Department of Biomedical Informatics, University at Buffalo
Grand Rounds September 9th, 2016

Knowledge acquisition and management for clinical decision-making
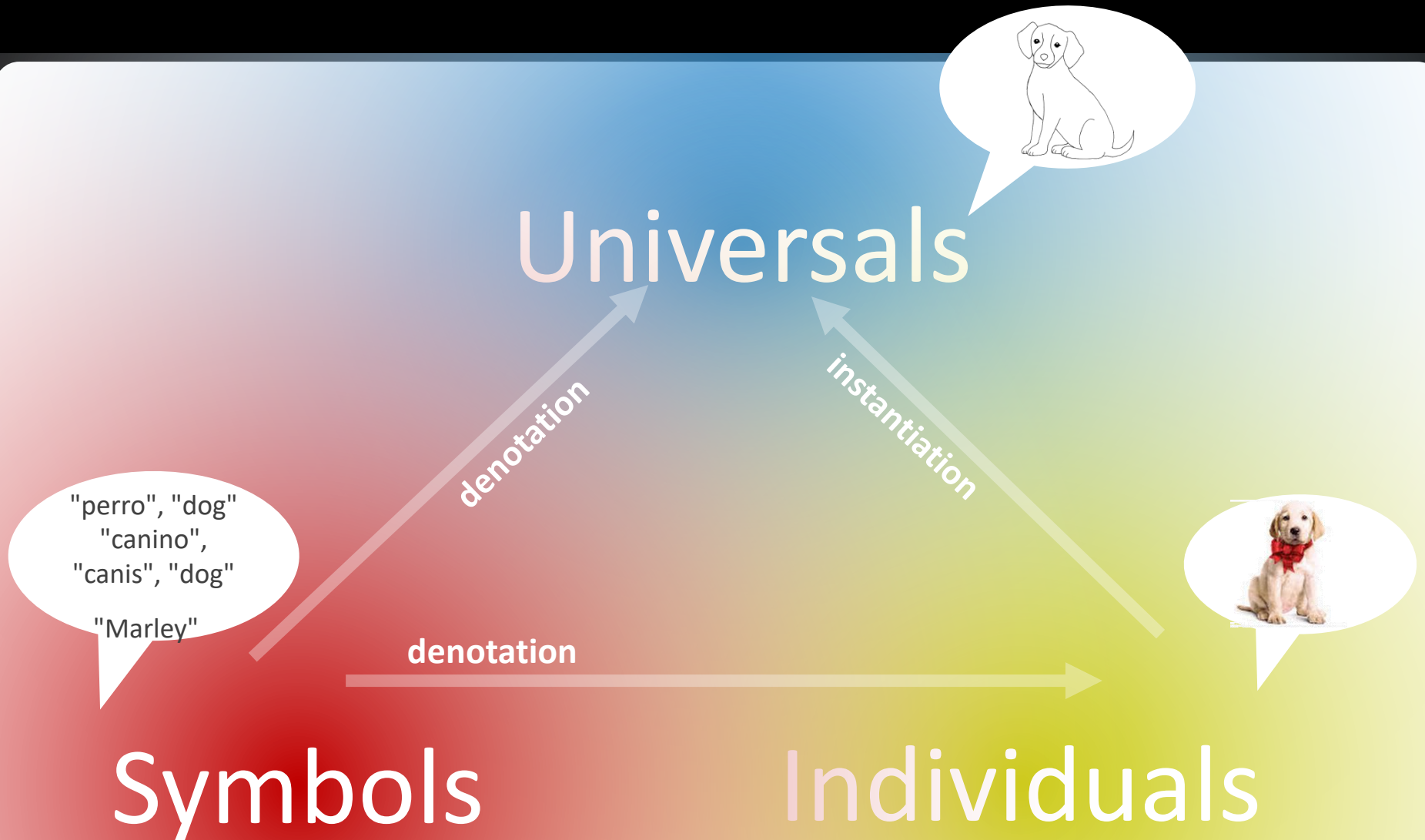
Stefan Schulz

Medical University of Graz (Austria)
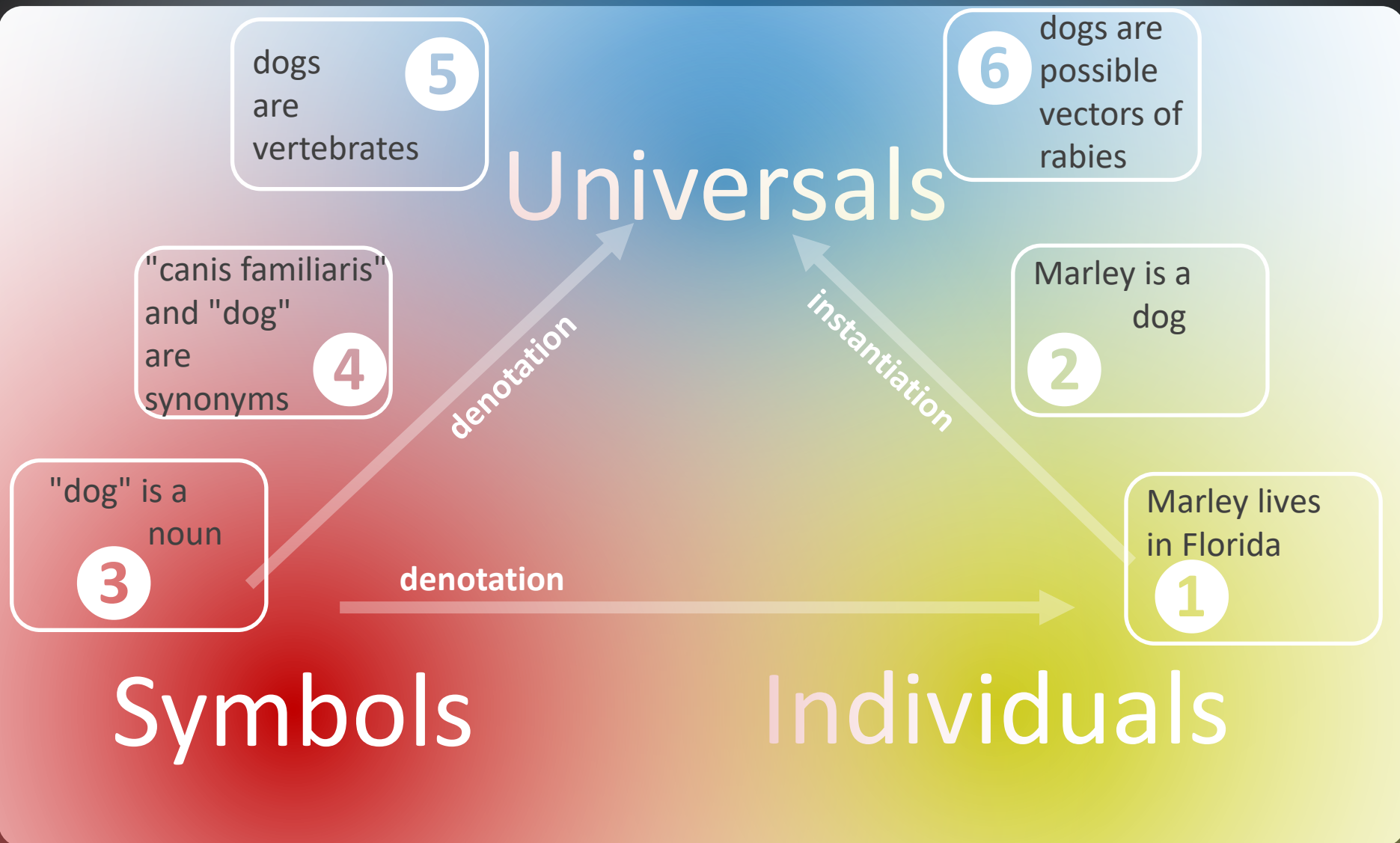
stefan.schulz@medunigraz.at

