Webul Semantic

Ștefan Trăuşan-Matu

Webul semantic

- Următoarea generație a WWW (Tim Berners-Lee)
- Utilizat nu numai pentru a fi parcurs de utilizatorii umani prin “browsere” ci și de programe care:
  - accesează selectiv (caută și filtrează) paginile (resursele) web,
  - efectuează anumite prelucrări ale acestora,
  - eventual generează alte resurse web.

Webul semantic necesită:

- adnotarea şi metadescrierea resurselor web folosind limbaje bazate pe XML;
- reprezentarea și prelucrarea cunoştinţelor (de exemplu, prin logici terminologice sau decripţionale) - ***ONTOLOGII***;
- prelucrări ale conținutului documentelor web – ***Prelucrarea* Limbajului Natural**

Ontologiile

- Constituie repertorii de termeni, vocabulare pentru metadescrieri
- Sunt baze de cunoștințe declarative

Ontologiile

- în filosofie denotă teoria asupra existenței, mai corect spus, asupra ceea ce consideră că există cel care întocmește teoria;
- construirea oricărui sistem filosofic pleacă de la o ontologie - definirea categoriile fundamentale de entități din realitate și a relaţiilor dintre ele;
- nu este întotdeauna explicită, orice demers însă necesitând-o.

Ontologiile

"An ontology is a specification of a conceptualization....That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents" (Gruber)

Ontologiile conţin :

- categoriile, conceptele fundamentale,
- proprietăţile conceptelor,
- relaţiile şi distincţiile între concepte,
- axiome.

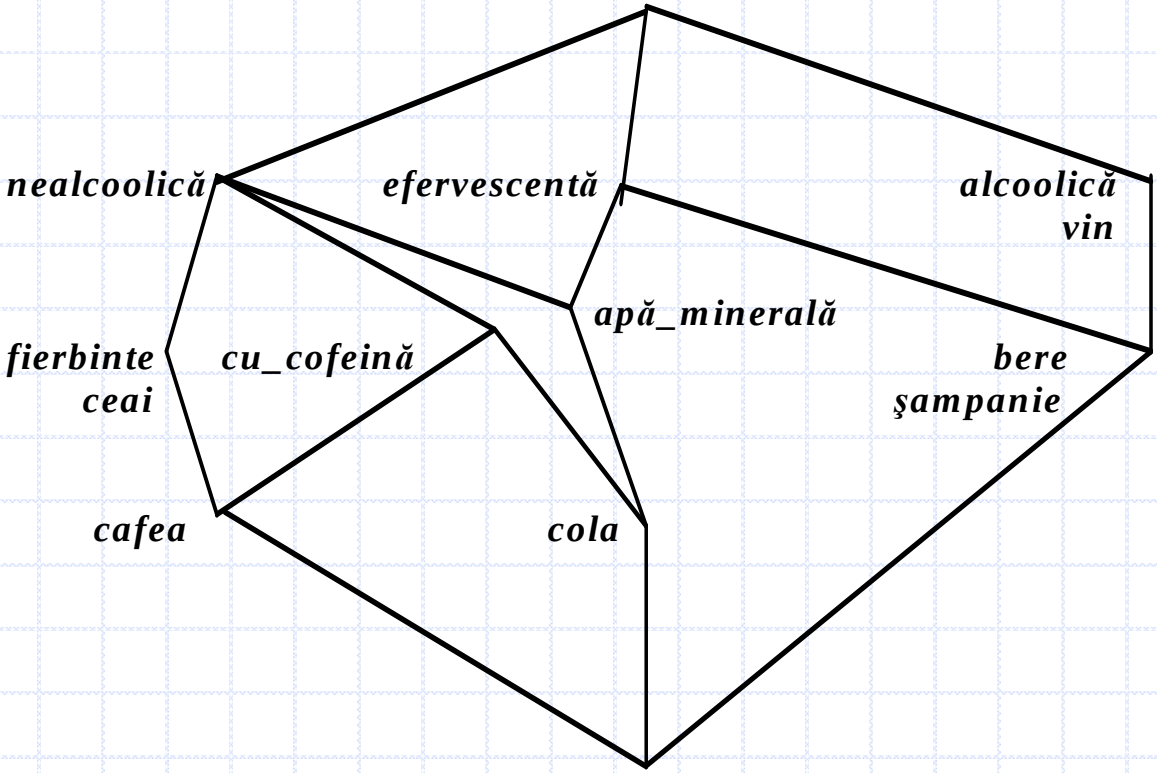
Ontologii

- Lexicalizate (WordNet, EuroWordNet, BalkanNet, FrameNet, MikroKosmos).
- Pentru reprezentarea cunoştinţelor.
- Implicite în, de exemplu, biblioteci de clase în OOP

