Ontologies

German Rigau i Claramunt

german.rigau@ehu.es

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU

Outline

- Introduction
- Mikrokosmos
- SUMO
- Cyc, OpenCyc

What is an Ontology?

- An ontology is an explicit specification of a conceptualization (Gruber 93)
- A conceptualization is an abstract, simplified view of the world represented for some purpose
- An ontology is a description (formal specification) of a set of concepts and relationships for enabling knowledge sharing and reuse (to perform logical commintments)
- An ontology commintment is an agreement to use a vocabulary in a way that is consistent with respect to the theory specified by the ontology

What is an Ontology?

- “A specific artifact designed with the purpose of expressing the intended meaning of a (shared) vocabulary” (Guarino 03)
- “In philosophy, **ontology** (from the Greek ων = *being* and λογοσ = *word/speech*) is the most fundamental branch of metaphysics. It is the study of being or existence as well as the basic categories thereof — trying to find out what entities and what types of entities exist. Ontology has strong implications for the conceptions of reality.” (from Wikipedia)
- “Ontology” dates to 17th century; meta-physics back to Aristotle

Disclaimer

- Ontologies is a **BIG** topic!
- Main focus on NLP and NLU
- Learn the basics from the experts:
  - Gruber papers: Various papers on the web
  - Sowa book: *Knowledge Representation*
  - Guarino tutorial: *Ontology-driven conceptual modeling* and various papers
  - Hovy tutorial!

Why use an Ontology?

- You need an unambiguous set of symbols for semantic representations
  - (*eat John tiramisu*): which *eat*? which *John*?
- You need to organize your symbols and/or variables according to the way they are processed
  - nouns act differently from verbs in general
- Can you do without an ontology?
  - Of course you can: *most* of (today’s) statistical NLP
- Our definition for this course: An ontology is a data structure in which symbols that represent conceptualizations are defined and manipulated by (NLP) software.

What is inside an Ontology?

- **Concepts**: represent a conceptualization; the class of all the examples of that event or entity
  - *Hapiness, children*
- **Relations**: represent a relationship between concepts
  - *Colour-of, location-of*
- **Axioms**: express a necessary fact holding between concepts and relationships
  - *If X is mortal then X will die one day*

T-Box

- **Instances**: represent a specific individual
  - *Albert Einstein*

A-Box

…but what about Beethoven's 9th symphony?

Content building steps (1)

- List **terms** that denote the entities, events, qualities, relationships, etc. in the domain
- Link them using one or more **relations**:
  - *structuring* relations (subsumption, others)
  - *definitional* relations
  - *additional info* relations
- Define **axioms** and properties
  - rules that specify what must be true about what
- Provide **additional information** resources:
  - lexicons, glossaries, documentation, etc.

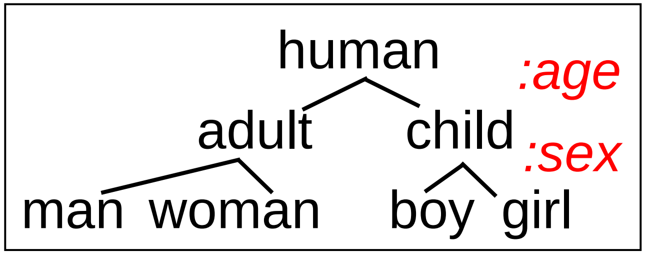
Terminology 'ontology' (e.g., WN)

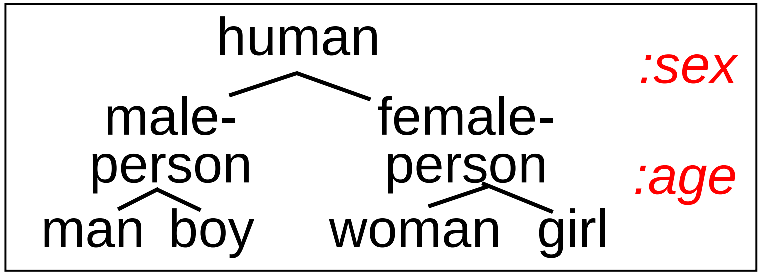
'True' ontology

External resources

Content building steps (2)

- Find a primitive concept (e.g. human)
- Specialize it in various ways by adding various differentiae
  - Ex: man, woman (:sex), adult, child (:age)
- Define these differentiae elsewhere in the ontology
- Don’t confuse **definitional** aspects with mere **properties**!
  - An *apple* is-a *fruit* with essential differentium *XXX* and with properties *:colour=red, :size=tennis-ball-sized…*
- Problems:
  - What are the differentiae?
  - How do you order them?





Ontologies
The prototypes (1)

- Functional purpose of classes: “provide maximum information with the least cognitive effort”
- Established experimental paradigms for determining how good an example of a category a member is judged to be
- **Basic Level** categories:
  - A basic category is the largest class of which we can form a fairly concrete image, like *chair* or *ball*. These are the first classifications that children make
  - Superordinate categories are collections of basic categories: *furniture* includes chairs, lamps, desks, beds, etc.; *toys* include balls, dolls, furry animals. No one object clearly represents them
  - Subordinate categories represent divisions of basic classes (*deck chairs, bar stools, teddy bears, school desks*)

Ontologies
The prototypes (2)

- people categorize using the common features of the members (differentiae)
- observations:

  (1) When people categorize, they cannot tell you what features they are using — often don’t know the differentiae!