Topics of the talk

- Context: representation of knowledge in computer science

- What are the types of knowledge to be distinguished

- How are they connected

- How can data help acquire and maintain knowledge?
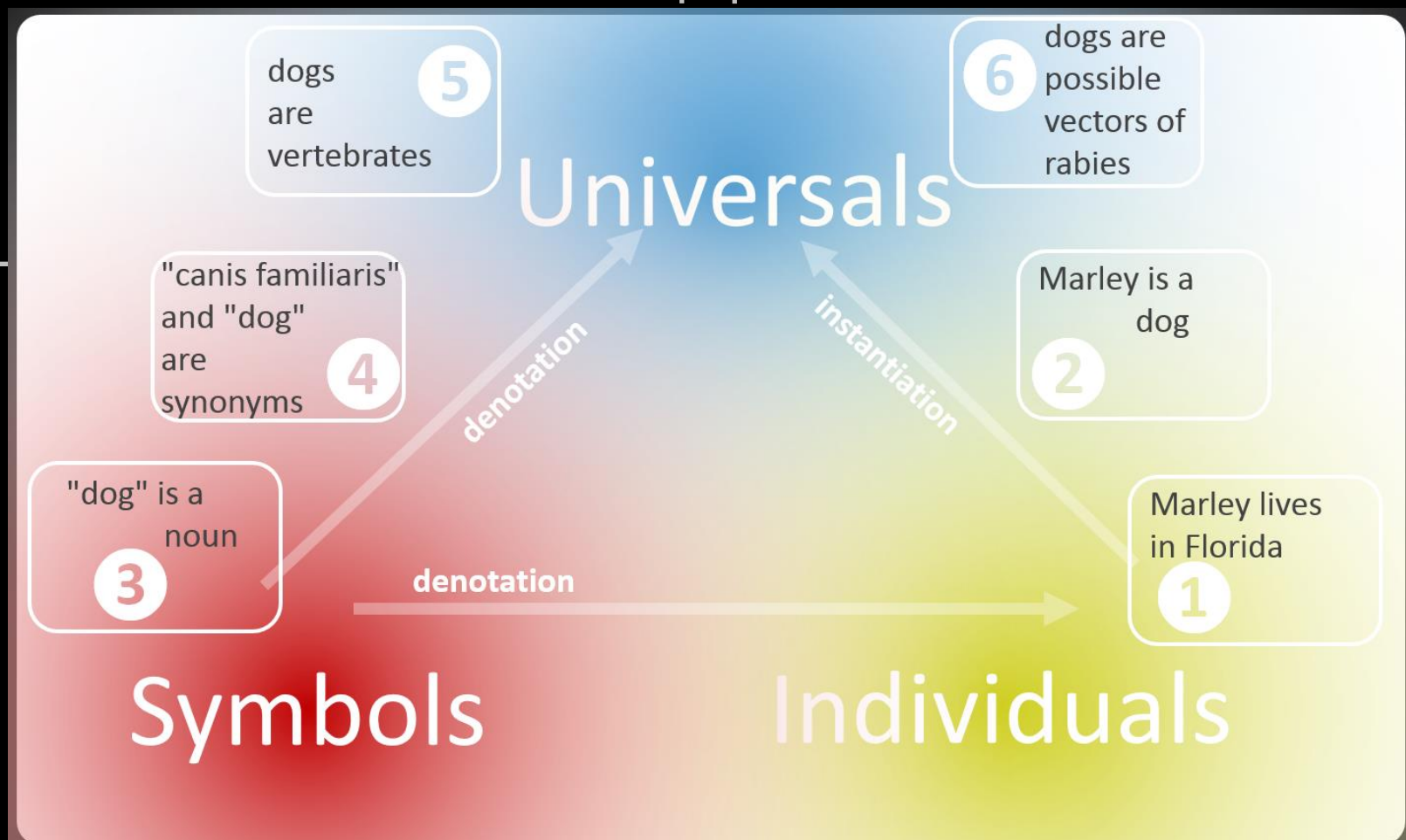
Knowledge map