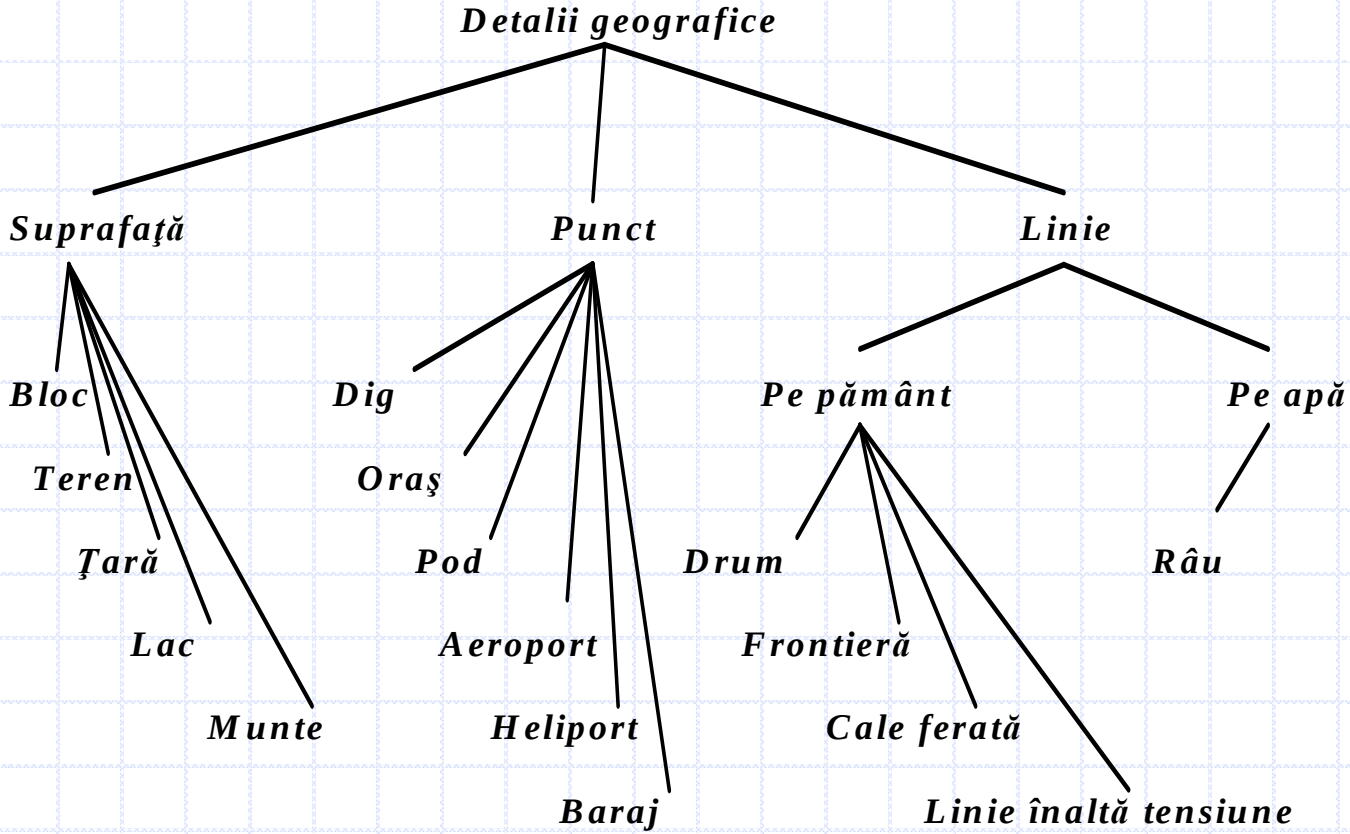
Metode pentru dezvoltarea ontologiilor

- Ad-hoc
- Analiza formală a conceptelor
- Psiholingvistică (WordNet)
- De la un tezaur, bază de date, o taxonomie
- Alinierera ontologiilor
- Extragere de cunoștințe din texte (text mining)
- Plecând de la categorii filosofice (e.g. Sowa)

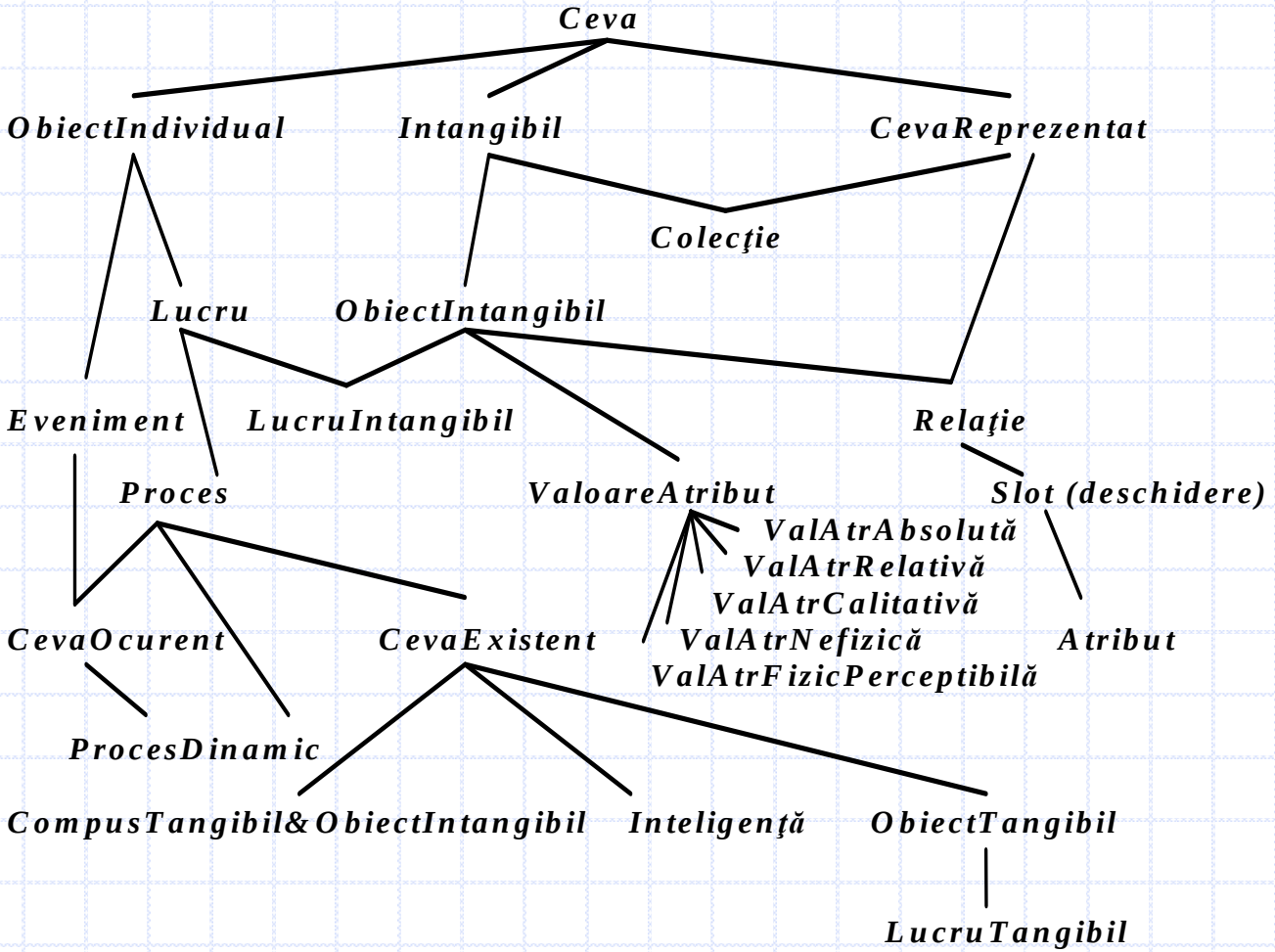
Latice construită prin metoda analizei formale a conceptelor (Sowa, 2001)