  (2) When people categorize, they usually find some members of categories more “typical” (“better”) than others (e.g., a *robin* is a better member of the category *Bird* than an *ostrich*)

  (3) When people categorize, they categorize more typical members more quickly than less typical ones
- suggestion:
  - Create ‘star structure’ of prototypes rather than (or in addition to?) a subsumption hierarchy with differentiae

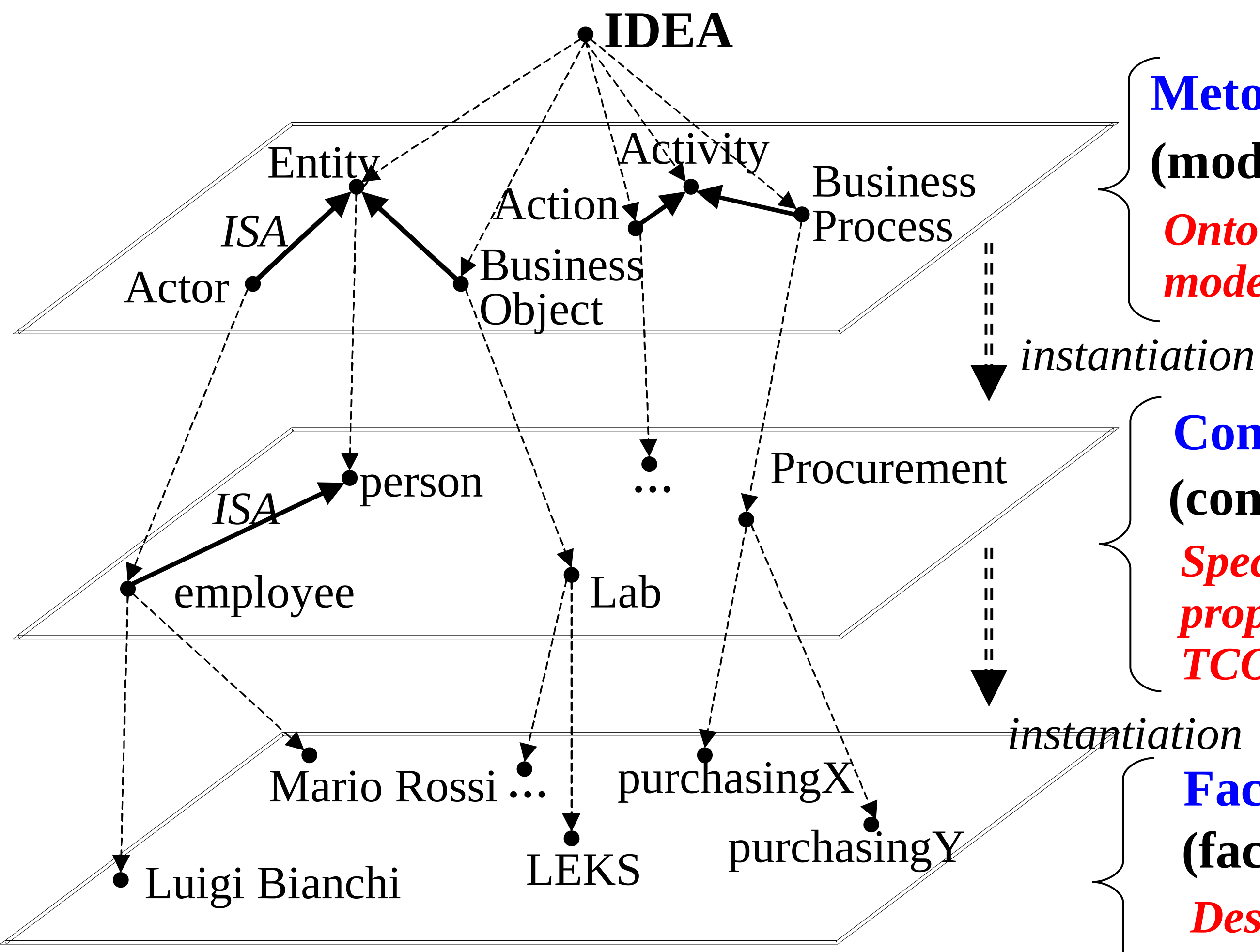
The prototypes (3)

- Base Concepts **BC** introduced in EuroWordNet.
- The BC are supposed to be the concepts that play the most important role in different languages. Two main criteria:
  - A high position in the semantic hierarchy (abstract)
  - Having many relations to other concepts (hub)
- Basic Level Concepts **BLC** are the result of a compromise between two conflicting principles of characterization:
  - Represent as many concepts as possible (abstract)
  - Represent as many distinctive features as possible (concrete)
- BC <> BLC

Domain ontologies

- Computational / expert systems:
  - **Protégé Ontologies Library**: Stanford University's collection of 18 influential ontologies (http://protege.stanford.edu/ontologies/ontologies.html)
  - **OntoSelect**: over 700 ontologies in various domains (http://views.dfki.de/Ontologies/)
- Medical:
  - **UMLS**: Metathesaurus (over 1 mill biomedical concepts and 5 mill concept names from over 100 controlled vocabularies and classifications (some in multiple languages) used in patient records, administrative health data, bibliographic and full-text databases, expert systems), the Semantic Network (*isa* for type hierarchy; *physically related, spatially related, temporally related, functionally related, conceptually related*), and the SPECIALIST lexicon (http://www.nlm.nih.gov/research/umls/)
- Industrial etc.:
  - **NAICS** (North American Industry Classification System): numerical classifications of construction, agriculture, technology, wholesale, retail, industry, etc., (http://www.census.gov/epcd/www/naics.html)

Ontologies
Levels of Knowledge



Metodological Knowledge

(modeling ideas: metatypes)

Ontology languages and models (OWL, KIF)

Conceptual Knowledge

(conceptual model: types)

Specific ontologies: classes and properties/relations (SUMO, TCO)

Factual Knowledge

(factual model: objects)

Description of specific data: Individuals and their relations

Ontologies

Conceptual and Factual Knowledge

Conceptual Knowledge (KR): Information to understand and process semantics:

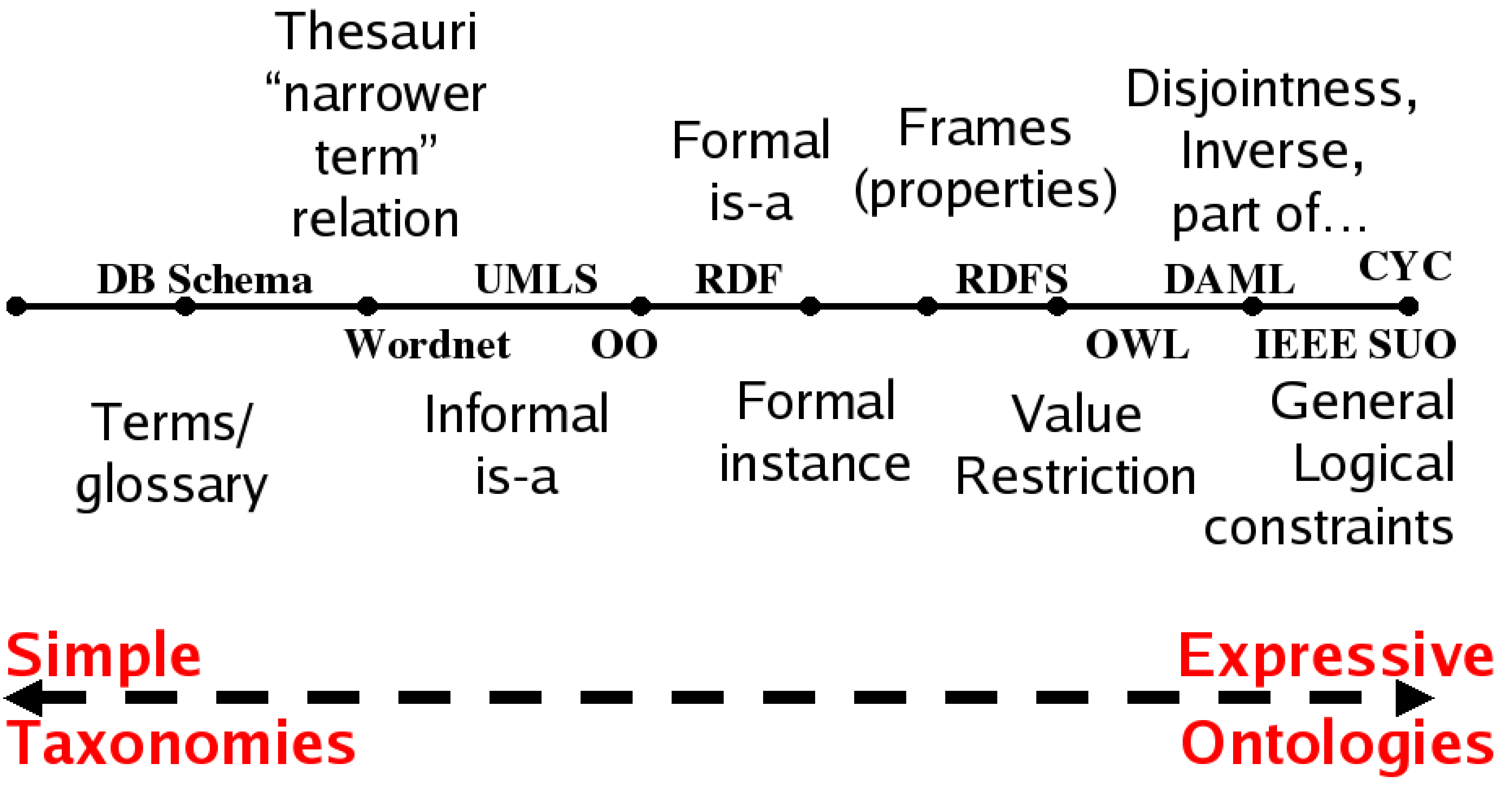
Knowledge, such as: an Hotel is composed by: a reception, some rooms, etc…

Factual Knowledge (FR): Information on the content of the concepts.

Data, such as: the Holyday Inn Hotel has 250 rooms, the prices are…

Ontologies

What kind of Ontologies?



Ontologies

Methodological knowledge: OWL Constructors

OWL Construct	DL	Example
intersectionOf	$C_1 \sqcap ... \sqcap C_n$	$Human \sqcap Male$
unionOf	$C_1 \sqcup ... \sqcup C_n$	$Doctor \sqcup Lawyer$
complementOf	$\neg C$	$\neg Male$
oneOf	$\{o_1, ..., o_n\}$	$\{john, mary\}$
allValuesFrom	$\forall P.C$	$\forall hasChild.Doctor$
someValuesFrom	$\exists P.C$	$\forall hasChild.Lawyer$
value	$\exists P.\{o\}$	$\exists citizenOf.USA$
minCardinality	$\geq nP.C$	$\geq 2 hasChild.Lawyer$
maxCardinality	$\leq nP.C$	$\leq 1 hasChild.Male$
cardinality	$= nP.C$	$= 1 hasParent.Female$

+ XML Schema datatypes: int, string, real, etc...

Conjunction
Disjunction
Negation
Choice between instances
Universal quantifier
Existential quantifier
Cardinality costraints
Inclusion between classes
Equivalence between classes
Inclusion between properties
Equivalence between propertie