C. K. Ogden and I. A. Richards (1923) The Meaning of Meaning

Knowledge map

Universals

Symbols

Individuals

dogs are vertebrates **5**

dogs are possible vectors of rabies **6**

"canis familiaris" and "dog" are synonyms **4**

Marley is a dog **2**

"dog" is a noun **3**

Marley lives in Florida **1**

denotation

instantiation

denotation

Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
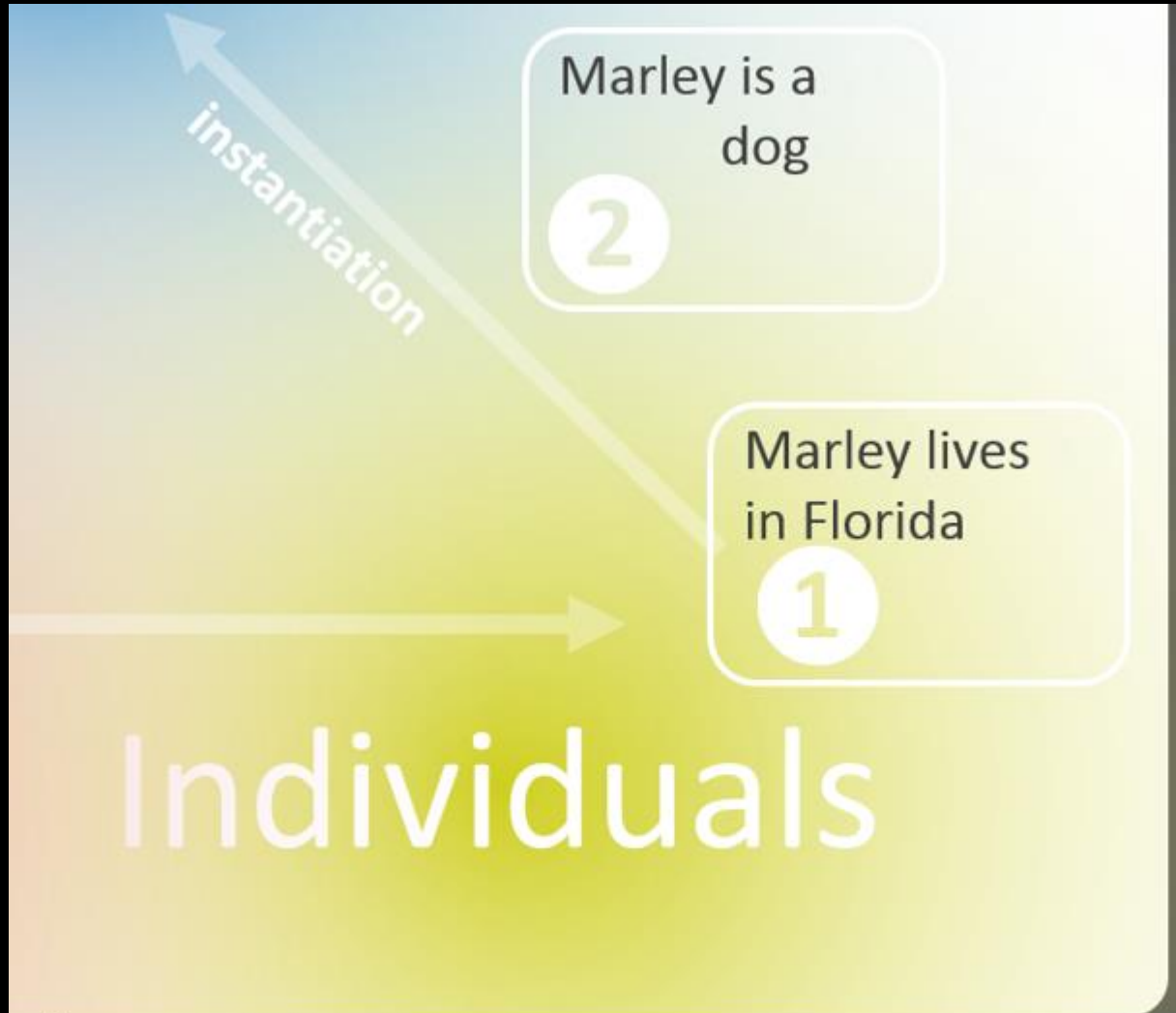typical, likely, possible

Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

dogs are vertebrates **5**

dogs are possible vectors of rabies **6**

"canis familiaris" and "dog" are synonyms **4**

Marley is a dog **2**

"dog" is a noun **3**

Marley lives in Florida **1**

Universals

Symbols

Individuals

denotation

instantiation

denotation

Factual knowledge: Statements about concrete entities and their relationships

Statements about individuals

\<Subject\> \<Predicate\> \<Object\>

```
:Florida rdf:type :state
:Marley rdf:type :dog
:Marley :lives :Florida
```

Syntax TURTLE : https://www.w3.org/TR/turtle/

Information extraction

Physical examination on admission revealed purpura of the upper and lower extremities, swelling of the gums and tonsils, but no symptoms showing the complication of myasthenia gravis. Hematological tests revealed leucocytosis: WBC count 68 700/μl (blasts 11.5%, myelocytes 0.5%, bands 2.0%, segments 16.0%, monocytes 65.5%, lymphocytes 4.0%, atypical lymphocytes 0.5%), Hb 7.1 g/dl (reticulocytes 12%) and a platelet count of 9.1 × 104/μl. A bone marrow aspiration revealed hypercelllar bone marrow with a decreased number of erythroblasts and megakaryocytes and an increased number of monoblasts

http://www.beverlyhillsneurology.com/mastheniagravis-symptoms.html

Instantiation

rdf:type

ADM{419620001
110714004
65124004 113279002
NEG{116223007 91637004 252275004
111583006} 767002[68700] 271040006[11.5]
313696224[0.5] 313696667[2.0]
313696009[16.0] 271037006[65.5]
271036002[4.0] 271036013[0.5]
365809007[7.1] 45995003[12]
365632008[91000] }
49401003 76197007 14016003