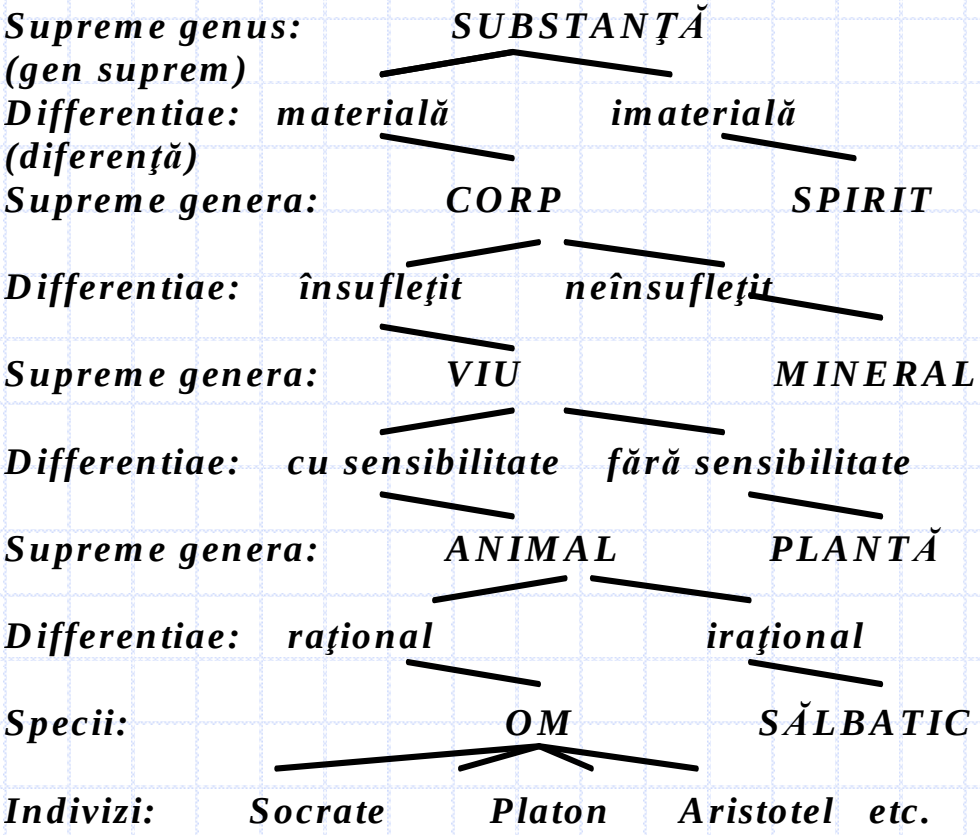
Categorii geografice - Chat-80
(Sowa, 2001)



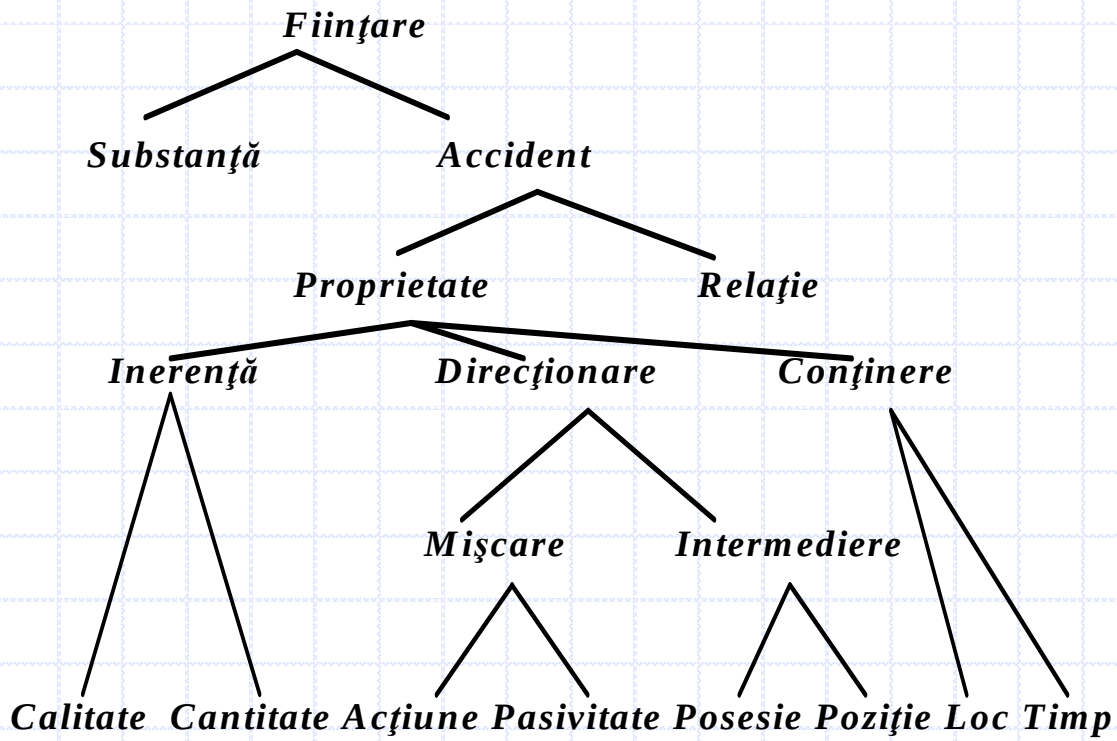
Categorii fundamentale în Cyc (Sowa, 2001)