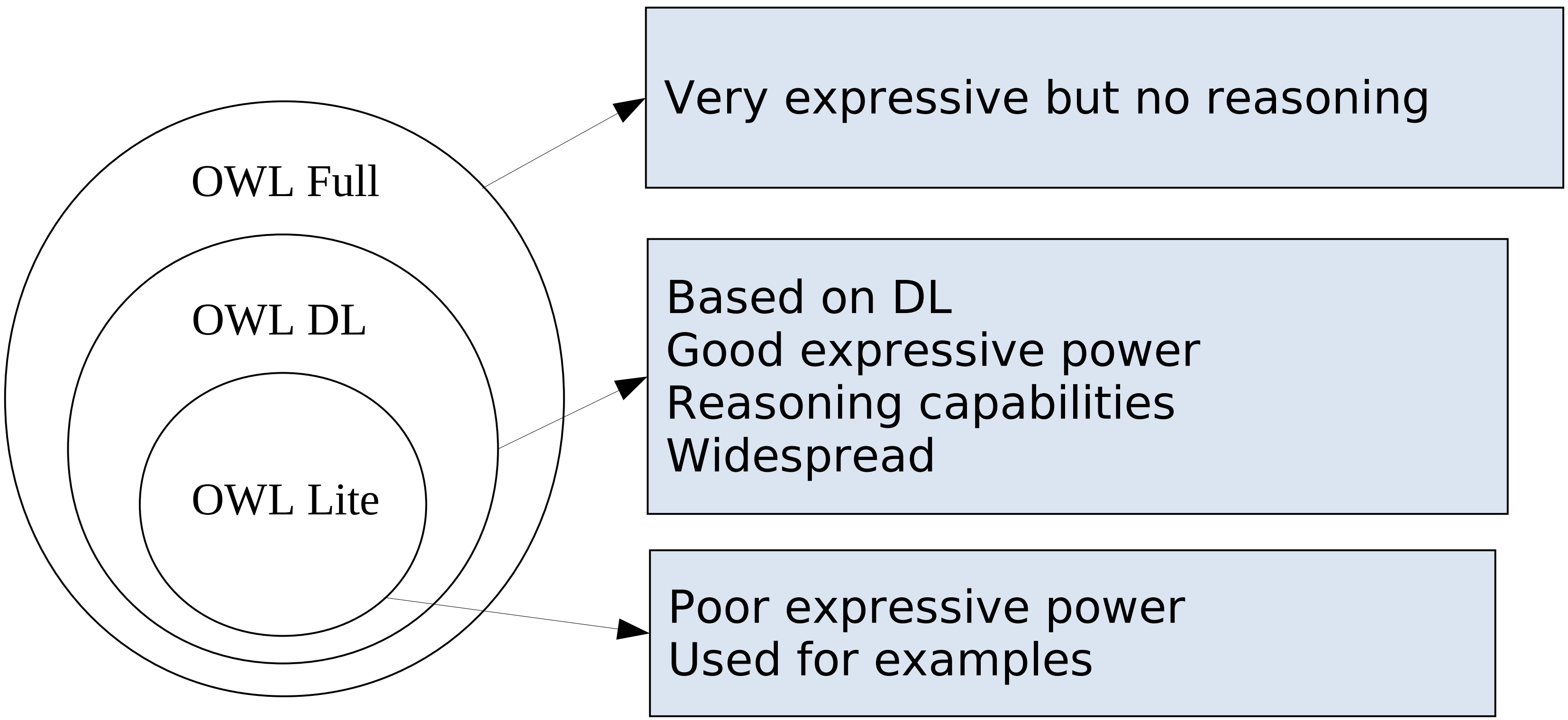
Ontologies
OWL: Web Ontology Language

OWL Axiom	DL	Example
SubClassOf	$C_1 \sqsubseteq C_2$	$Human \sqsubseteq Animal \sqcap Biped$
EquivalentClasses	$C_1 \equiv \ldots \equiv C_n$	$Man \equiv Human \sqcap Male$
SubPropertyOf	$P_1 \sqsubseteq P_2$	$hasDaughter \sqsubseteq hasChild$
EquivalentProperties	$P_1 \equiv \ldots \equiv P_n$	$cost \equiv price$
SameIndividual	$o_1 = \ldots = o_n$	$President\_Bush = G\_W\_Bush$
DisjointClasses	$C_i \sqsubseteq \neg C_j$	$Male \sqsubseteq \neg Female$
DifferentIndividuals	$o_i \neq o_j$	$john \neq peter$
inverseOf	$P_1 \equiv P_2^-$	$hasChild \equiv hasParent^-$
Transitive	$P^+ \sqsubseteq P$	$ancestor^+ \sqsubseteq ancestor$
Symmetric	$P \equiv P^-$	$connectedTo \equiv connectedTo^-$

