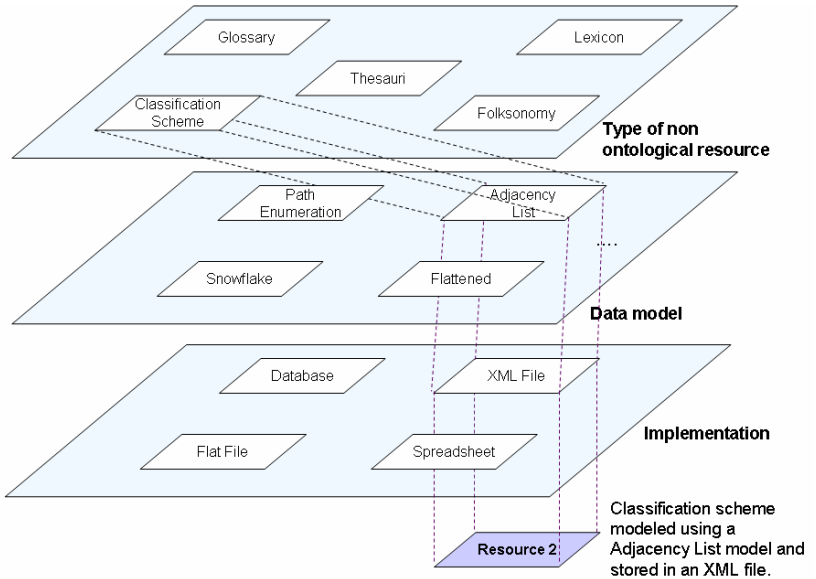


- **Flattened Data Model, is a de-normalized structure. The hierarchy is represented with a table where each hierarchical level is stored in a different column.**

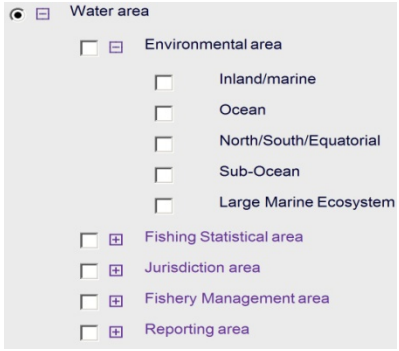
First Level		Second Level		Third Level	
ID	CSI_Name	ID	CSI_Name	ID	CSI_Name
20000	Water area	21000	Environmental area	21001	Inland/marine
20000	Water area	21000	Environmental area	21002	Ocean
20000	Water area	21000	Environmental area	21003	North/South/Equatorial
20000	Water area	22000	Fishing Statistical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



Example - FAO Water Areas Classification Scheme



Classification Scheme



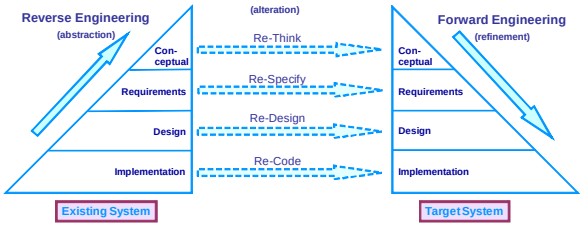
Adjacency List

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

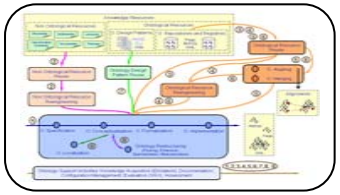
XML

```
<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>
```

NeOn Approach for Re-engineering NORs



General Model for Software Reengineering



Patterns for Reengineering Non Ontological Resources (PR-NOR)

Reverse Engineering

Con-ceptual

Requirements

Design

Implementation

Transformation

Speci-fication

Conceptua-lization

Formalization

Implementation

Forward Engineering

Non Ontological Resource

Ontology

http://ontologydesignpatterns.org/

Ontology Engineering Group

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 [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.
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

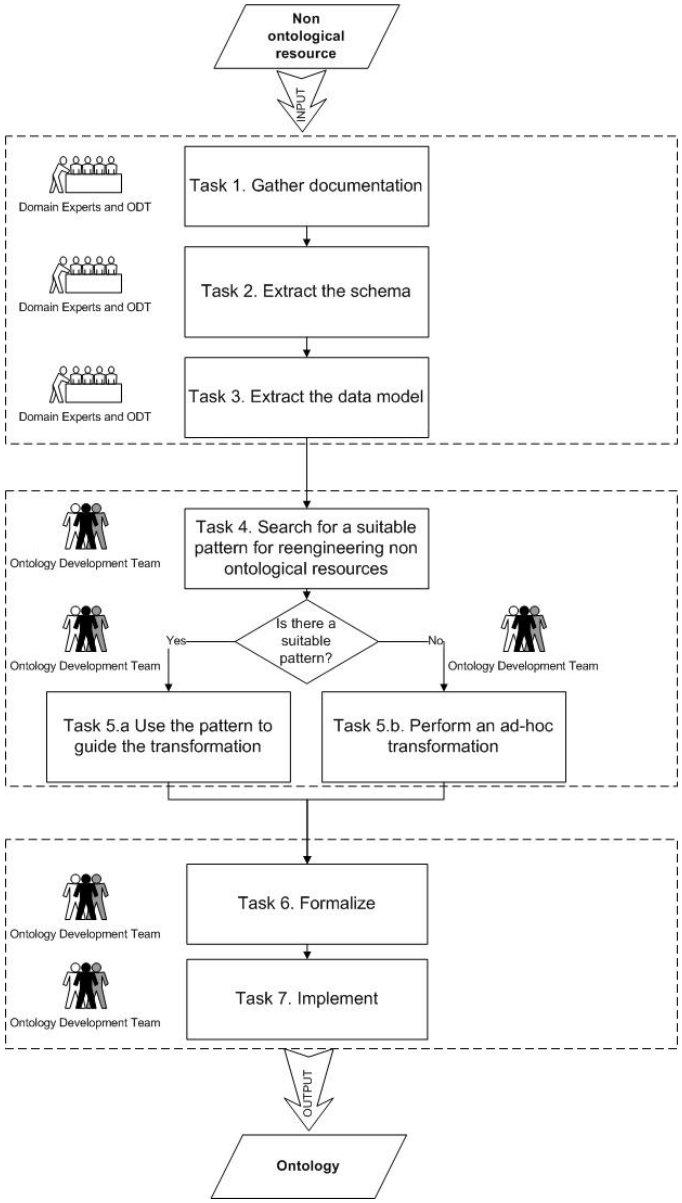
PR-NOR Template

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 2
Non-Ontological Resource Transformation

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 form texts can be the starting point in the development of ontologies

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

Recommended papers
Terminology and ontologies, October 2009

- Read three papers from the six proposed in the wiki. Comment on the ideas proposed and give your opinion in your own words.
- 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