Arborele lui Porfir
(traducere după Petrus Hispanus)



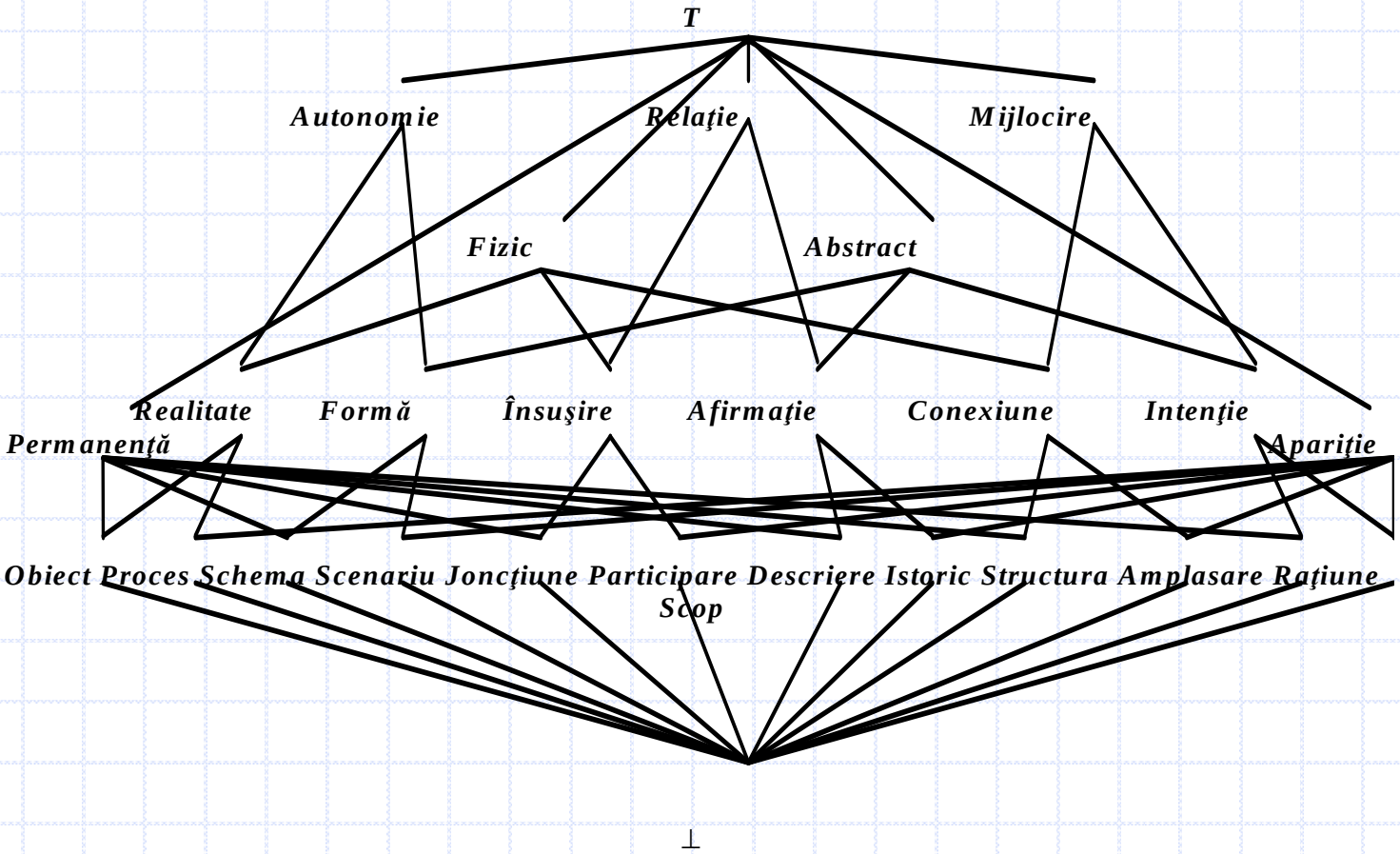
Arborele lui Brentano pentru categoriile lui Aristotel

