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**

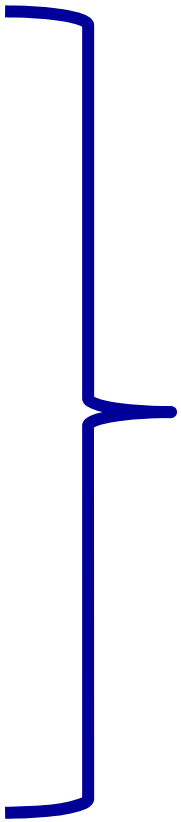
Terminology in technical & scientific domains

- 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)

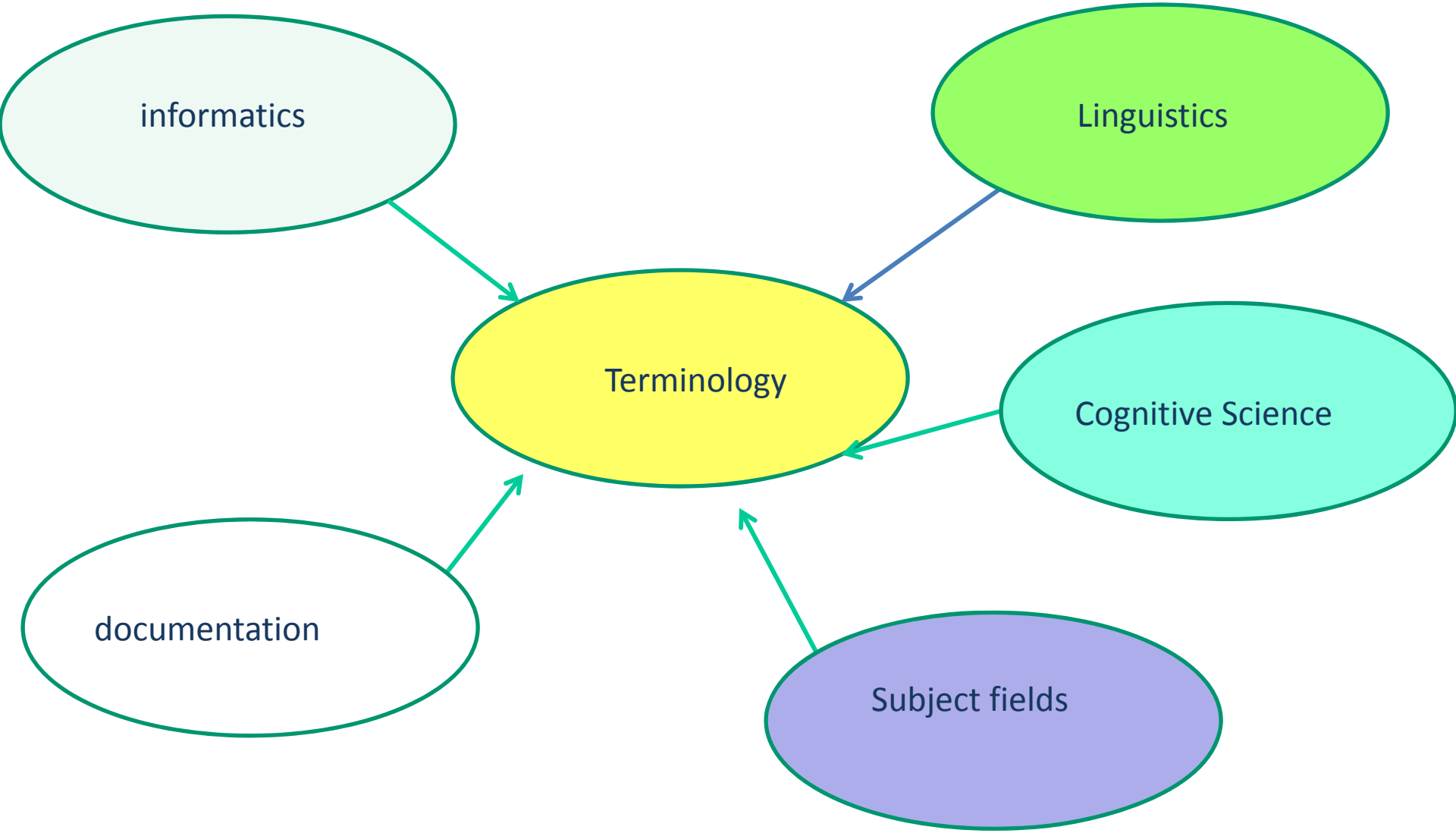
Terminological data as content carriers

- Record
- Store
- Order
- Manage
- Represent
- Retrieve
- Disseminate
- Communicate
- Transfer

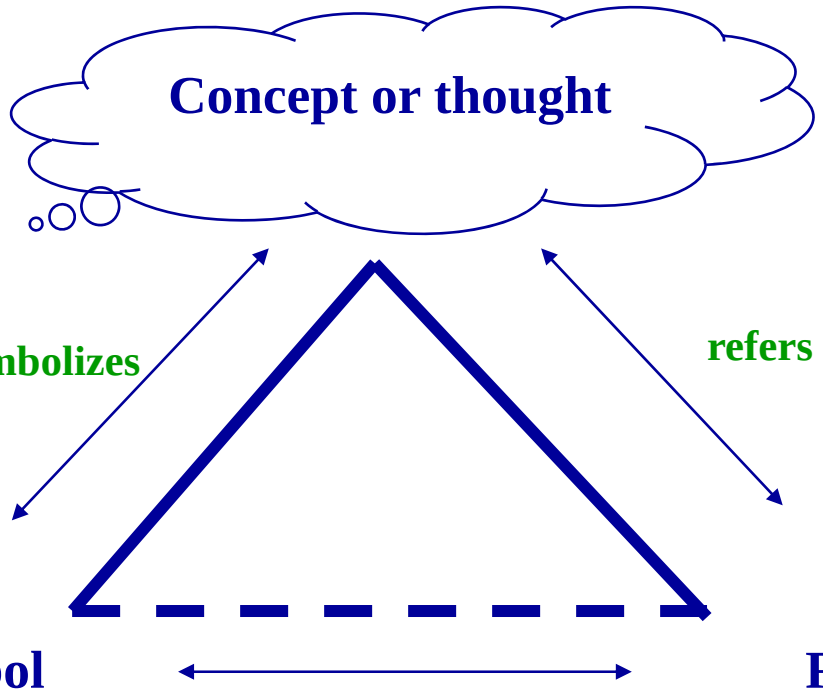


Specialized information and knowledge

Terminology: interdisciplinary subject field



How do we relate concepts and terms?