Myasthenia gravis (disorder) ☆ ↗

SCTID: 91637004

91637004 | Myasthenia gravis (disorder) |

Myasthenia gravis
Erb-Goldflam disease
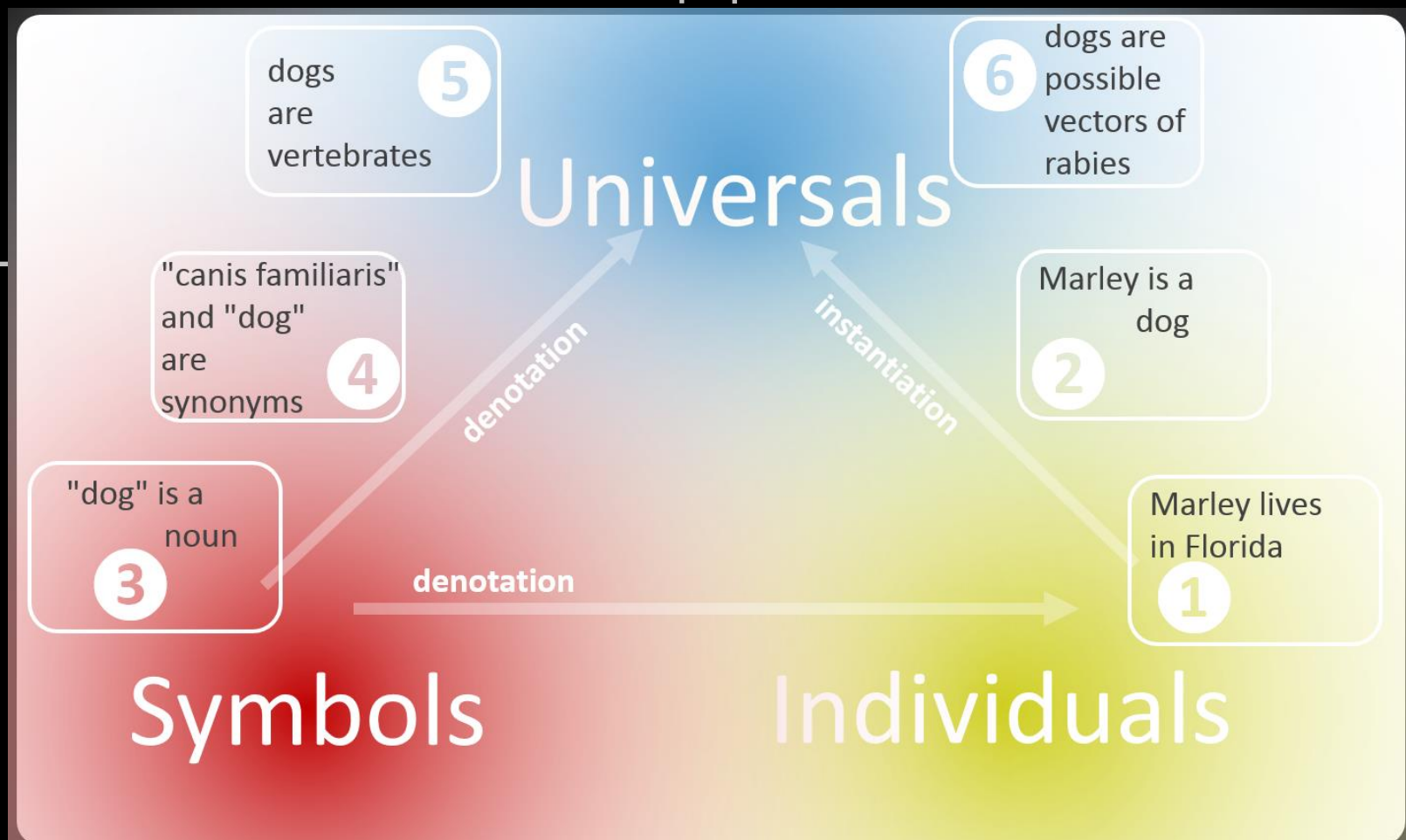Myasthenia gravis (disorder)
MG - Myasthenia gravis

Schulz S, Daumke P, Stenzhorn H, Markó K, Poprat M: Incremental Semantic Enrichment of Narrative Content in Electronic Health Records. World Congress on Medical Physics and Biomedical Engineering, 2009

Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible

Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

"canis familiaris" and "dog" are synonyms **4**

"dog" is a noun **3**

denotation

denotation

Symbols

Linguistic knowledge:
Statements about properties and meaning of signs of language

Representations SKOS / Linked Data

```
:ex:Dog rdf:type skos:Concept
:ex:Dog skos:prefLabel "dog"@en;
:ex:Dog skos:prefLabel "perro"@es;


:ex:Animal rdf:type skos:Concept
:ex:Animal skos:broader ex:Dog


wr:dog       lemon:sense      wr:dog-English-Noun-1
wr:dog       lemon:sense      wr:dog-English-Verb-1
wr:dog-English-Noun-1    wt:hasPoS wt:Noun
```

Syntax TURTLE : https://www.w3.org/TR/turtle/
Wiktionary: http://wiki.dbpedia.org/wiktionary-rdf-extraction

③ Linguistic knowledge: POS tagging
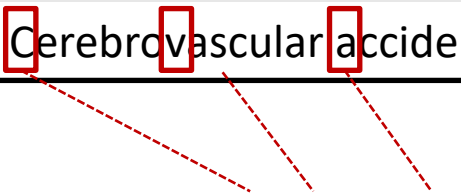
"Myasthenia" is a noun

JJ Physical **NN** examination **IN** on **NN** admission **VB**... of **DT** the **JJ** upper **CC** and **JJR** lower **NNS** extremities **VBG** swelling **IN** of **DT** the **NN** gums **CC** and **NNS** tonsils **,** , **CC** but **DT** no **NNS** symptoms **VBG** showing **DT** the **NN** complication **IN** of **NN** myasthenia **NN** gravis **.** . **JJ** Hemato logical **NNS** tests **VBD** revealed **NN** leucocytosis **:** : **NNP** WBC **VBP** count **CD** 68 **CD** 700 **NN** / **NN** μ **NN** l **-LRB-**
(**NNS** blasts **CD** 11.5 **NN** % **,** , **VBZ** myelocytes **CD** 0.5 **NN** % **,** , **NNS** bands **C D** 2.0 **NN** % **,** , **NNS** segments **CD** 16.0 **NN** % **,** , **VBZ** monocytes **CD** 65.5 **NN** % **,** , **VBZ** lymphocytes **CD** 4.0 **NN** % **,** , **NN** atypical **VBZ** lymphocytes **CD** 0.5 **NN** % **-RRB-**) **,** , **NNP** Hb **CD** 7.1 **NN** g **NN** / **NN** dl **-LRB-**
(**VBZ** reticulocytes **CD** 12 **NN** % **-RRB-**
) **CC** and **DT** a **NN** platelet **NN** count **IN** of **CD** 9.1 **NN** × **CD** 104 **NN** / **NN** μ **N N** l **.** . **DT** A **NN** bone **NN** marrow **NN** aspiration **VBD** revealed **JJ** hypercelllar **NN** bone **NN** marrow **IN** with **DT** a **VBN** decreased **NN** number **IN** of **NNS** eryt hroblasts **CC** and **NNS** megakaryocytes **CC** and **DT** an **VBN** increased **NN** nu mber **IN** of **NNS** monoblasts