Ontologies

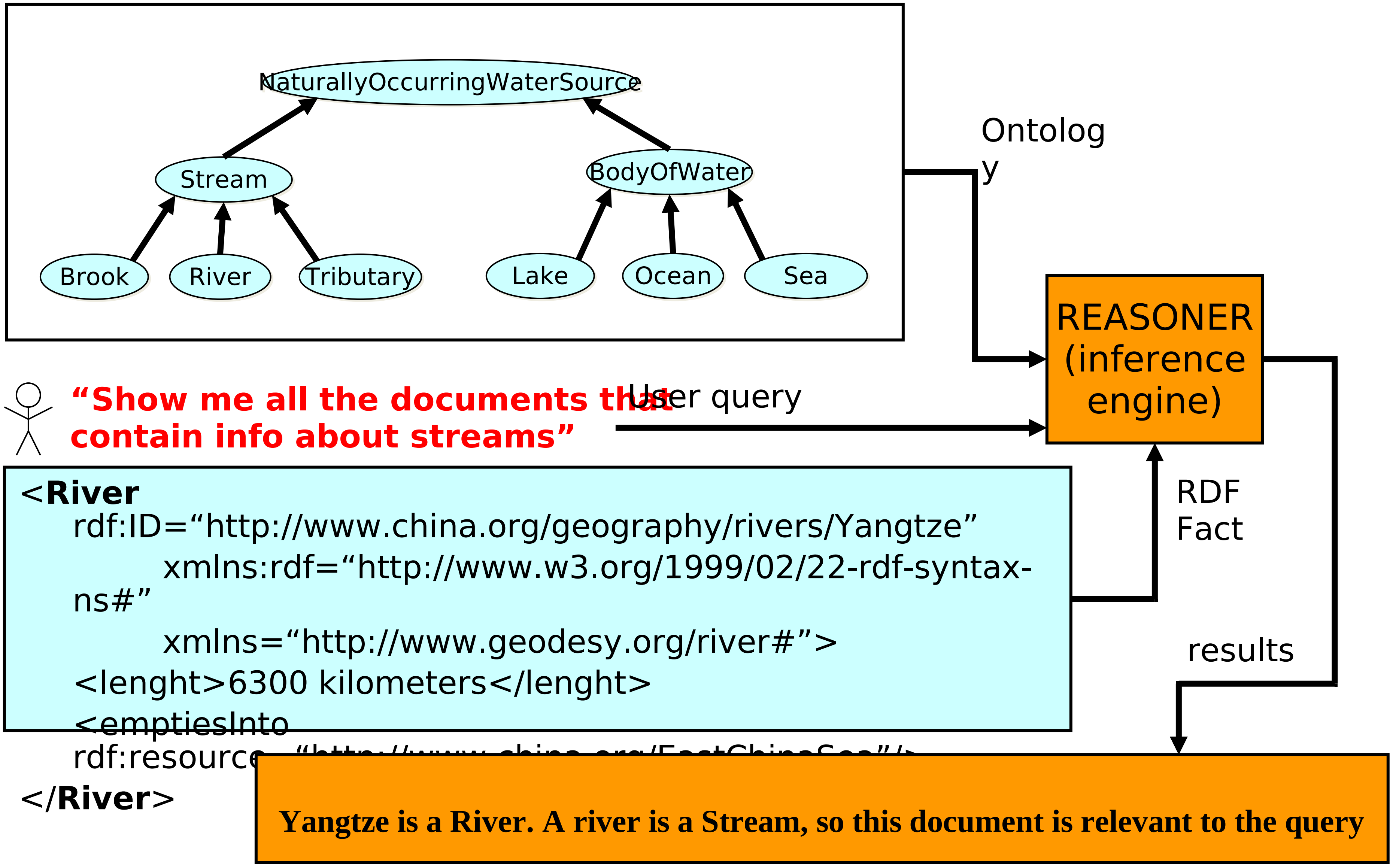
OWL: Web Ontology Language

3 versions with different complexity and expressive power



Ontologies

Processing knowledge through reasoning



Ontologies

Ontology languages

- First Order Logic
- Cycl, F-Logic, Loom, KIF, Ontolingua, Shoe, RDFs, OIL, OWL, ...
- Trade-off between
  - Expressive power
  - Reasoning power

- The following statements are not expressible in OWL-DL ...

  homeWorker(x) <- work(x,y) ^ live(x,z) ^ loc(y,w) ^loc(z,w)

  r(x,z) <- r(x,y) ^ r(y,z)

Ontologies
OWL tools

- Editor and Browsers
  - Protégé
  - SWOOP
  - Ontotrack
- Owl compliant reasoners
  - Pellet
  - Fact++

Ontologies

KIF: Knowledge Interchange Format

- More expressive than FOL
- Few tools for KIF
  - Sigma editor and browser
  - Vampire (Riazanov & Voronkov 2002)
  - E-prover
  - ...

Authority

- Who decides?
- Which features are the determinate ones?
- Why?

- There is no **authority**: it can be tradition, the law, social consensus, or simply ad hoc purpose-driven.
- The point is to know which to adopt and to **be careful and consistent**.

Outline

- Introduction
- Mikrokosmos
- SUMO
- CyC

Ontologies
Mikrokosmos

- Representational Issues
  - The Lexicon
  - The Ontology
- Acquisition Process
  - Lexicon Acquisition
  - Guidelines
  - Ontology/Lexicon Trade-off
- Semantics in Action

Mikrokosmos

Introduction

- Knowledge Base Machine Translation (KBMT)
- CRL, NMSU (Viegas et al. 96)
- 5,000 concepts
  - Events
  - Objects
  - Properties
- 7,000 Spanish word senses
- 40,000 word senses
  - after expansion with productive Lexical Rules
  - comprar -> comprador, comprable, ...
- Text Meaning Representation

Mikrokosmos

Representational Issues: The Lexicon

- Typed Feature Structures (Pollard and Sag 87)
- language-dependant
- 10 zones
  - phonology
  - orthography
  - morphology
  - Syntactic (subcategorization)
  - Semantic (Lexical Semantic Representation)
  - syntax-semantic linking
  - stylistics
  - paradigmatic
  - syntacmatic

Representational Issues: The Lexicon

Adquirir-V1

syn:	subj:	cat:	NP
	obj:	cat:	NP
sem:	acquire		
	agent:		HUMAN
	theme:		OBJECT

Adquirir-V2

syn:	subj:	cat:	NP
	obj:	cat:	NP
sem:	acquire		
	agent:		HUMAN
	theme:		INFORMATION

Representational Issues: The Ontology

- Taxonomic multi-hierarchical
- 14 local or inherited links in average
- language-impartial
- EVENTS, OBJECTS, PROPERTIES
- Methodology & Guidelines

Mikrokosmos

Representational Issues: The Ontology

ACQUIRE

DEFINITION	“The transfer of possession event where the agent transfers an object to its possession”
IS - A	TRANSFER-POSSESSION
SOURCE	HUMAN PLACE
THEME	OBJECT (NOT HUMAN)
AGENT	ANIMAL (DEFAULT HUMAN)
DESTINATION	ANIMAL PLACE (DEFAULT HUMAN)

INHERITED

BENEFICIARY	HUMAN

Mikrokosmos

Acquisition Process: The Lexicon

- Multi-lingual
  - French, English, Japanese, Russian, Spanish, etc.
- Multi-media
- Multi-process
  - Analysis
  - Generation (mono and multilingual)
  - MT
  - Summarization
  - IE
  - Speech Processing
- Tools
  - corpus-search, lookup dictionary, ontology browser