Semantic triangle by Ogden & Richards

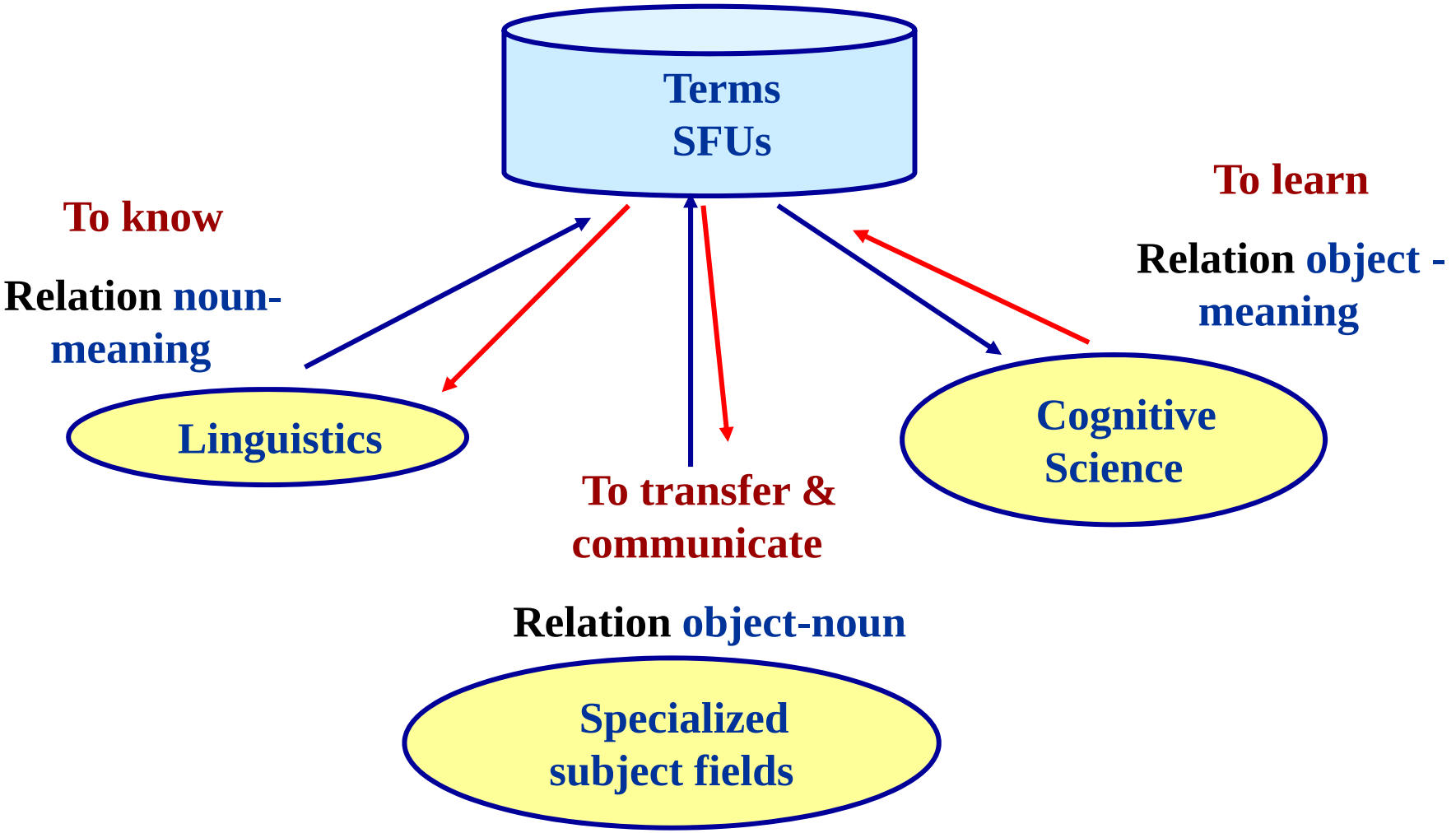
Object of study in terminology

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

NOWADAYS

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

The object of study of terminology
Linguistic approach (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

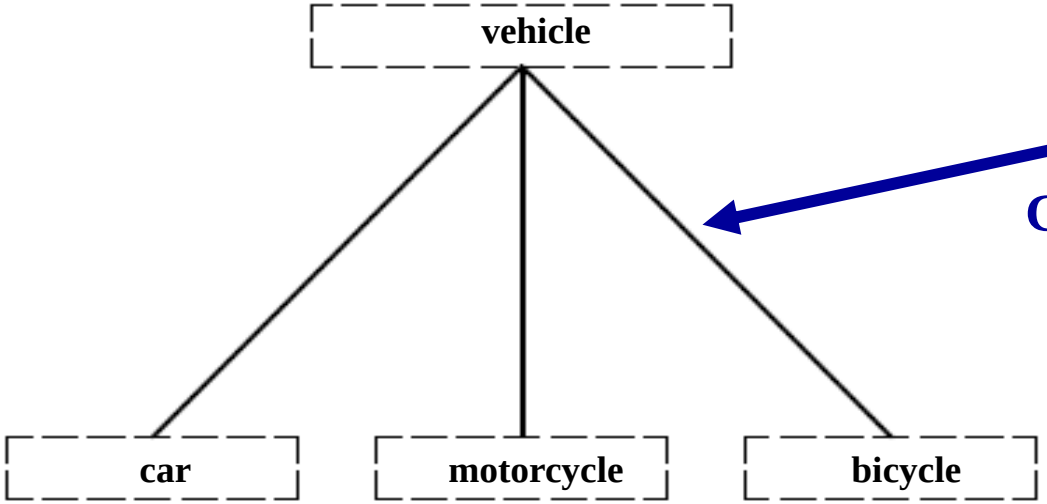
Close relation between a concept and its characteristics

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)

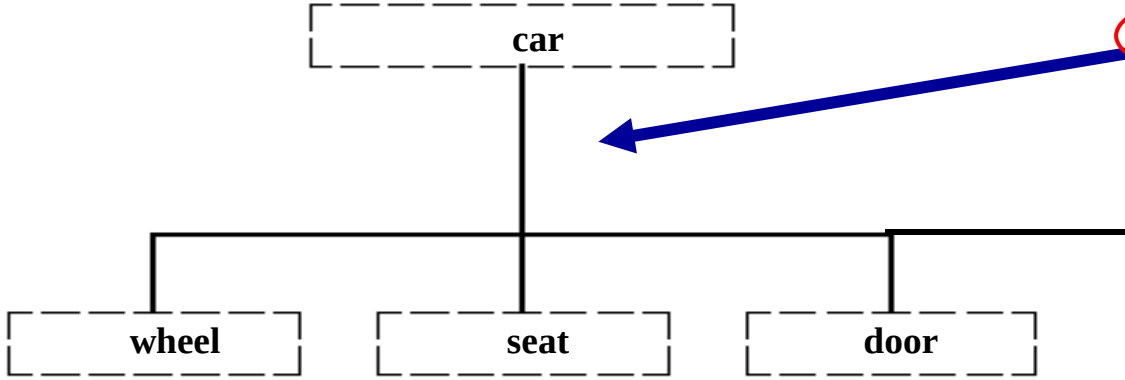
B. PARTITIVE RELATIONS (part-whole relation) PART_OF

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



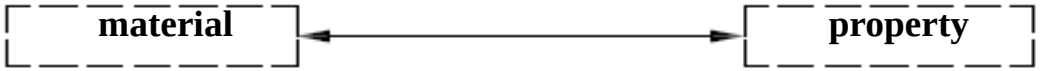
Tree diagram

Generic concept relations



Rake diagrams

Partitive relations



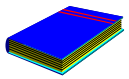
Line with arrowheads

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

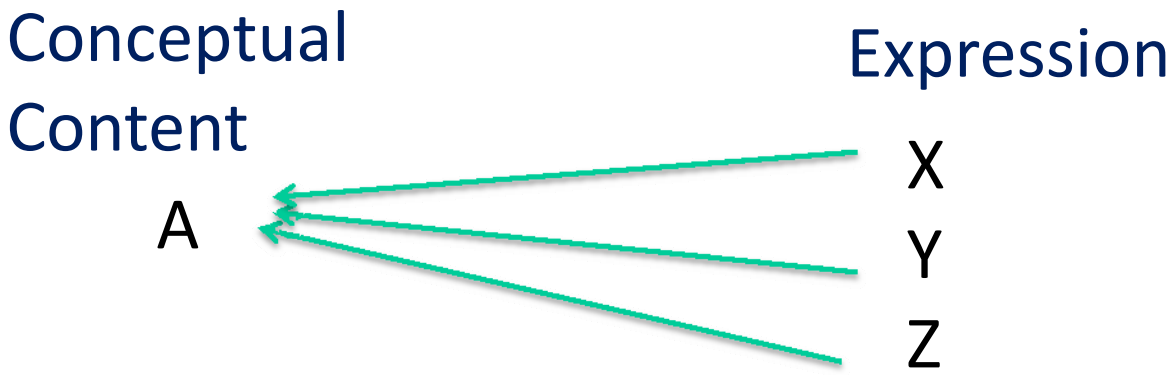
- According to its **origin**:
  - – Borrowings from other languages: *hardware, software, football, cookies, folksonomies*
  - – Adapted borrowings: *formatear, inicializar, fútbol*, etc.
  - – Loans of structure : *inteligencia artificial, lógica difusa, programación orientada a objetos, anotación social, kindergarten= jardín de infancia*
  - – Semantic loans: *aplicación, utilidades, editar, icono, ratón, menú, nube, 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