- pattern to extract acronym definitions from a corpus

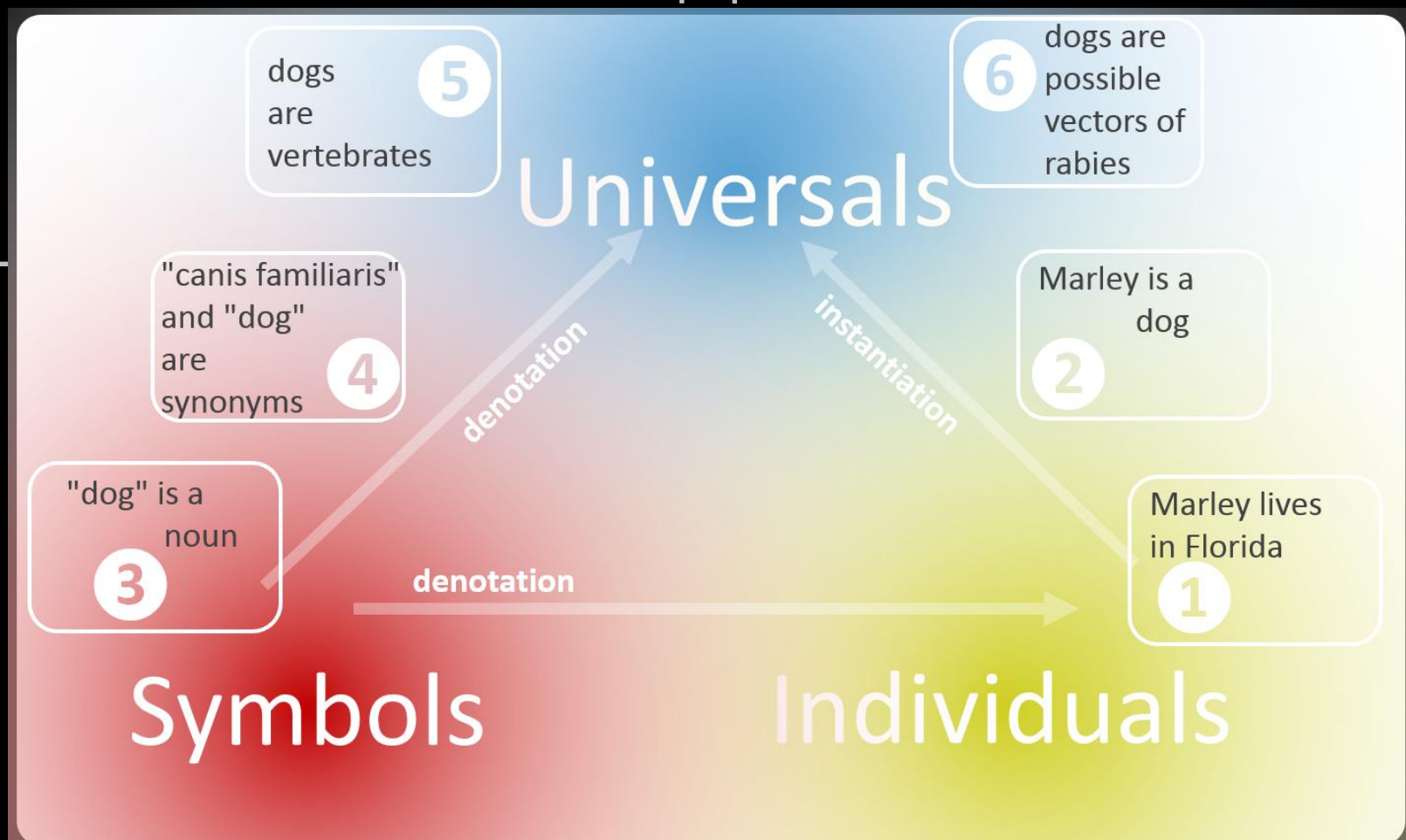| Pattern	Example
Acronym (Definition)	CVA (Cerebrovascular accident)
Definition (Acronym)	Cerebrovascular accident (CVA)
Acronym – Definition	CVA – Cerebrovascular accident–
Definition – Acronym	Cerebrovascular accident–ACV–

C < V < A

Sánchez D. & Isern D. Automatic extraction of acronym definitions from the Web. Appl Intell (2011) 34: 311–327

Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible

Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

Universals

Symbols

Individuals

5 dogs are vertebrates

6 dogs are possible vectors of rabies

4 "canis familiaris" and "dog" are synonyms

Marley is a dog 2

3 "dog" is a noun

Marley lives in Florida 1

denotation

instantiation

denotation

Ontological knowledge:

Axioms that are universally true

dogs
are
vertebrates

5

Universals

Representation OWL

`Dog subclassOf Vertebrate`

`Vertebrate subclassOf Animal`

`Vertebra subclassOf Bone`

`Vertebrate equivalentTo Animal and`
`            has-part some Bone`
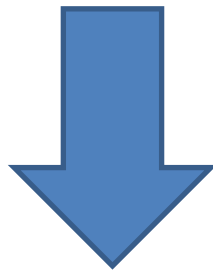
computable inference
(e.g. HermiT or Fact++
OWL reasoner)

There is no dog that has no bones

OWL Manchester Syntax: https://www.w3.org/TR/owl2-manchester-syntax/
HermiT reasoner: http://www.hermit-reasoner.com/
Fact++ reasoner: http://owl.man.ac.uk/factplusplus/

Extraction of taxonomic relations

- Lexical-semantic patterns Hearst:

Patrón	Ejemplo
NP such as {**NP**}* (and|or) **NP**	"AGE-binding proteins, such as GAL-3"
NP {**NP**}* (and|or) other **NP**	"fractures or other Injuries"
NP including {**NP**}* (or|and) **NP**	"mental illnesses including schizophrenia"

AGE-binding protein		injury		mental illness
subclass-of		subclass-of		subclass-of
GAL-1		fracture		schizophrenia

 Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In Proceedings of the Fourteenth International Conference on Computational Linguistics, pages 539--545, Nantes, France, July 1992.
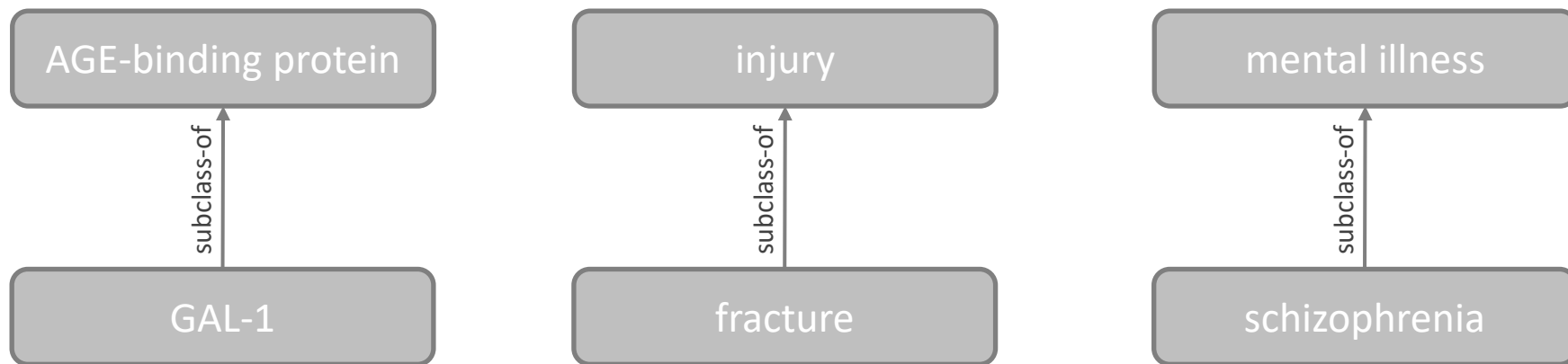
⑤ Extracción de otras relaciones