Mikrokosmos

Acquisition Process: The Ontology

Guidelines:

1) Do not add instances as concepts
   - Instances do not have their own instances
   - Concepts do not have fixed position in space/time
2) Do not decompose concepts further
3) Use close concepts
4) Do not add EVENTs with particular arguments
5) Do not add concepts with instance-specific aspects, temporal relations
6) Do not add language-specific concepts
7) Do not add ontologycal concepts for collections

Acquisition Process: Ontology/Lexicon Trade-off

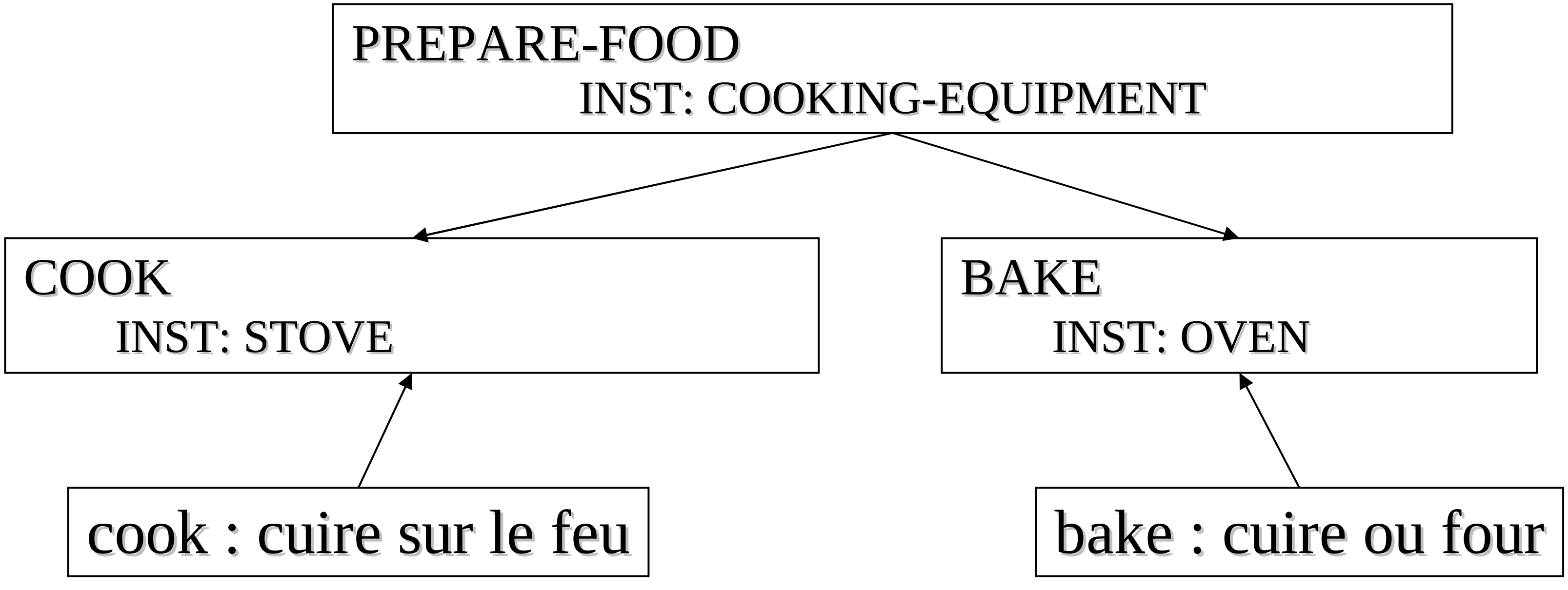
- Daily negociations
  - lexicon acquirers
  - ontology acquirers

- Possibilities
  - one-to-one mapping
  - lexicon unspecification
  - lexicon ontology balance