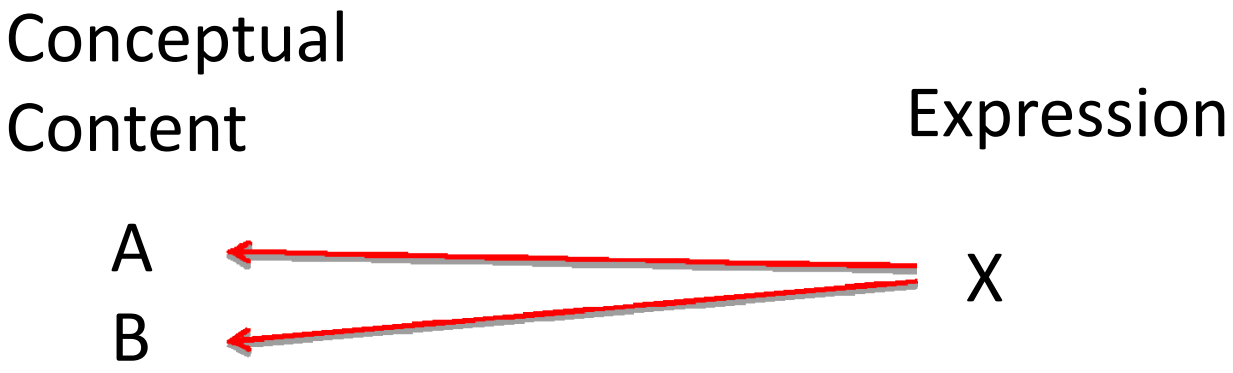
- According to **its components**:
  - **Suffixation:**
    - *teca/tica*: *animática, burótica, indumática, ofimática, robótica, telemática, turismática*
    - *ware: hard-, soft-, middle-,*
    - *itis: a)* inflamación*: bronquitis, faringitis, amigdalitis, otitis*
    - *b)* obsesión*: madriditis, mamitis, futbolitis*
  - **Prefixation:** ciber: *cibercoffee, cibermedicine, cibercrime, etc*
  - **Composition**: *screensaver = salvapantallas, reposapiés = footrest,*
  - **Abbreviation (acronyms):** *PC, PDF, TCP/IP, blog, MP3, wysiwyg, P2P*
  - **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

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



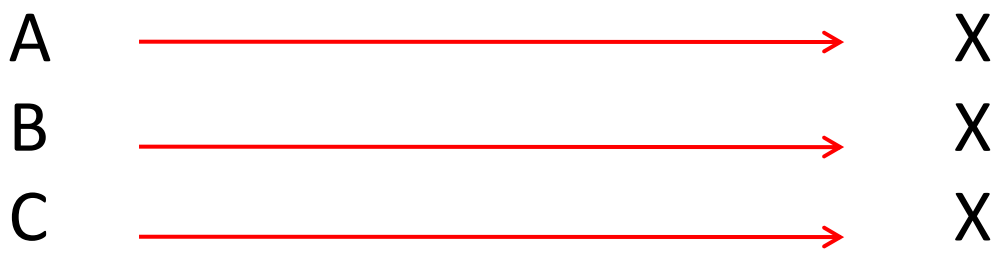
1. A group of loosely coupled computers that work together closely (HW)
2. A group of disk sectors used in a File Allocation Table (SW)

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 (aleta)
The end parts of an anchor (uña)
The fins on a whale's tail.
A stroke of luck (chiripa)

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

Concept systems

- model *concepts* and relations between them based on specialized knowledge of a *subject field*;
- clarify the relations between *concepts*;
- form the basis for a uniform and standardized *terminology*;
- facilitate the comparative analysis of *concepts* and *designations* across languages and across *subject fields*;
- facilitate the writing of *definitions*;
- facilitate the inclusion of all relevant *concepts* while developing a terminological resource.

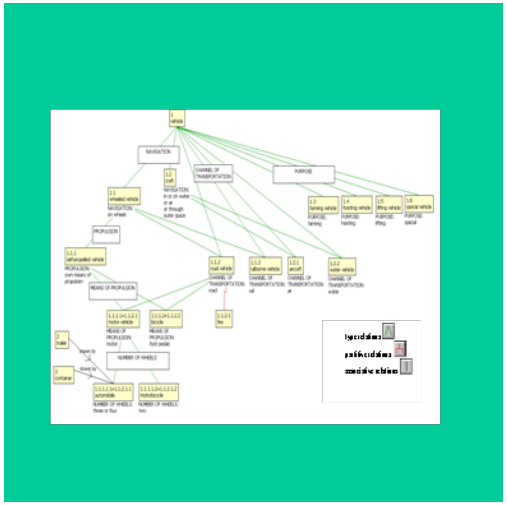
Example of conceptual hierarchy "entidad de defensa costera" with semantics

- **Obra de defensa costera**: construcción de hormigón, madera y acero destinada a reflejar y absorber la energía del oleaje, previniendo así la erosión.
  - **Espigón**: **obra de defensa**, [**is-a**], perpendicular a la costa, [**location-of**], construida a base de madera, hormigón o escollera [**made-of**], que retarda la deriva litoral y el proceso de erosión [**has-function**]
  - **Dique rompeolas: obra de defensa**, generalmente paralela a la costa,. Construida a base de madera, hormigón o escollera, para la protección del impacto del oleaje y para proporcionar abrigo a una zona marítima.
    - **dique exento**: **dique rompeolas** que no está conectado a la costa

(Faber et al. 2007)

Classification & concepts

Concept system



Classification system(s)

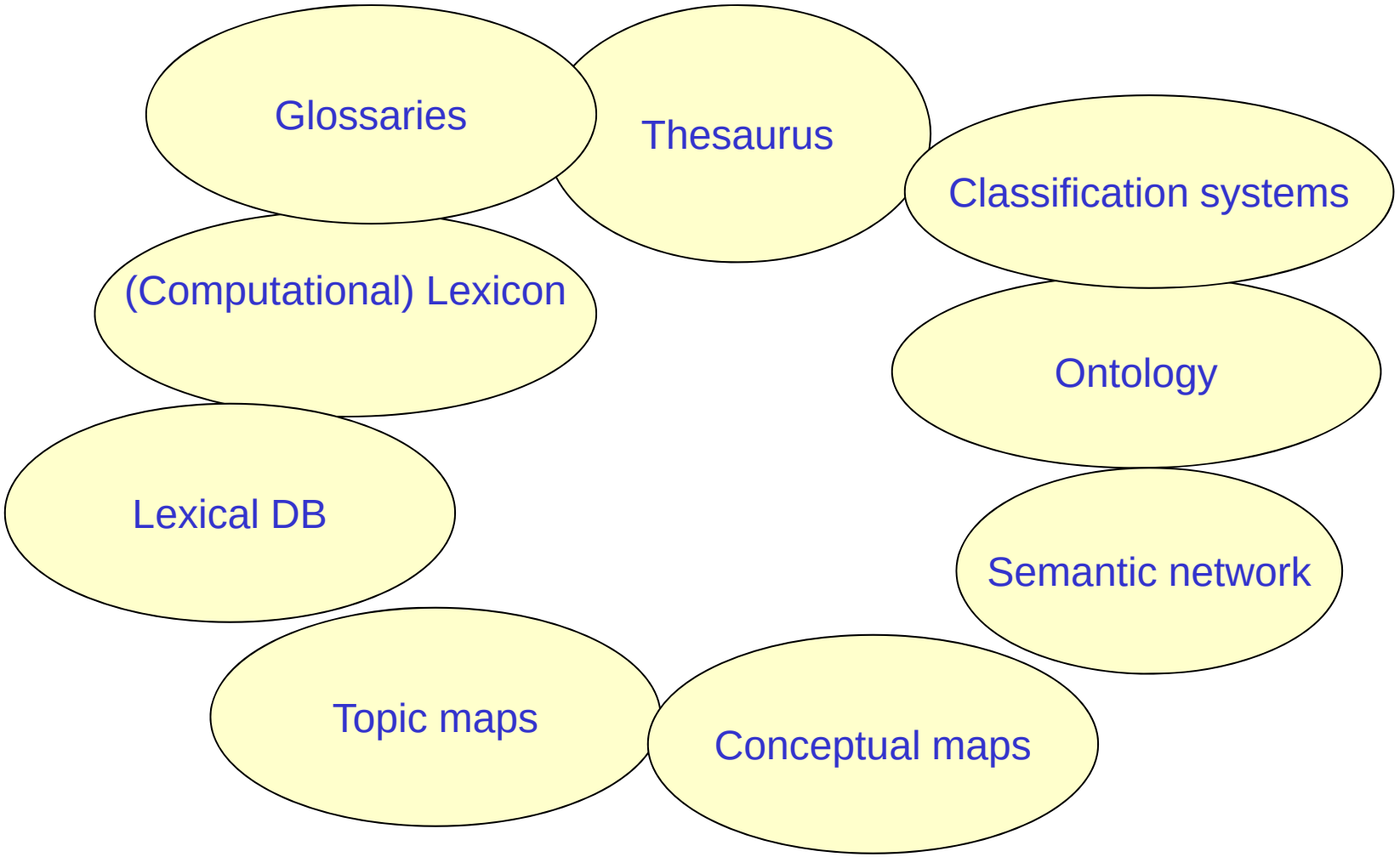
- Automotive technology
  - Automobile
  - Hoisting, lifting vehicle
  - Aircraft
  - Vehicle (railborne)
  - Water vehicle
  - Special vehicle
  - Trailer
  - Tires
  - Farming vehicle