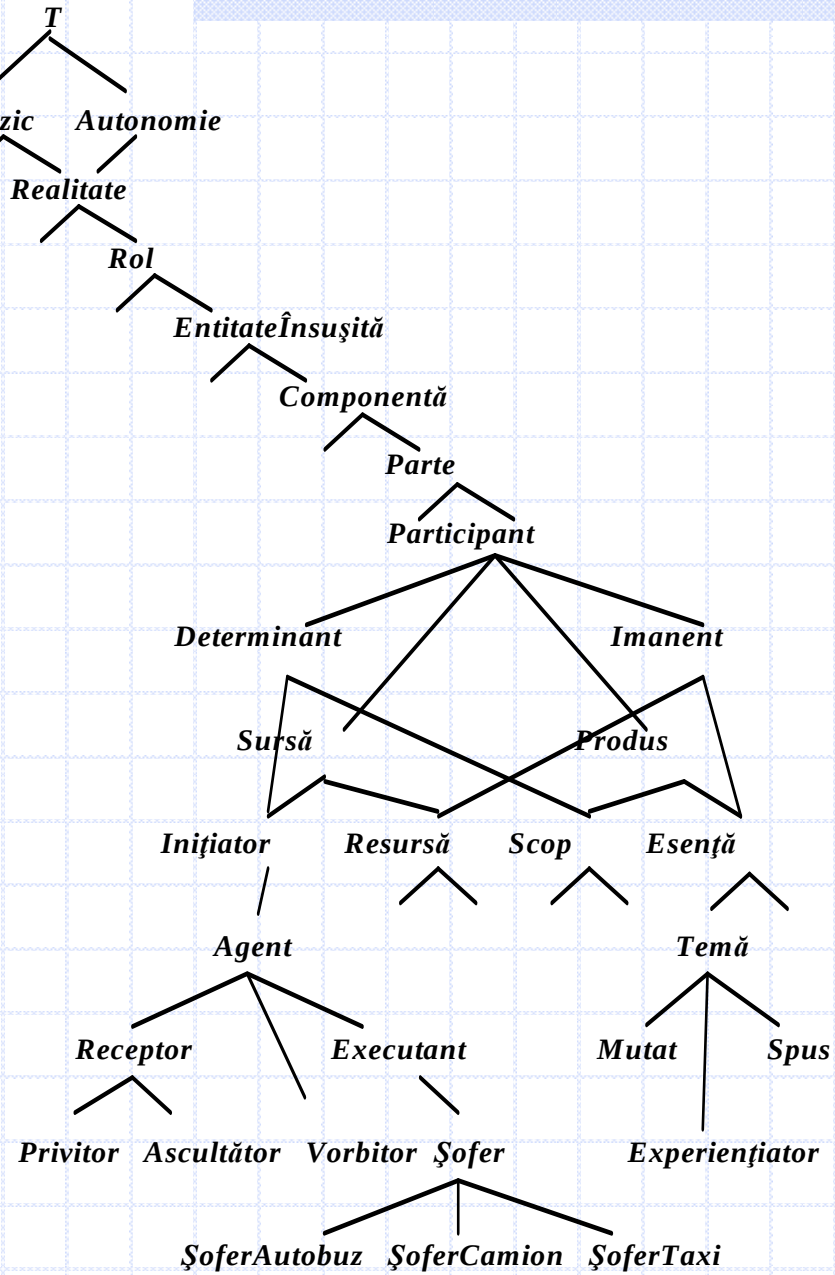


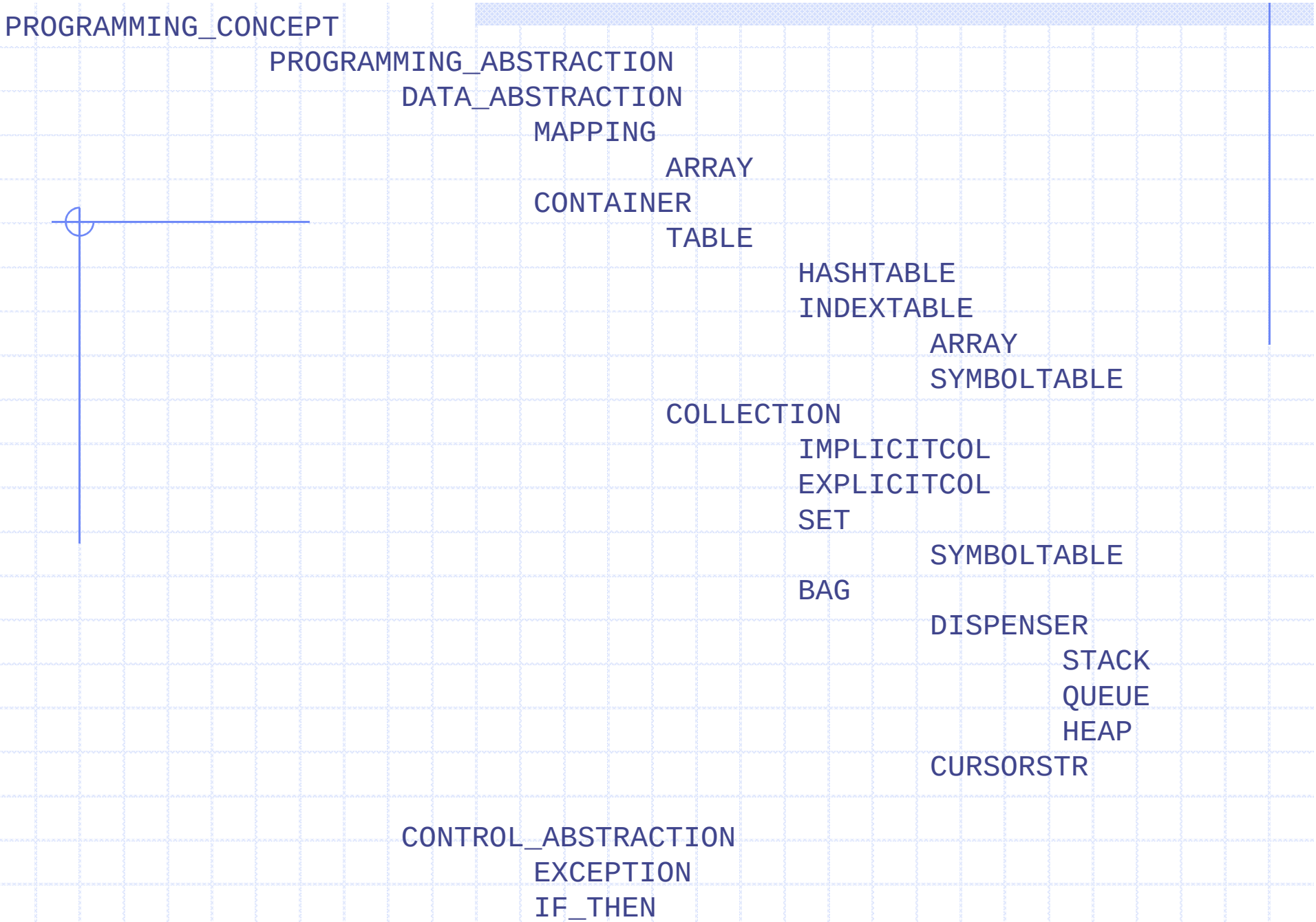
Categoriile lui Kant

Cantitate	*Calitate*	*Relaţie*	*Modalitate*
Unitate	*Existenţă*	*Substanţă*	*Posibilitate*
Pluralitate	*Inexistenţă*	*Cauzalitate*	*Realitate*
Totalitate	*Limitaţie*	*Comunitate*	*Necesitate*

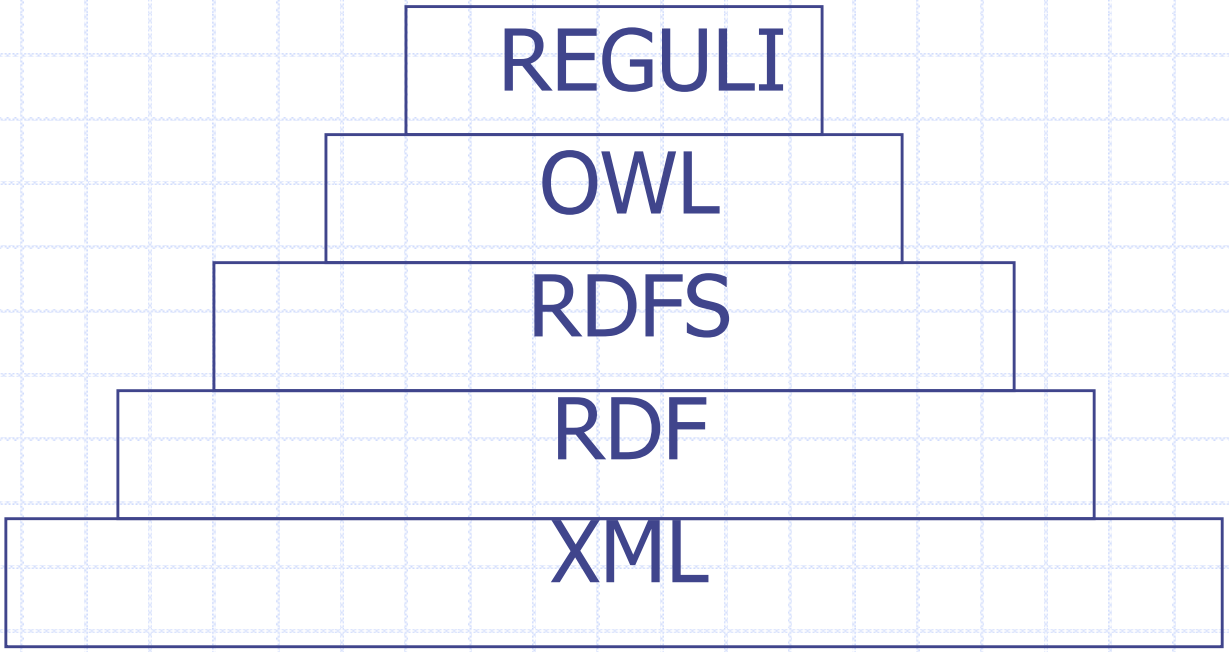
Ontologia lui John Sowa
(Sowa, 2001)




T
Fizic
Autonomie
Realitate
Rol
EntitateÎnsuşită
Componentă
Parte
Participant
Determinant
Imanent
Sursă
Produs
Iniţiator
Resursă
Scop
Esenţă
Agent
Temă
Receptor
Executant
Mutat
Spus
Privitor
Ascultător
Vorbitor
Şofer
Experienţiator
ŞoferAutobuz
ŞoferCamion
ŞoferTaxi


PROGRAMMING_CONCEPT
PROGRAMMING_ABSTRACTION
DATA_ABSTRACTION
MAPPING
ARRAY
CONTAINER
TABLE
HASHTABLE
INDEXTABLE
ARRAY
SYMBOLTABLE
COLLECTION
IMPLICITCOL
EXPLICITCOL
SET
SYMBOLTABLE
BAG
DISPENSER
STACK
QUEUE
HEAP
CURSORSTR
CONTROL_ABSTRACTION
EXCEPTION
IF_THEN

Webul semantic



RDF – Resource Description Framework

- Triplete
  - Resurse
  - Nume proprietăți
  - Valori proprietăți
- URI
- Literali

RDF

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
   ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://purl.org/rss/1.0/">
 <rdf:Description rdf:about="http://esp.ro/republica">
    <title>Republica</title>
    <dc:subject>Filosofie</dc:subject>
    <dc:creator>Platon</dc:creator>
    <dc:publisher>Ed.                St.                si
   Enciclopedica</dc:publisher>
    <dc:date>1986</dc:date>
    <dc:language>ro</dc:language>
    ...
  </rdf:Description>
</rdf:RDF>
```

Individuali