Mikrokosmos

Acquisition Process: Ontology/Lexicon Trade-off

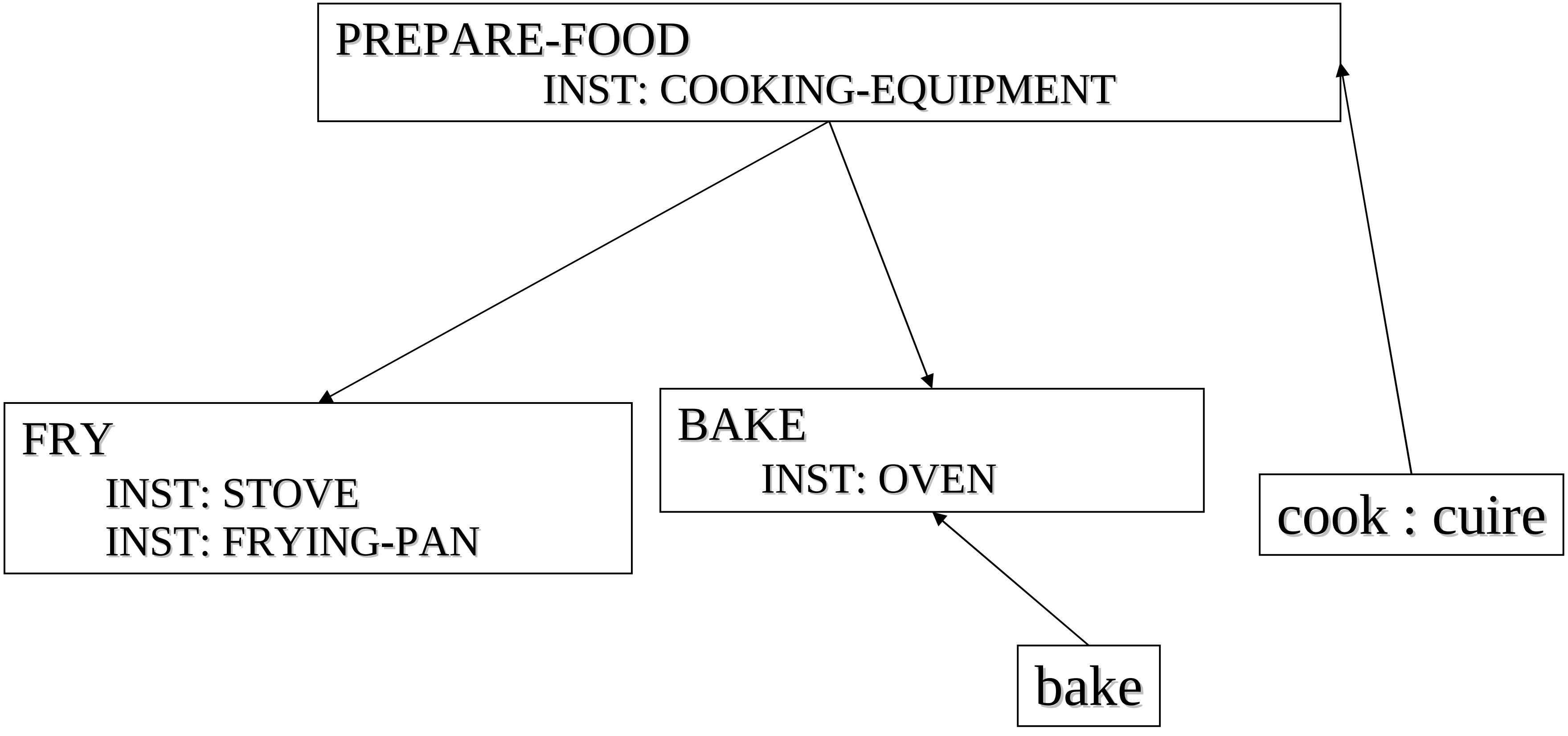
- one-to-one mapping



- Problems
  - Lexical: every word in a language is a concept
  - conceptual: *cuire* in french is not ambiguous

Acquisition Process: Ontology/Lexicon Trade-off

- Lexicon-Ontology Balance



Mikrokosmos

Semantics in Action

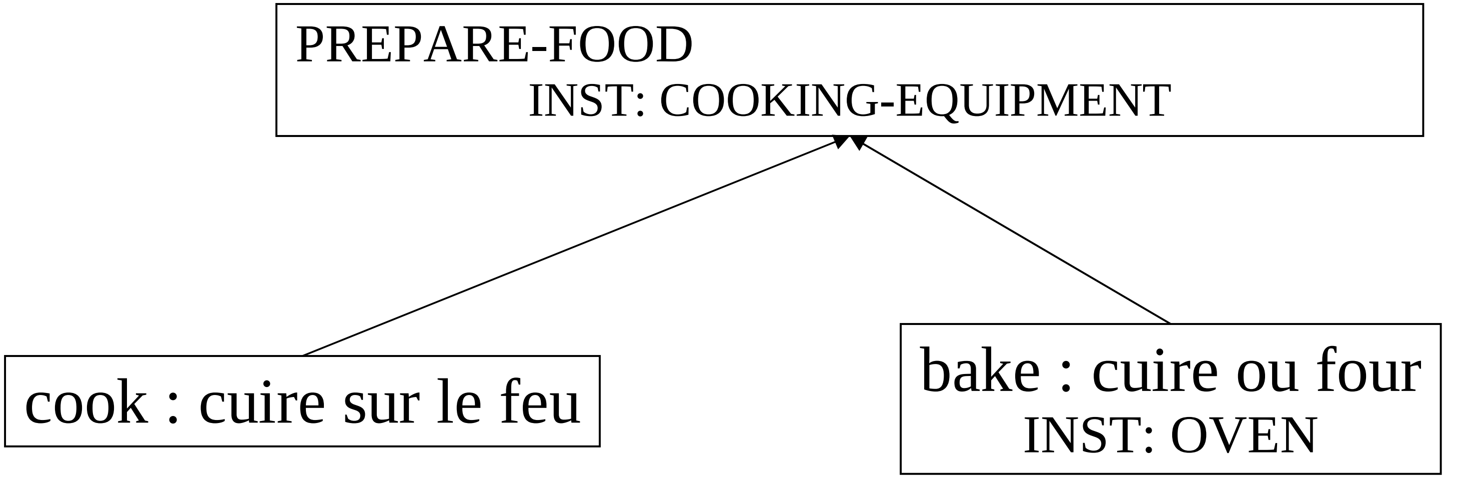
- El grupo Roche, a través de su compañía en España, adquirió Doctor Andreu.
- El grupo Roche adquirió Doctor Andreu a través de su compañía en España.
- La adquisición de Doctor Andreu por el grupo Roche fue hecha a través de su compañía en España.

ACQUIRE-1
Agent: ORGANIZATION-1
Theme: ORGANIZATION-2
Instrument: ORGANIZATION-3

ORGANIZATION-1	Object-Name: Grupo Roche
ORGANIZATION-2	Object-Name: Doctor Andreu
ORGANIZATION-3	Location: España

Acquisition Process: Ontology/Lexicon Trade-off

- Lexicon Unspecification



- Problems
    - BAKE is not in the ontology

Mikrokosmos

Semantics in Action

- Onto-Search:

  Ontological search mechanism to check constraints

- check-onto(ACQUIRE, EVENT) = 1
  - since ACQUIRE is a type of EVENT
- check-onto(ORGANIZATION, HUMAN) = 0.9
  - since ORGANIZATION HAS-MEMBER HUMAN