- Automotive technology
  - Automobile
  - Hoisting, lifting vehicle
  - Aircraft
  - Vehicle (railborne)
  - Water vehicle
  - Special vehicle
  - Trailer
  - Tires
  - Farming vehicle

Taken from Reinhard Nerke
ISO TC SC3, 22274

28

Linguistic and conceptual resources: terminological chaos?



Some definitions

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]

of terms and their relations (hierarchical,
in a subject domain.

information developed by several

a data model that allows the storage ,

s semantic relations between concepts

ns and scopes thet may exist in some

on for an arrangement of objects into
s.

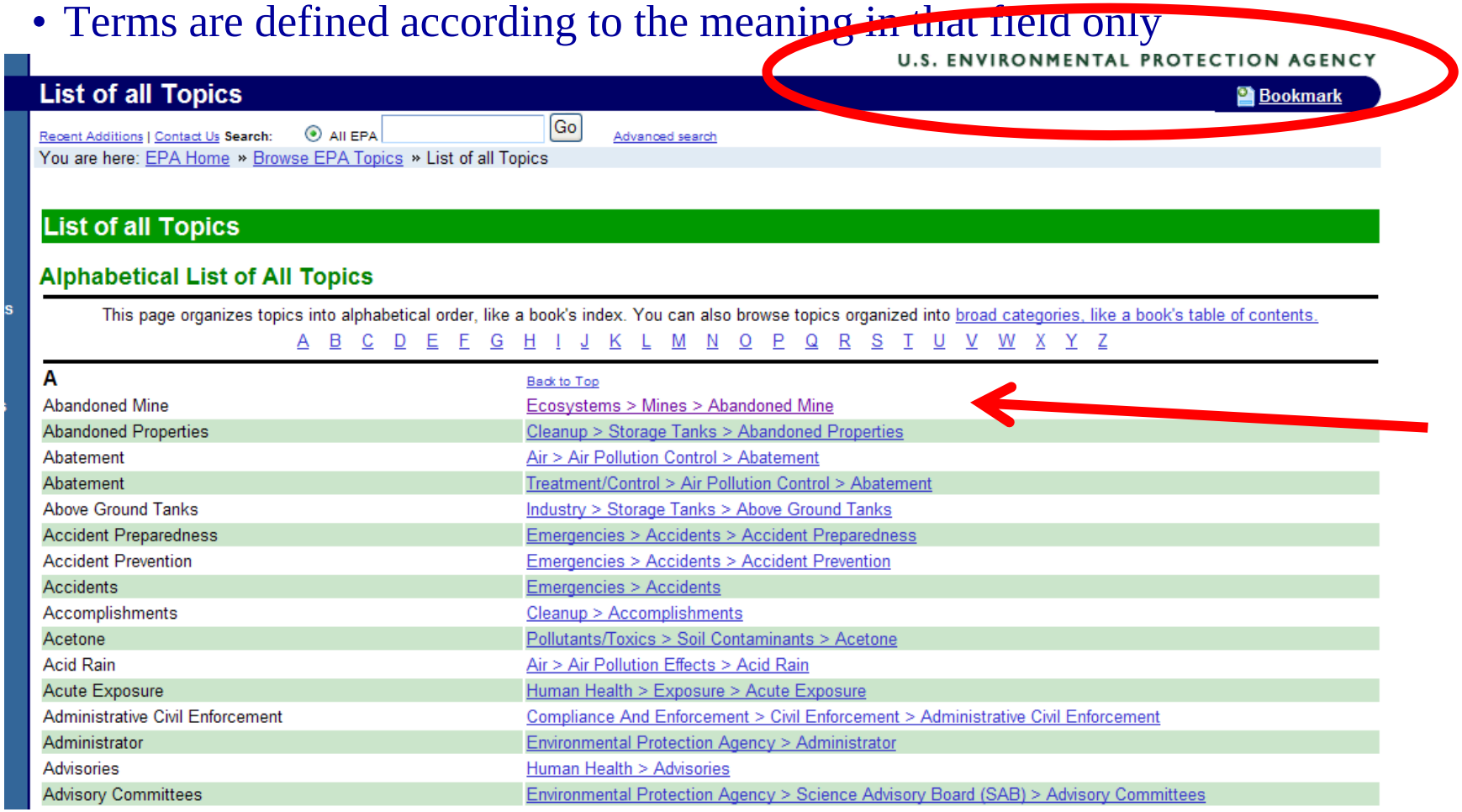
to the semantics of the grammatical units

– Relate specific linguistic information with other kinds of information

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

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)*

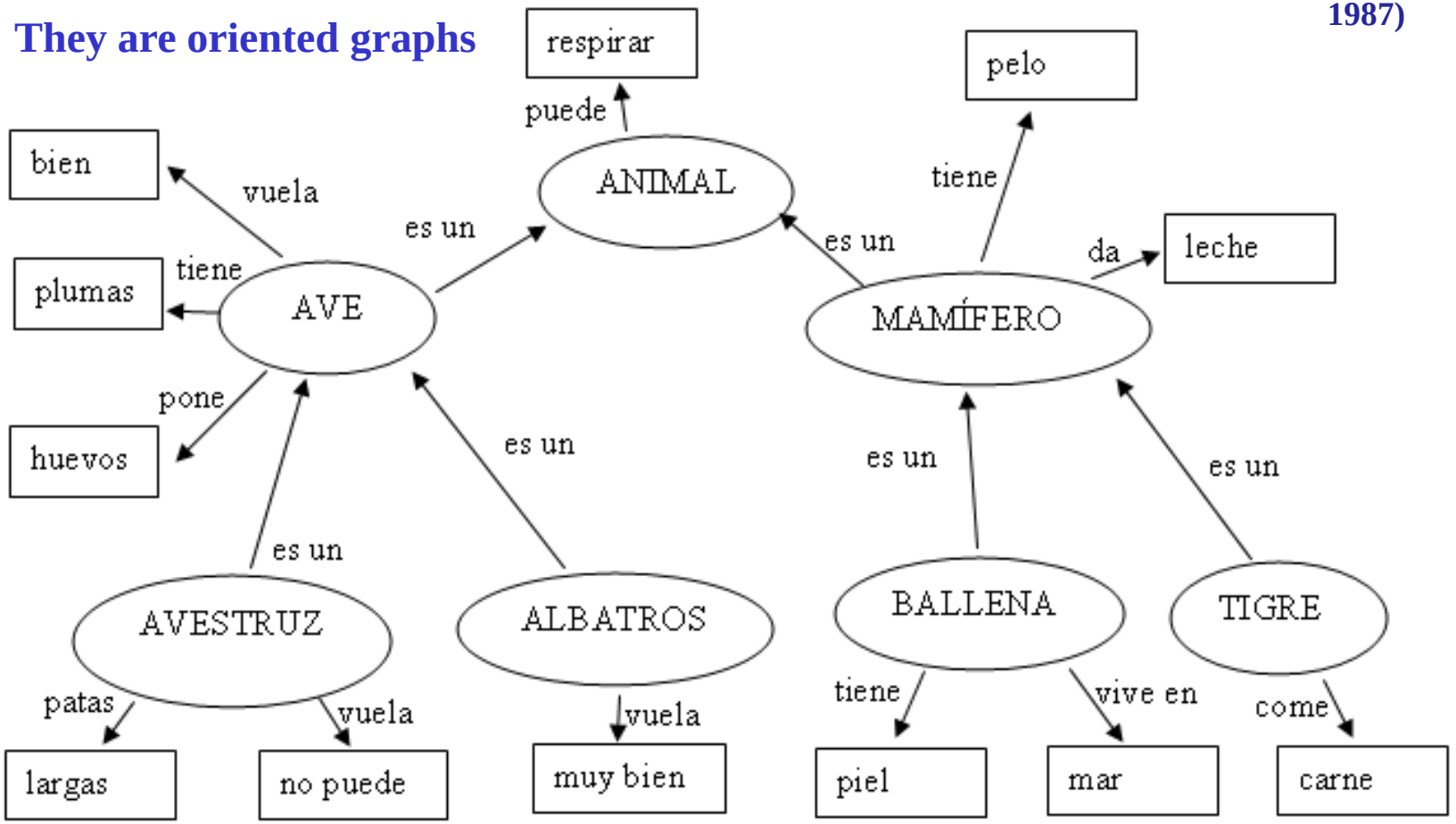
1 2 3 4 5 6 [Siguiente] [Fin]