- Validation of SNOMED CT axioms by Web mining

- Query:
  "**Gastritis is an inflammation of**"

Current Concept:
Fully Specified Name: Gastritis (disorder)
ConceptId: 4556007

Defining Relationships:

Is a	Disorder of stomach (disorder)
Is a	Inflammation of specific body organs (disorder)
Is a	Inflammatory disorder of digestive tract (disorder)

Group 1
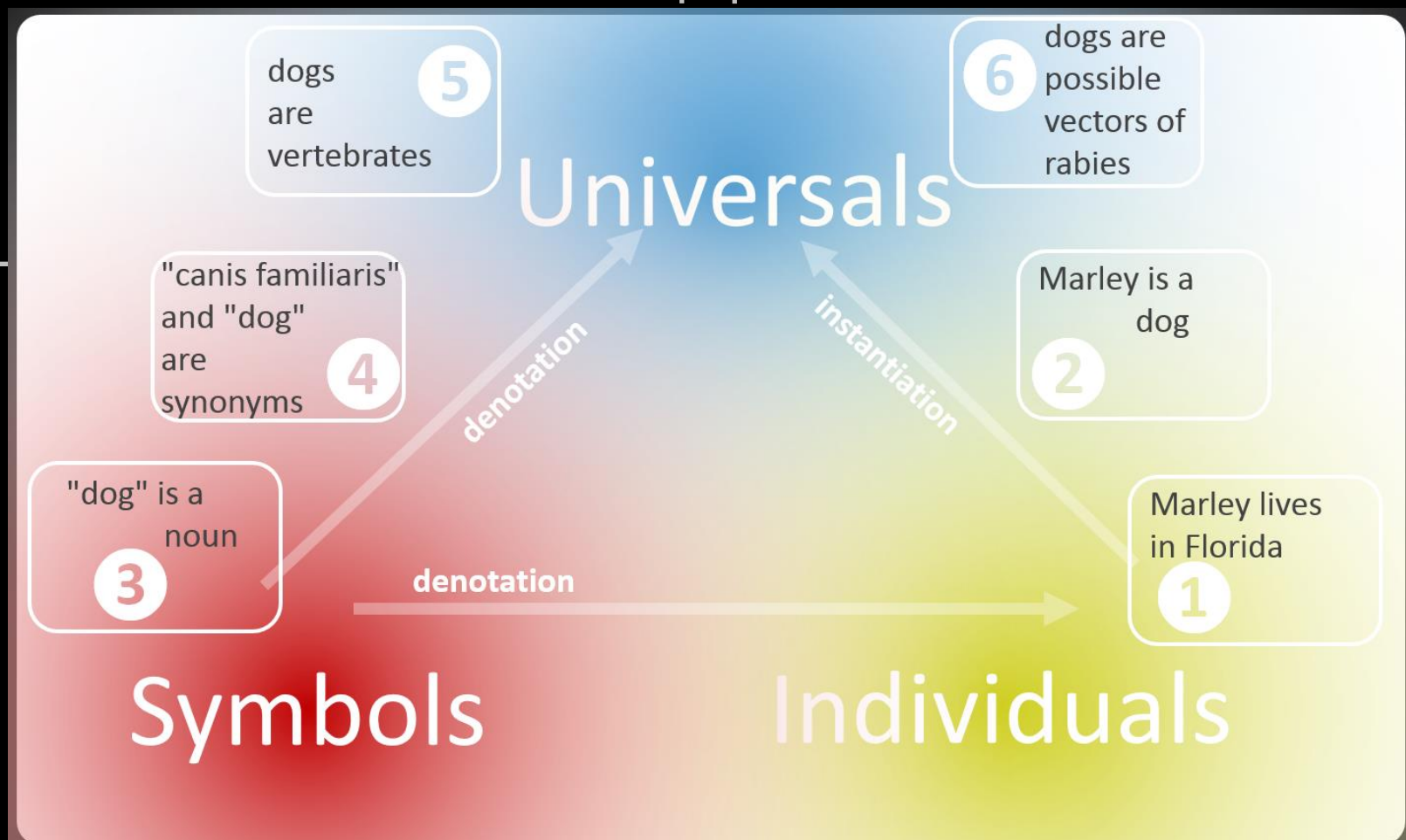|	
Associated morphology	Inflammation (morphologic abnormality)
Finding site	Stomach structure (body structure)

This concept is fully defined.

Result	Frequency
"stomach lining"	44
"lining of the stomach"	22
"lining of your stomach"	3
"lining of stomach"	1

Kreuzthaler M, Schulz S. Metonymies in medical terminologies. A SNOMED CT case study. AMIA Annu Symp Proc. 2012;2012:463-46.

Ontological knowledge:
Axioms that are universally true

Contingent knowledge:
typical, likely, possible

Linguistic knowledge:
Statements about properties and meaning of signs of language

Factual knowledge:
Statements about concrete entities and their relationships

Contingent knowledge

Alan Rector (2008):

" very few interesting items of knowledge that are truly ontological…". Much current work on informatics ontologies is aimed at integrating probabilistic and typical reasoning with universal "ontological" reasoning effectively. Hence, the background information for a clinical system often goes well beyond the ontology, in this strict sense. Brachman introduced the notion of the ontology as a "conceptual coat rack" on which other information is held."

Don't do this !

Dog subclassOf
  vector-of some Rabies

Tobacco subclassOf
  causes some Cancer

Aspirin subclassOf
  treats some Pain

Rabies subclassOf
  has-vector some Dog

Cancer subclassOf
  caused-by some Tobacco

Pain subclassOf
  treated-by some Aspirin

Triple representation

- No formal semantics!
- Different, mostly complex interpretations

```
<Subject>     <Predicate>     <Object>
:Dog          :vector-of      :Rabies
:Tobacco      :causes         :Cancer
:Aspirin      :treats         :Pain
:Fever        :suggests       :Malaria
:Bird         :capable-of     :Flying
```

- Co-occurrence analysis

Source concept	*Name*	Bipolar disorder
	Type	Disorder
Target concept	*Name*	Tricyclic antidepressant
	Type	Substance
MeSH subheadings		DT=9,CI=7,DI=5,PX=4,CO=2, EP=2,GE=2,BL=1,ET=1,PA=1, PC=1,PP=1,TH=1
Absolute co-occurrence		17
Log-likelihood		54.57