```
<Person rdf:ID="John"/>
<Person rdf:ID="Mary"/>
<rdf:Description rdf:about="#John">
  <hasParent:resource="#Mary"/>
  <age>25</ age>
</rdf:Description>
<rdf:Description rdf:about="#John">
  <differentIndividualFrom:resource="#Mary"/>
</rdf:Description>
<rdf:Description rdf:about="#Clinton">
  <sameIndividualAs:resource="#BillClinton"/>
</rdf:Description>
```

RDFS- clase (tipuri) pentru RDF

```
<rdfs:Class rdf:ID="Book"/>
<rdfs:Class rdf:ID="Man"/>
<rdfs:Class rdf:ID="Philosopher">
   <rdfs:subClassOf
   rdf:resource="#Man"/>
    ...
</rdfs:Class>
```

OWL –Ontology Web Language

```
<owl:Class rdf:ID="Man">
  <rdfs:subClassOf
  rdf:resource="#Person"/>
  <owl:disjointWith
  rdf:resource="#Woman"/>
</owl:Class>
```

OWL (cont.)

```
<owl:Class rdf:ID="EyesColor">
  <rdfs:subClassOf
  rdf:resource="#ManDescr"/>
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing rdf:about="#Blue"/>
    <owl:Thing rdf:about="#Green"/>
    <owl:Thing rdf:about="#Brown"/>
    <owl:Thing rdf:about="#Black"/>
  </owl:oneOf>
</owl:Class>
```

OWL constraint

```
<owl:ObjectProperty rdf:ID="author">
  <rdfs:domain rdf:resource="#Book" />
  <rdfs:range  rdf:resource="#Person" />
</owl:ObjectProperty>
```

OWL constraint

```
<owl:Restriction>
  <owl:onProperty
  rdf:resource="#author"/>
       <owl:minCardinality

  rdf:datatype="&xsd;nonNegativeInteger">
       1
       </owl:minCardinality>
</owl:Restriction>
```

Nivele OWL

- Lite
- DL
- Full

Logici descripționale

- KL/ONE
- CLASSIC
- KRYPTON
- LOOM
- FaCT
- Racer
- ...

Logici descripționale

- Concepte
- Roluri
- Axiome
- Individuali

Logici descripţionale

- Subsumare
- Verificare consistenţei
- Clasificare
- Tractabilitate vs. expresivitate

Definiții

- Extensionale
  - E(man)={John, Fred, Bob, Dan}
- Intensionale

Definiții intensionale

(**and** *concept1 concept2*)
(**or** *concept1 concept2*)
(**not** *concept*)
(**some** *proprietate1 concept1*)
(**all** *proprietate1 concept1*)
(**atleast** *numar1 proprietate1 concept1*)
(**atmost** *numar1 proprietate1 concept1*)

Descrieri extensionale

$$E[(:\text{and } c_1\ c_2)] = E(c_1) \cap E(c_2)$$

$E[(\text{all r v})]=\{d \in D \mid E[r(d)] \subseteq E(v)\}$

Concepte și roluri

```
(concept algorithm :primitive)
(concept                data-struct
  :primitive)
(disjoint   algorithm   data-
  struct)
(concept container (and data-
  struct))
(role input
  (:domain algorithm)
```

Concepte și roluri (cont.)

```
(concept container-algorithm
         (and algorithm (:the
  input container)))
```

Concepte și roluri (cont.)

```
(:the r c)
      echivalent cu:
   (and  (all  r  c)(atleast  1
  r)(atmost 1 r))