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

Semantic network

- First approach: Quillian 1968 in AI.
- Concept structure with nodes and relations, not hierarchically organized
- Can include BT, NT, RT relations or other associative relations
- They are oriented graphs

Used in MT:
Fujitsu's ATLAS En-Jp Translation System (Uchida 1987)



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

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

How are lexicons used in NLP?

- They contain the necessary linguistic information to construct meaning representations

Lexicon

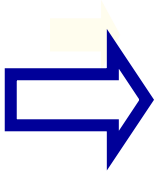
Account *n.* Domain [financial]

Account *v.* …

Bank_1 *n domain* [financial]

Bank_2 *n domain* [geography]

Money n. domain [financial]



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))))

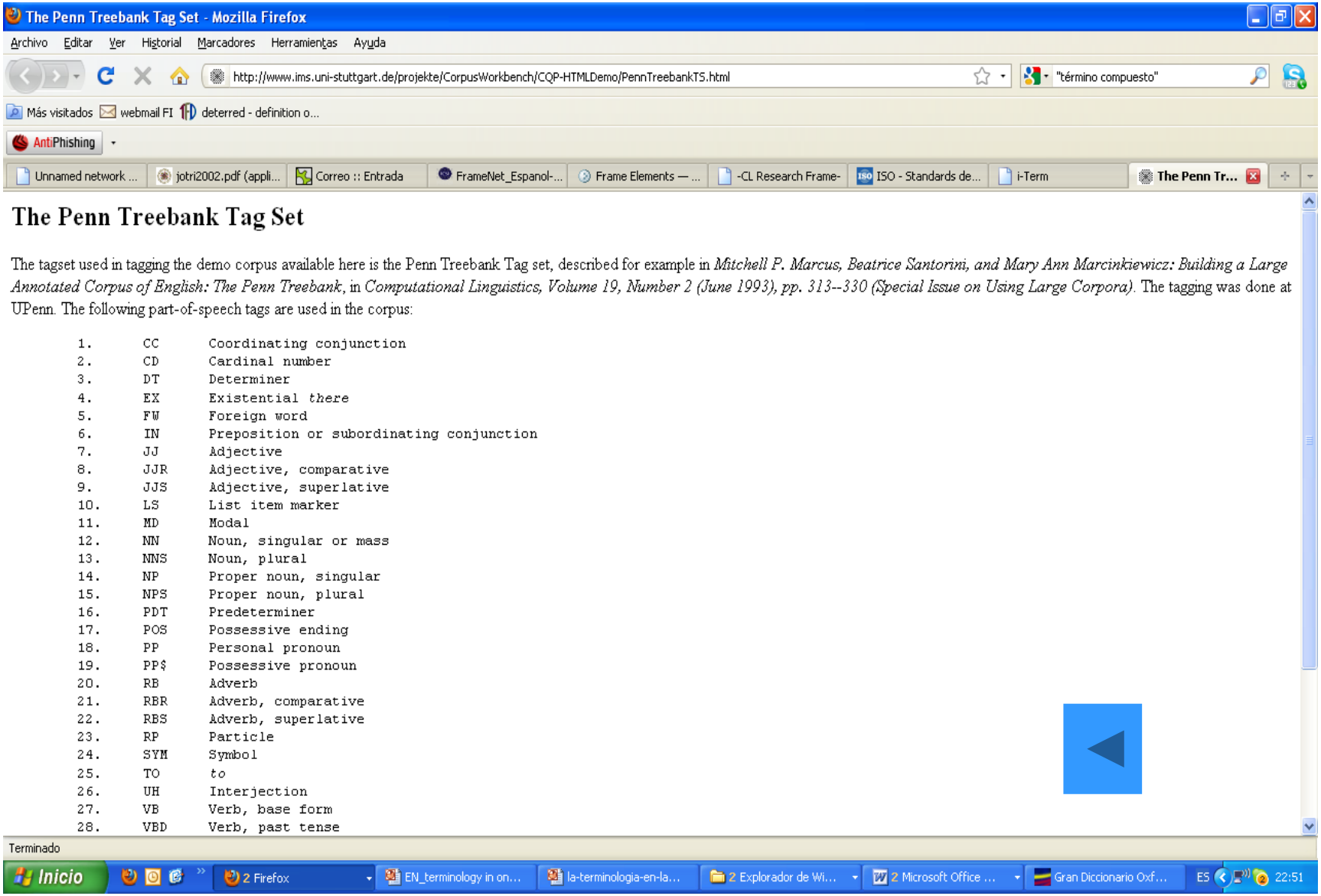
John *n.* sem: human

Store *n.* sem: loc

John went to the store

Adapted from Nancy Ide

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

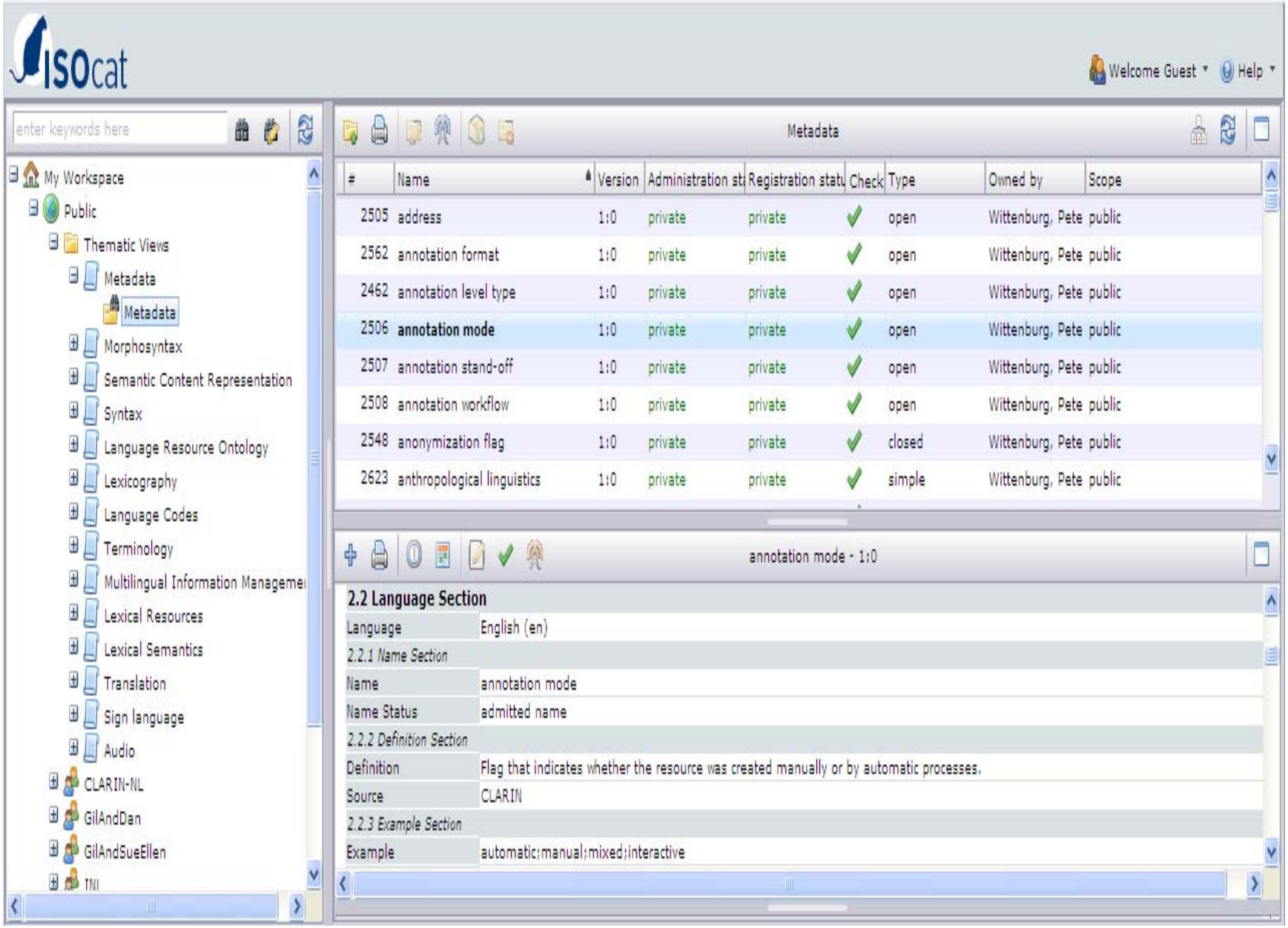
befoulment, defilement, **pollution** -- (the state of being polluted)

- 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)


ISOcat
Welcome Guest
Help
enter keywords here
My Workspace
Public
Thematic Views
Metadata
Metadata
Morphosyntax
Semantic Content Representation
Syntax
Language Resource Ontology
Lexicography
Language Codes
Terminology
Multilingual Information Manageme
Lexical Resources
Lexical Semantics
Translation
Sign language
Audio
CLARIN-NL
GilAndDan
GilAndSueEllen
TNI
Metadata
Name Version Administration st Registration statu Check Type Owned by Scope
2505 address 1:0 private private open Wittenburg, Pete public