Qualifies source concept, e.g:
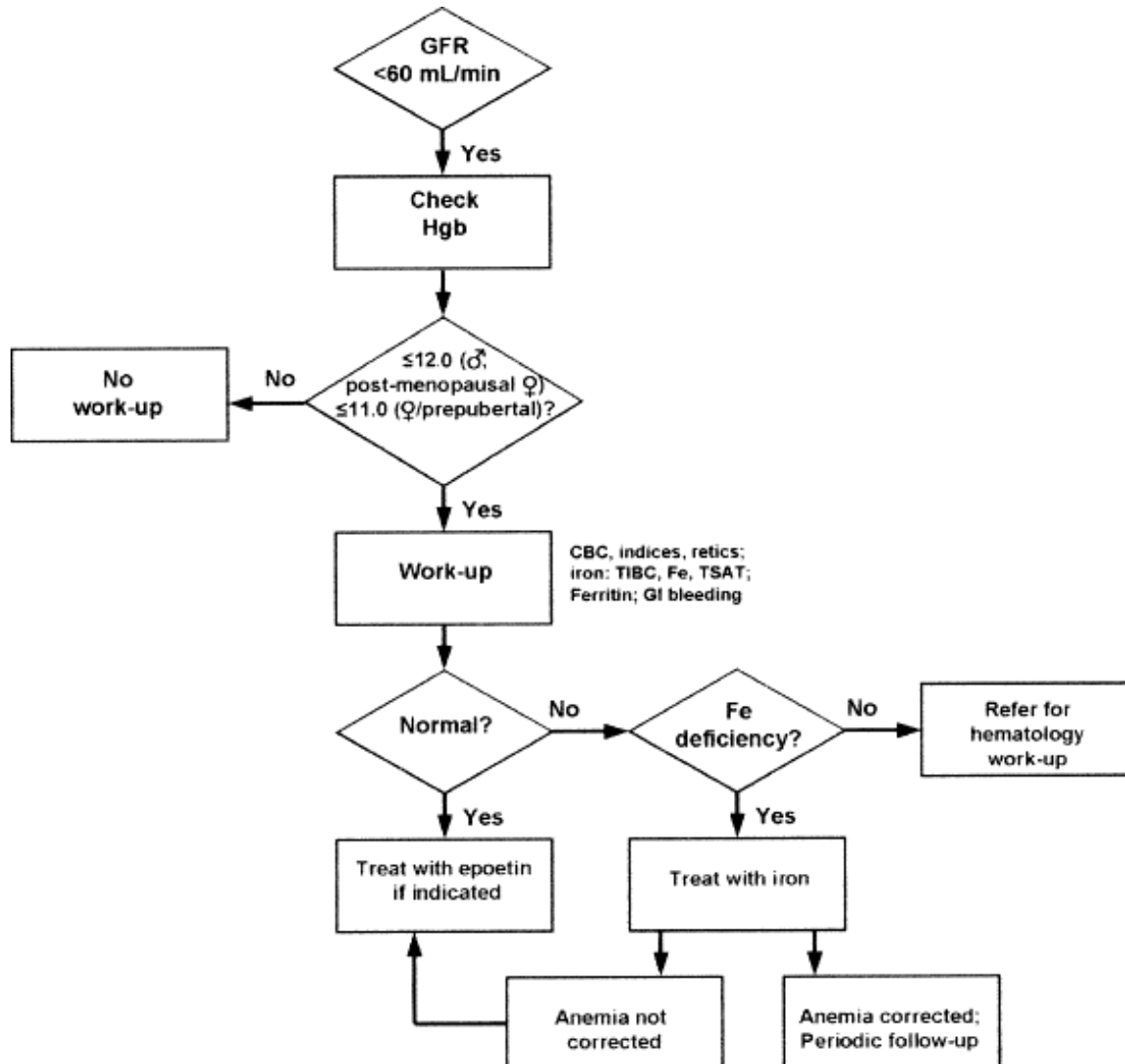
DT = "drug therapy"

PC = "prevention and control"

CO = "complication"

Predicate extractions according to MeSH subheadings analysis

	Object			
Subject	**Disease**	**Finding**	**Substance**	**Organism**
Finding	sign of symptom of	accompanied by	treated by	affects caused by
Substance	causes treats prevents metabolite of	causes treats prevents	Interacts with	affects produced by
Organism	causes affected by	causes	sensitive to	interacts with
Body part	possible location of	possible location of	targeted by	targeted by

Simplified: high frequency of " TU" ("therapeutic use") suggests the predicate "treats", high frequency of "PC" (prevention & control) suggests "prevents"

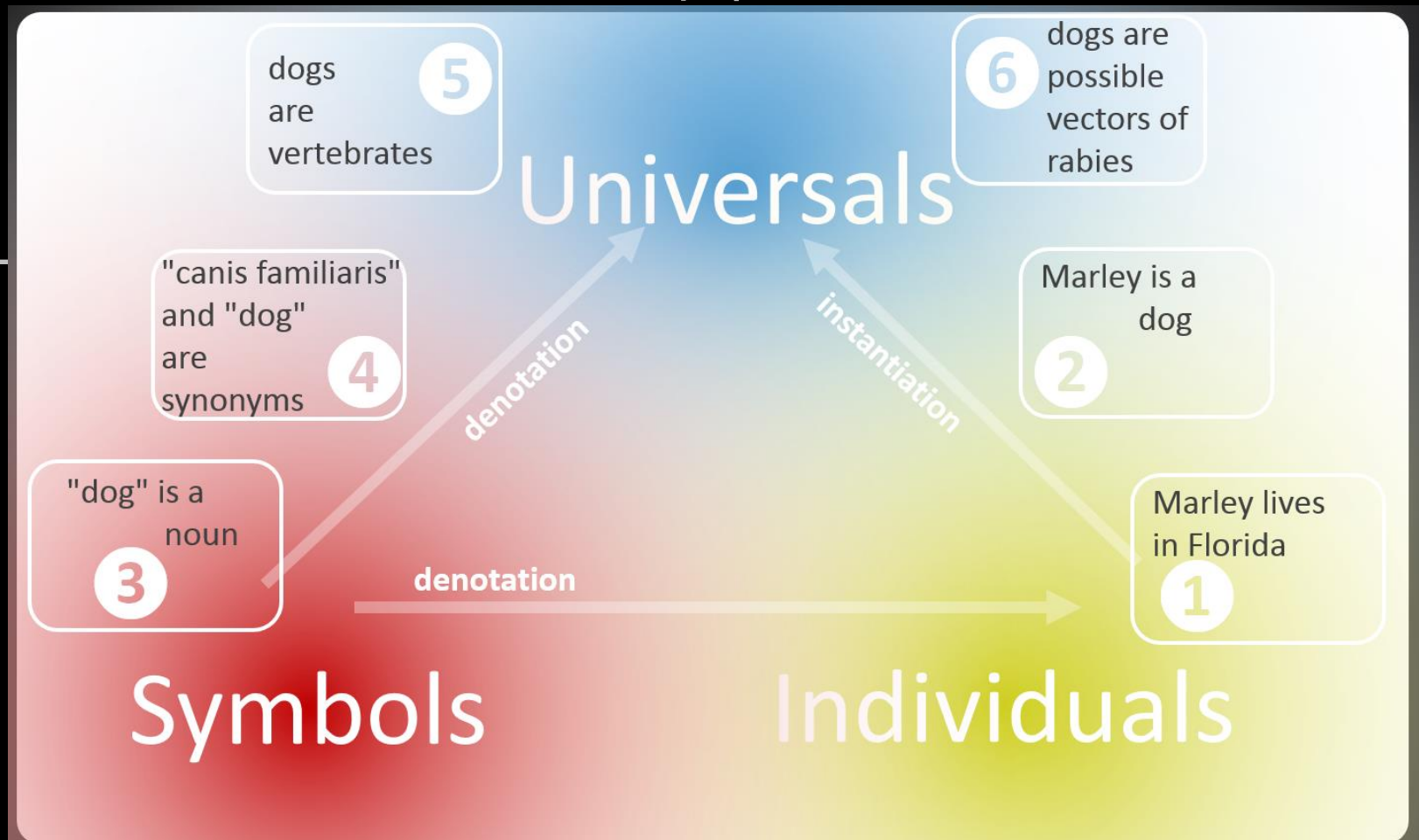Miñarro-Giménez JA, Kreuzthaler M, Schulz S. Knowledge Extraction from MEDLINE by Combining Clustering with Natural Language Processing. AMIA Annu Symp Proc. 2015 Nov 5;2015:915-24. eCollection 2015