(some r c)
      echivalent cu:
(and (all r c)(atleast 1 r))
```

Servicii terminologice

- Subsumare
  - c1 subsumes c2 ⇔ E(c1) includes E(c2)
- Clasificare

Ontologii lexicale pe web

- WordNet (http://wordnet.stanford.edu)
- EuroWordNet
- BalkanNet
- FrameNet (http://framenet.icsi.berkeley.edu/)
- VerbNet

WordNet

- > 100,000 concepte
- Substantive, verbe, adjective, adverbe
- Din experimente psicholingvistice → o rețea semantică a conceptelor comune din limbaj

WordNet (cont.)

- 1 concept (e.g. share) → 1 synset → o mulțime de sinonime:
  share, portion, part, percentage
- 1 word → mai multe sensuri (synsets)

Sense 1 share -- (any of the equal portions into which the capital stock of a corporation is divided and ownership of which is evidenced by a stock certificate; "he bought 100 shares of IBM at the market price")

Sense 2 share, portion, part, percentage -- (assets belonging to or due to or contributed by an individual person or group; "he wanted his share in cash")

Sense 3 parcel, portion, share -- (the allotment of some amount by dividing something; "death gets more than its share of attention from theologians")

Sense 4 contribution, part, share -- (any one of a number of individual efforts in a common endeavor; "I am proud of my contribution to the team's success"; "they all did their share of the work")

Sense 5 plowshare, ploughshare, share -- (a sharp steel wedge that cuts loose the top layer of soil)

WordNet (cont.)

- Număr fix de relații:
- Hypernyms

share, portion, part, percentage -- (assets belonging to or due to or contributed by an individual person or group; "he wanted his share in cash")

=> assets -- (anything of material value or usefulness that is owned by a person or company)

=> possession -- (anything owned or possessed)

=> relation -- (an abstraction belonging to or characteristic of two entities or parts together)

=> abstraction -- (a general concept formed by extracting common features from specific examples)

=> abstract entity -- (an entity that exists only abstractly)

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

WordNet (cont.)

Hyponyms

share, portion, part, percentage

=> tranche -- (a portion of something (especially money))

=> dispensation -- (a share that has been dispensed or distributed)

=> dole -- (a share of money or food or clothing that has been charitably given)

=> way -- (a portion of something divided into shares; "the split the loot three ways")

=> ration -- (a fixed portion that is allotted (especially in times of scarcity))

=> allowance -- (an amount allowed or granted (as during a given period); "travel allowance"; "my weekly allowance of two eggs"; "a child's allowance should not be too generous")

=> slice, piece -- (a share of something; "a slice of the company's revenue")

=> split -- (a promised or claimed share of loot or money; "he demanded his split before they disbanded")

=> interest, stake -- ((law) a right or legal share of something; a financial involvement with something; "they have interests all over the world"; "a stake in the company's future")

=> profit sharing -- (a system in which employees receive a share of the net profits of the business)

=> cut -- (a share of the profits; "everyone got a cut of the earnings")

=> allotment, allocation -- (a share set aside for a specific purpose)

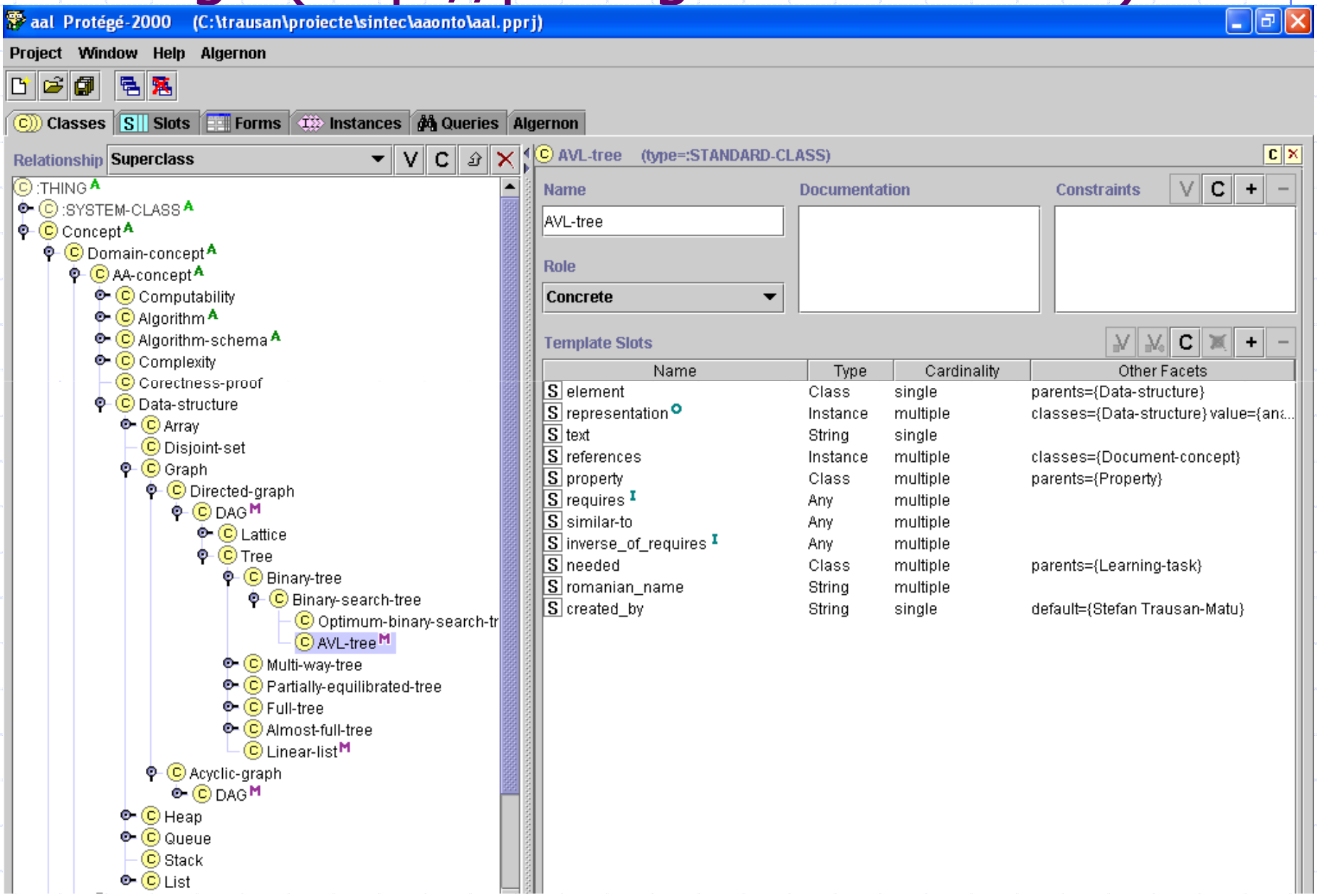
WordNet (cont.)

- Holonyms

share, portion, part, percentage -- (assets belonging to or due to or contributed by an individual person or group; "he wanted his share in cash")

PART OF: net income, net, net profit, lucre, profit, profits, earnings -- (the excess of revenues over outlays in a given period of time (including depreciation and other non-cash expenses))

Protégé (http://protege.stanford.edu)



Prelucrări semantice ale limbajului natural

- Vecinătate semantică
- Distanțe semantice în ontologii sau în rețele semantice
- Spații semantice – Latent Semantic Analysis (vezi lsa.colorado.edu)

Bibliografie

- Constandache, G.G., Ștefan Trăușan-Matu, Ontologia şi hermeneutica calculatoarelor, Ed. Tehnică, 2001
- Constandache, G.G., Ștefan Trăușan-Matu, M. Albu, C. Niculescu, Filosofie şi ştiinţe cognitive, MatrixRom, 2002
- Gruber, T., What is an Ontology, http://www.kr.org/top/definitions.html
- Ștefan Trăușan-Matu, Interfațarea evoluată om-calculator, Ed. MatrixRom, 2000
- http://www.w3.org/2004/OWL/
- http://wordnet.princeton.edu