2562 annotation format 1:0 private private open Wittenburg, Pete public
2462 annotation level type 1:0 private private open Wittenburg, Pete public
2506 annotation mode 1:0 private private open Wittenburg, Pete public
2507 annotation stand-off 1:0 private private open Wittenburg, Pete public
2508 annotation workflow 1:0 private private open Wittenburg, Pete public
2548 anonymization flag 1:0 private private closed Wittenburg, Pete public
2623 anthropological linguistics 1:0 private private simple Wittenburg, Pete public
annotation mode - 1:0
2.2 Language Section
Language English (en)
2.2.1 Name Section
Name annotation mode
Name Status admitted name
2.2.2 Definition Section
Definition Flag that indicates whether the resource was created manually or by automatic processes.
Source CLARIN
2.2.3 Example Section
Example automatic;manual;mixed;interactive

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

UF

BT

TT

RT

PT

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

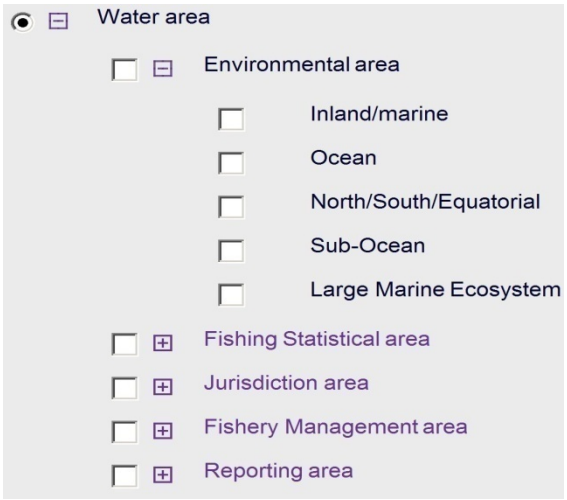
Types of relations

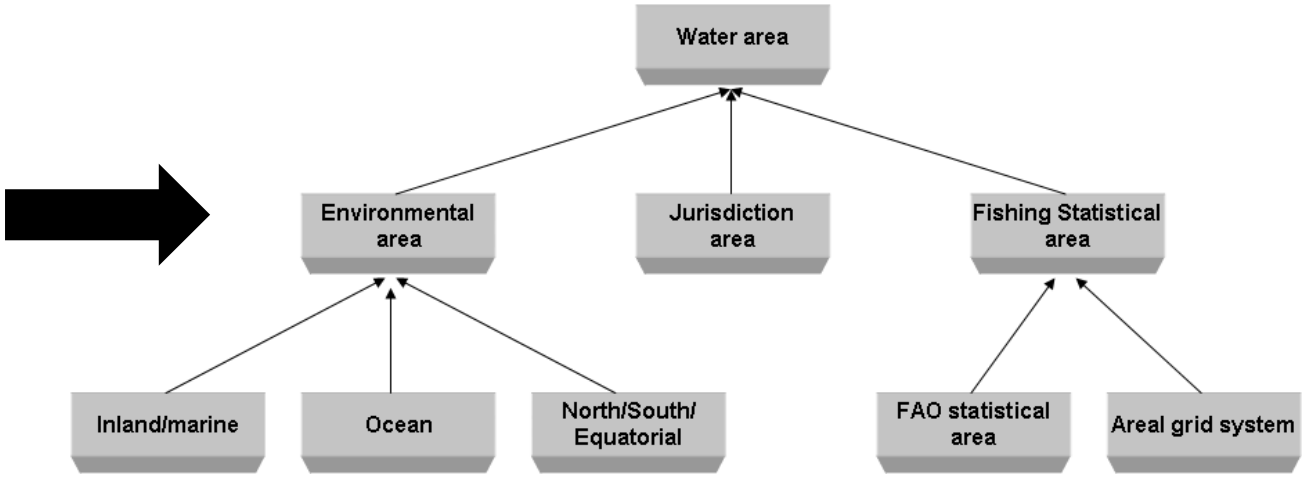
Classification systems (CS)

- Used as taxonomies:
  - Systema Naturae by Linné used in biology
  - International Statistical Classification of Diseases and Related Health Problems by the WHO
  - World Dewey Decimal Classification used for library classification
  - International Classification for Standards by ISO
- Purpose
  - e-business and e-procurement
  - Semantic Web (also WWW)
- Construction, maintenance and use of CS influenced by:
  - the classes and their properties
  - the class names and identifiers
  - the structure of the classification system.
  - BUT this knowledge should be made accessible to computers