Clinical practice guidelines

GFR <60 mL/min → Yes → Check Hgb →

≤12.0 (♂, post-menopausal ♀) ≤11.0 (♀/prepubertal)? → No → No work-up

Yes → Work-up — CBC, indices, retics; iron: TIBC, Fe, TSAT; Ferritin; GI bleeding

Normal? → No → Fe deficiency? → No → Refer for hematology work-up

Yes → Treat with epoetin if indicated

Fe deficiency? Yes → Treat with iron

Anemia not corrected

Anemia corrected; Periodic follow-up

Ontological knowledge: Axioms that are universally true

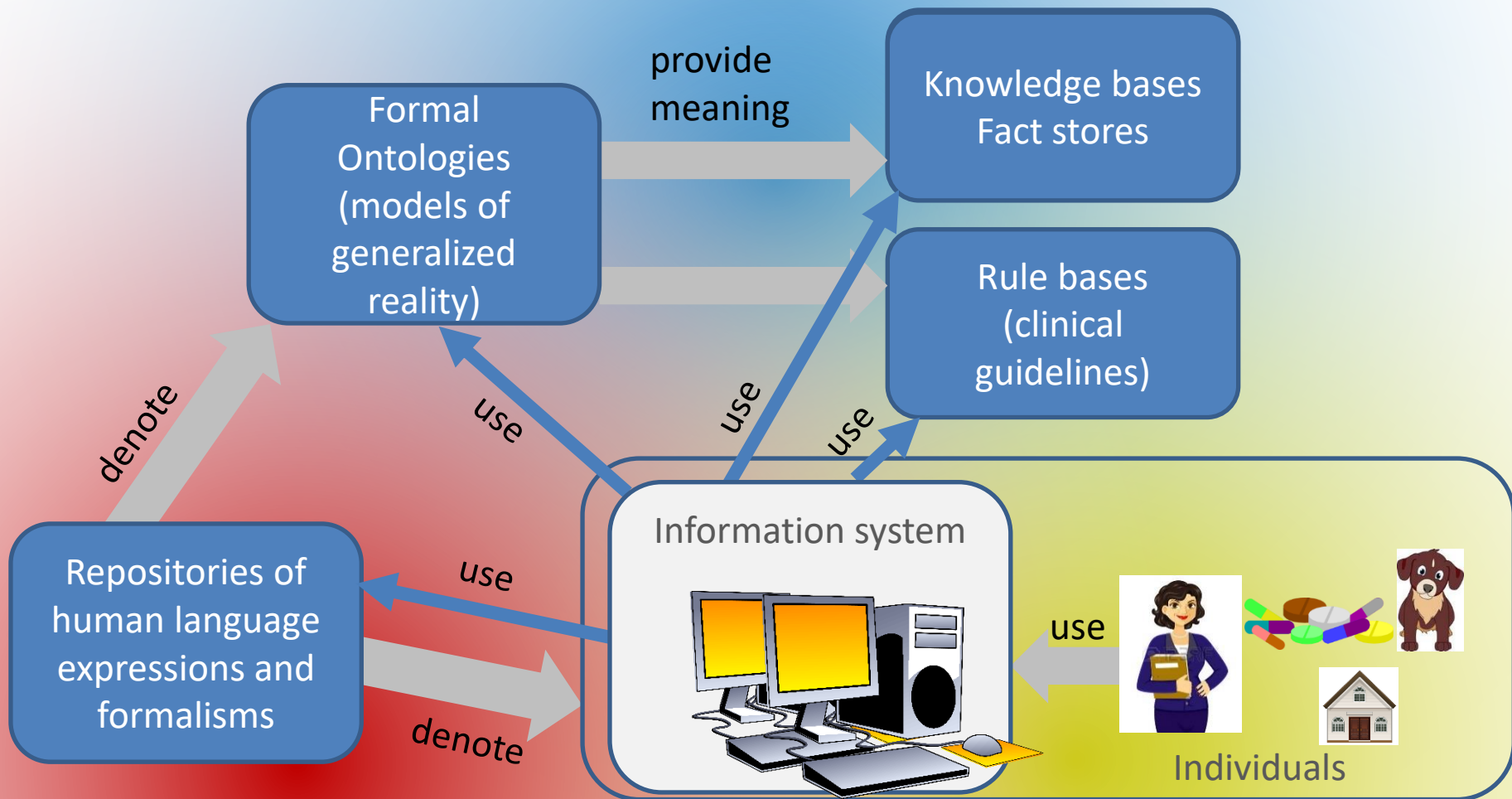**Contingent knowledge:** typical, likely, possible

Linguistic knowledge: Statements about properties and meaning of signs of language

Factual knowledge: Statements about concrete entities and their relationships

Knowledge assets, users and information systems

Perspectives

- Use big data to help construct and maintain large knowledge resources

- Importance of unstructured data in clinical data management
    - Used for knowledge construction
    - Used as input for decision-making

- Sensitize knowledge engineers to distinguish between qualitatively different knowledge assets

- Distinction between sustainable and permanent knowledge assets

Questions?

stefan.schulz@medunigraz.at