Mikrokosmos

Semantics in Action

1) *a-través-de* <u>INSTRUMENT</u>, LOCATION
   adquirir require PHYSICAL-OBJECT
2) *en* <u>LOCATION</u>, TEMPORAL
   España is not a TEMPORAL-OBJECT
3) *adquirir* <u>ACQUIRE</u>, LEARN
   Doctor Andreu is not an INFORMATION
4) *Doctor Andreu* <u>ORGANIZATION</u>, HUMAN
   the Theme of ACQUIRE is not HUMAN
5) *compañía* <u>CORPORATION</u>, SOCIAL-EVENT
   ORGANIZATIONs typically fill the INSTRUMENT slot of ACQUIRE acts

Experiment: WSD

Text	1	2	3	4	Mean
words	347	385	370	353	364
words/sentence	16.5	24.0	26.4	20.8	21.4
open-class words	183	167	177	177	176
ambiguous words	57	42	57	35	48
syntax	21	19	20	12	18
correct	51	41	45	34	43
%	97	99	93	99	97

Experiment: WSD

Text	Mean	Mean Unseen
words	364	390
words/sentence	21.4	26
open-class words	176	104
ambiguous words	48	26
syntax	18	9
correct	43	23
%	97	97

Outline

- Introduction
- Mikrokosmos
- SUMO
- CyC

Ontologies
SUMO

- Introduction
- Mapping SUMO to WordNet
- SIGMA
- Vampire & other Theorem provers

SUMO
Introduction

- The Suggested Upper Merged Ontology (SUMO)
- IEEE Standard Upper Ontology Working Group
- An upper ontology is limited to concepts that are meta, generic, abstract, general enough to address a broad range of domain areas.
- To promote:
  - Interoperability
  - Information Search and retrieval
  - Automated inference
  - NLP
  - Developement of Domain ontologies

SUMO

Introduction

- Incorporates over 50 publicly available sources of high-level ontological content
- May be used without fee for any purpose (including for profit)
- Refined extensively on the basis of input from SUO mailing list participants
- 42 publicly released versions (approximately 1,000 concepts, 4000 assertions, and 600 rules so far)

Mapping SUMO to WordNet

- Facilitate uses of the SUMO by those who lack extensive training in logic and mathematics
- Allows the SUMO to be used automatically by applications that process free text
- Completeness check on SUMO content
- Testing the SUMO with a state of the art theorem-prover
  - Redundancy
  - Contradiction

Mapping SUMO to WordNet

- Align noun, verb and adjective database (96,000 synsets) of WordNet 1.6 to SUMO concepts
  - synonymousExternalConcept =
  - subsumingExternalConcept +
  - Instance @

00008864 03 n 03 plant 0 flora 0 plant_life 0 . . . | a living organism lacking the power of locomotion &%Plant=

00048640 04 n 01 insider_trading 0 001 @ 00047814 n 0000 | buying or selling corporate stock by a corporate officer or other insider &%FinancialTransaction+

00821498 04 n 01 Actium 0 002 @ 00614512 n 0000 #p 06449758 n 0000 | naval battle where Antony and Cleopatra were defeated by Octavian's fleet under Agrippa in 31 BC &%Battle@

Domain Specific Ontologies

- Air force planning
- Finance and investment
- Real Estate
- Terrain features
- Computers and Networks (Quality of Service)
- Army planning
- ECommerce services
- Ontologies developed outside Teknowledge
  - Biological viruses
  - Intellectual property
  - Linguistic elements

Example: Boiling

- (subclass Boiling StateChange)
- (documentation Boiling "The Class of Processes where an Object is heated and converted from a Liquid to a Gas.")
-
```
(=>
    (instance ?BOIL Boiling)
    (exists
        (?HEAT)
        (and
            (instance ?HEAT Heating)
            (subProcess ?HEAT ?BOIL))))
```
- "if instance BOIL Boiling, then there exists HEAT such that instance HEAT Heating and subProcess HEAT BOIL"

SUMO

Example: Boiling

- ```
(=>
    (and
        (instance ?BOIL Boiling)
        (patient ?BOIL ?OBJ))
    (exists
        (?PART)
        (and
            (part ?PART ?OBJ)
            (holdsDuring
                (BeginFn
                    (WhenFn ?BOIL))
                (attribute ?PART Liquid))
            (holdsDuring
                (EndFn
                    (WhenFn ?BOIL))
                (attribute ?PART Gas)))))
```
- "if instance BOIL Boiling and patient BOIL OBJ, then there exists PART such that part PART OBJ and holdsDuring BeginFn WhenFn BOIL attribute PART Liquid and holdsDuring EndFn WhenFn BOIL attribute PART Gas"
```

SUMO

Reasoning with Sigma (Vampire)

(forall (?X) (or (not (instance ?X FloweringPlant)) (not (instance ?X BodyPart)))) YES.

(exists (?X) (and (instance ?X FloweringPlant) (instance ?X BodyPart)))
NO.

(forall (?Y) (forall (?X) (=> (equal ?X ?Y) (equal ?X ?Y)))) YES.

(forall (?X) (=> (instance ?X Flower) (instance ?X Organ))) YES.

(instance Ear FloweringPlant) NO.
(instance Ear Organ) NO.
(subclass Ear Organ) YES.

(forall (?X) (=> (not (subclass ?X Organ)) (subclass ?X Flower))) NO.

(forall (?X) (=> (subclass ?X Flower) (subclass ?X Organ))) YES.

SUMO
Reasoning with Sigma (Vampire) but ...

(forall (?X) (=> (instance ?X Organ) (instance ?X Flower))) YES!!!

(forall (?X) (=> (subclass ?X Organ) (subclass ?X Flower))) YES!!!

(forall (?X) (=> (not (subclass (?X) Organ)) (subclass (?X) Flower)))
YES !!!

(forall (?X) (or (not (subclass (?X) Organ)) (subclass (?X) Flower))))
YES!!!

(forall (?X) (or (not (instance ?X Organ)) (instance ?X Flower))) YES !!!

Ontologies

German Rigau i Claramunt
german.rigau@ehu.es
IXA group
Departamento de Lenguajes y Sistemas Informáticos
UPV/EHU