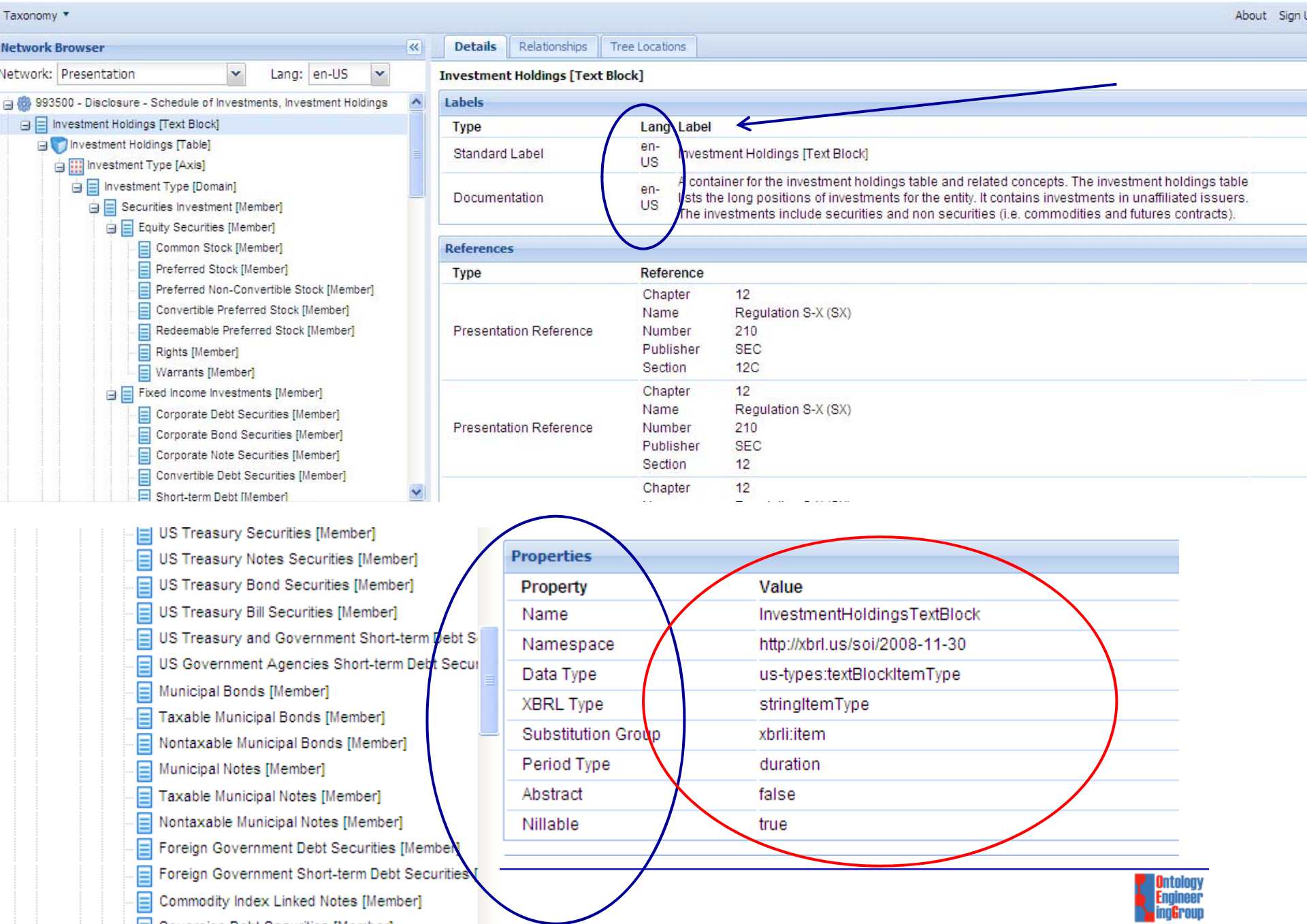
Classification Scheme

- 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


Taxonomy
About
Network Browser
Network: Presentation
Lang: en-US
993500 - Disclosure - Schedule of Investments, Investment Holdings
Investment Holdings [Text Block]
Investment Holdings [Table]
Investment Type [Axis]
Investment Type [Domain]
Securities Investment [Member]
Equity Securities [Member]
Common Stock [Member]
Preferred Stock [Member]
Preferred Non-Convertible Stock [Member]
Convertible Preferred Stock [Member]
Redeemable Preferred Stock [Member]
Rights [Member]
Warrants [Member]
Fixed Income Investments [Member]
Corporate Debt Securities [Member]
Corporate Bond Securities [Member]
Corporate Note Securities [Member]
Convertible Debt Securities [Member]
Short-term Debt [Member]
US Treasury Securities [Member]
US Treasury Notes Securities [Member]
US Treasury Bond Securities [Member]
US Treasury Bill Securities [Member]
US Treasury and Government Short-term Debt S
US Government Agencies Short-term Debt Secu
Municipal Bonds [Member]
Taxable Municipal Bonds [Member]
Nontaxable Municipal Bonds [Member]
Municipal Notes [Member]
Taxable Municipal Notes [Member]
Nontaxable Municipal Notes [Member]
Foreign Government Debt Securities [Member]
Foreign Government Short-term Debt Securities
Commodity Index Linked Notes [Member]
Details
Relationships
Tree Locations
Investment Holdings [Text Block]
Labels
Type
Lang
Label
Standard Label
en-US
Investment Holdings [Text Block]
Documentation
en-US
A container for the investment holdings table and related concepts. The investment holdings table lists the long positions of investments for the entity. It contains investments in unaffiliated issuers. The investments include securities and non securities (i.e. commodities and futures contracts).
References
Type
Reference
Presentation Reference
Chapter 12
Name Regulation S-X (SX)
Number 210
Publisher SEC
Section 12C
Presentation Reference
Chapter 12
Name Regulation S-X (SX)
Number 210
Publisher SEC
Section 12
Chapter 12
Properties
Property
Value
Name
InvestmentHoldingsTextBlock
Namespace
http://xbrl.us/soi/2008-11-30
Data Type
us-types:textBlockItemType
XBRL Type
stringItemType
Substitution Group
xbrli:item
Period Type
duration
Abstract
false
Nillable
true

Reusing and sharing KOS in the Semantic Web: SKOS

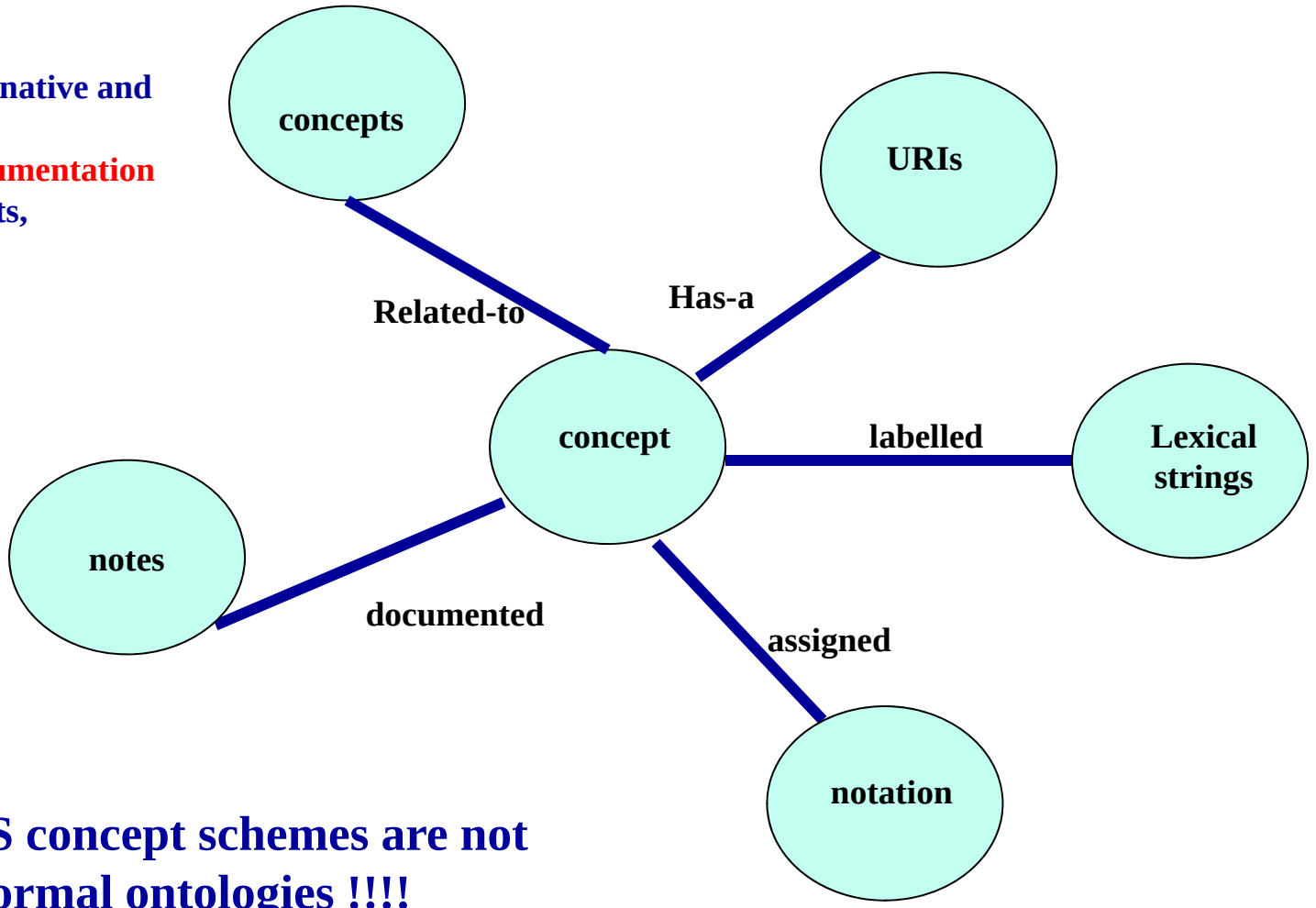
- Simple Knowledge Organization System
  - Simple, flexible, extensible, machine-understandable representation for sharing KOS
  - Goal: to enable easy publication of controlled structured vocabularies for the semantic web
    - Thesauri
    - Classification schemes
    - Subject heading systems
   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

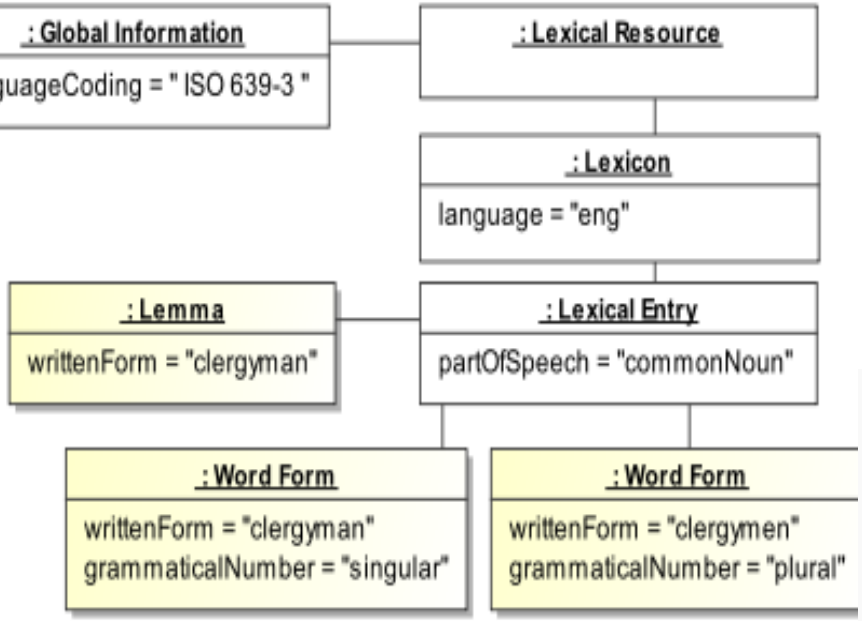
Lexical Markup Framework
LMF (1)

- ISO TC 37 standard for natural language processing (NLP) and machine-readable dictionary (MRD) lexicons
- AIM: standardization of principles and methods relating to language resources in the contexts of multilingual communication and cultural diversity

Lexical Markup Framework
LMF (2)

- HOW:
  - providing a common model for the creation and use of lexical resources
  - managing the exchange of data between and among these resources
  - enabling the merging of large number of individual electronic resources to form extensive global electronic resources.

Lexical Markup Framework
LMF (3)




```
<LexicalResource dtdVersion="15">
    <GlobalInformation>
        <feat att="languageCoding" val="ISO 639-3"/>
    </GlobalInformation>
    <Lexicon>
        <feat att="language" val="eng"/>
        <LexicalEntry>
            <feat att="partOfSpeech" val="commonNoun"/>
            <Lemma>
                <feat att="writtenForm" val="clergyman"/>
            </Lemma>
            <WordForm>
                 <feat att="writtenForm" val="clergyman"/>
                 <feat att="grammaticalNumber" val="singular"/>
            </WordForm>
            <WordForm>
                <feat att="writtenForm" val="clergymen"/>
                <feat att="grammaticalNumber" val="plural"/>
            </WordForm>
        </LexicalEntry>
    </Lexicon>
</LexicalResource>
```

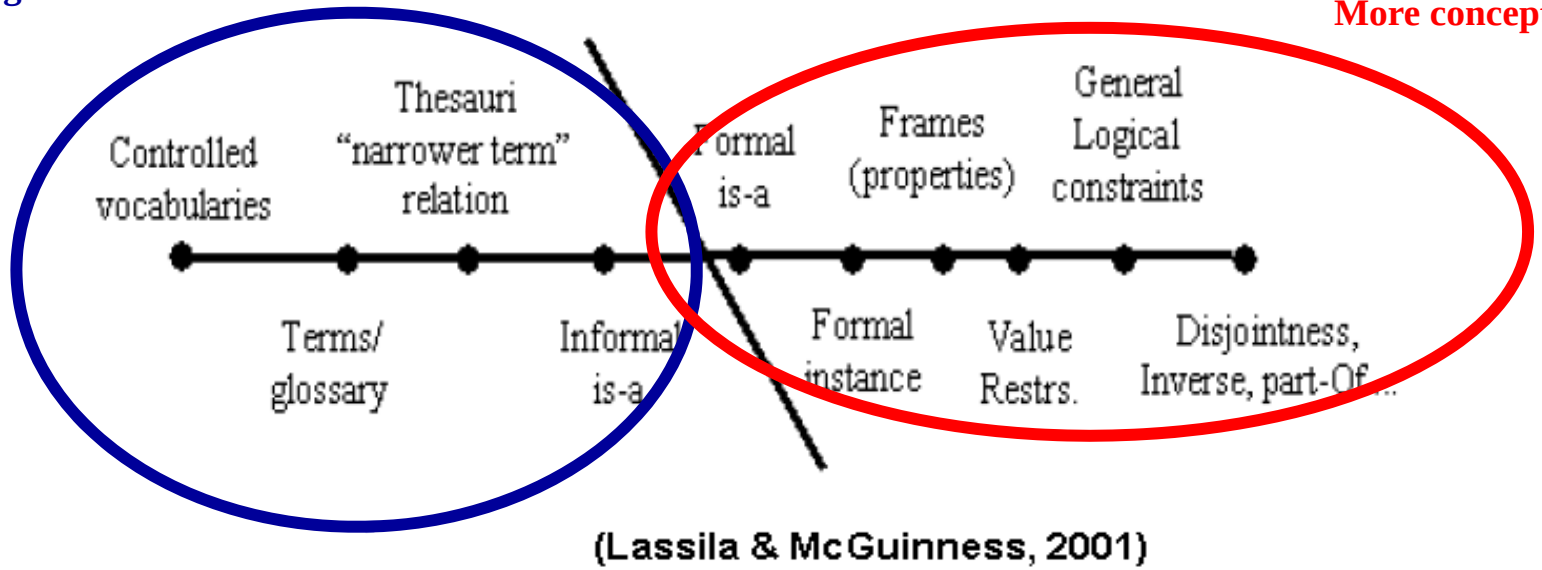
What is an ontology?

- "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.**

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

Automatization in building taxonomies

http://www.multites.com/onlinethesauri/

http://www.termtree.com.au/

http://www.autonomy.com/content/Products/products-idol-server/index.en.html

Recommended papers
Terminology and ontologies, October 2010

- Read three papers from the six proposed in the wiki. Comment on the ideas proposed and give your opinion in your own words.
- Vallez, Mari; Rovira, Cristòfol, Codina, Lluís; Pedraza, Rafael (2010). "Procedimientos para la extracción de palabras clave de páginas web basados en criterios de posicionamiento en buscadores". *Hipertext.net*, 8, http://www.upf.edu/hipertextnet/numero-8/extraccion_keywords.html
- 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