Universitas Saraviensis
Project Seminar „Text Mining for Historical Documents“
Antonia Scheidel • February 2009

An Introduction To Ontologies

- What are Ontologies?
- What do they look like?
- How can they be used?
- Why should we want to use them?

- Introduction to Ontologies
  - Philosophy and Information Science
  - Implementation of ontologies
  - Problems in ontology building
- Various uses of ontologies
- An automatically assembled ontology: YAGO
- Conclusion

Why we should be interested in Ontologies, Part One

Because they can help us solve some of the problems presented so far:

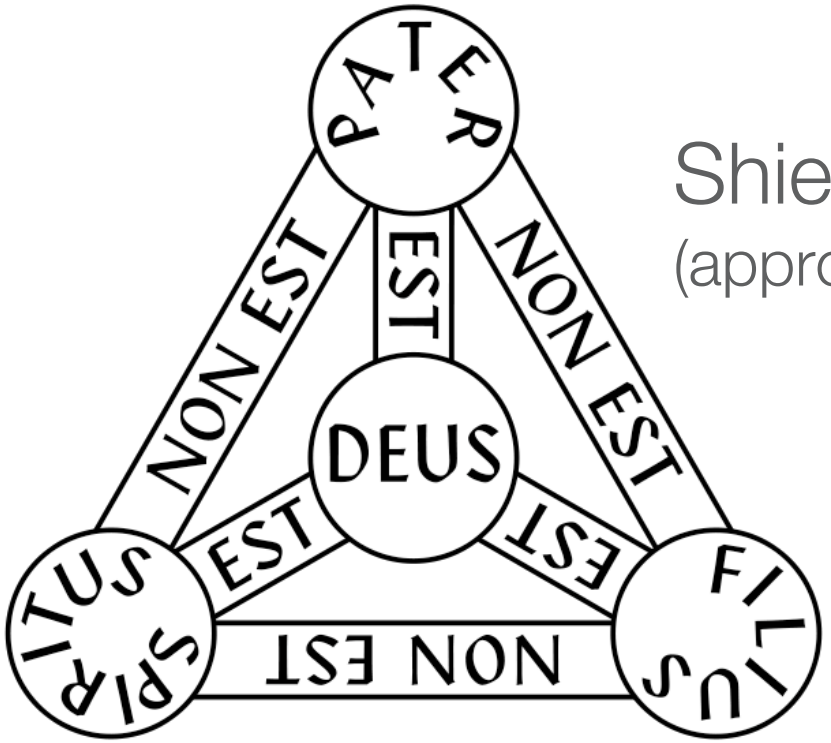
Ontologies can be used in Named Entity Disambiguation, are hugely important for the Semantic Web, etc...

Ontology in Philosophy

- „the science of what is“
- Among the first ontologists: Plato, Aristotle
- Questions: „What am I?“, „What is a physical object?“, „How do the properties of an object relate to the object itself?“
- *„…the theory of objects and their ties“*

Raul Corazzon, www.formalontology.it

Ontology in Philosophy - cont.



Shield of the Trinity
(approx. 12th century)

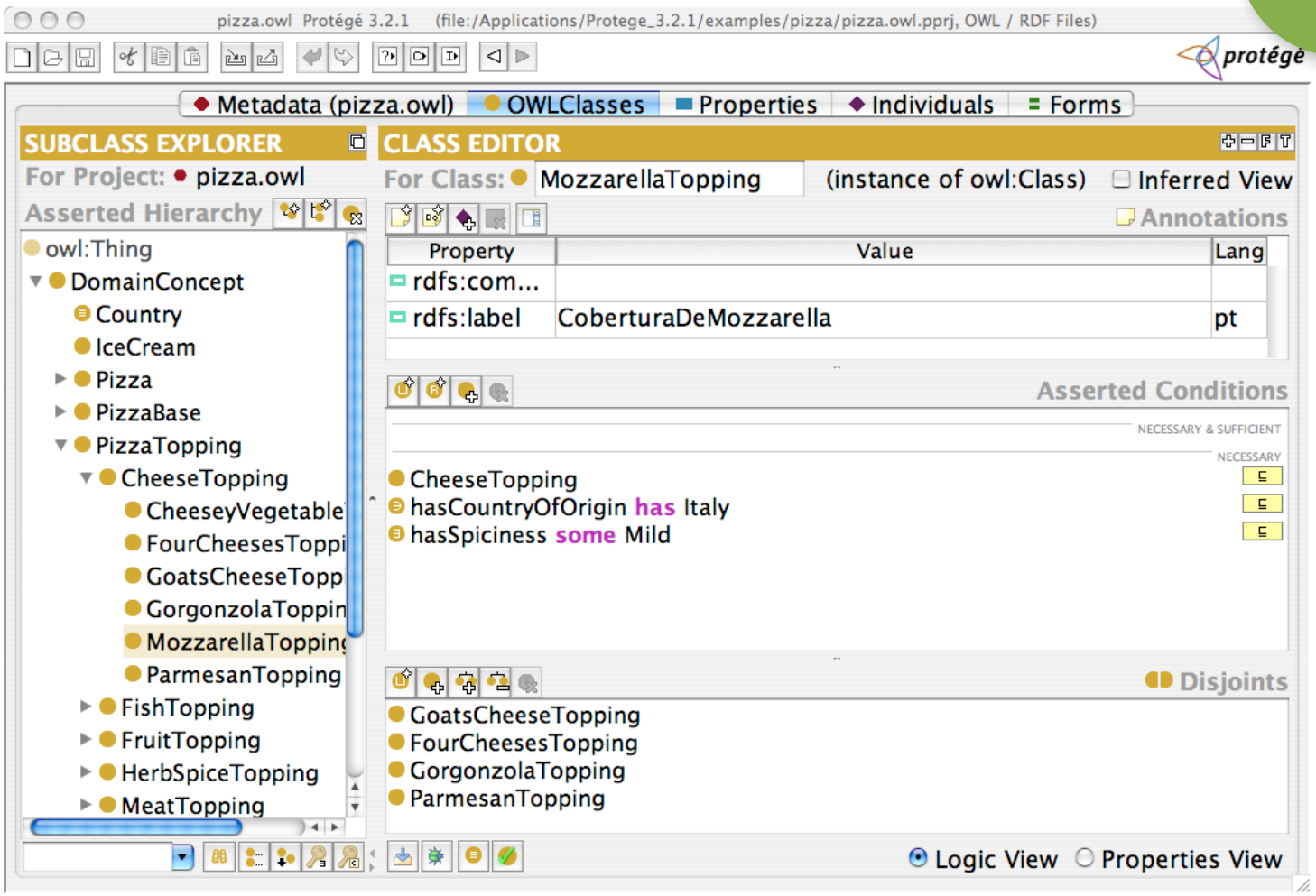
Ontologies in Information Science

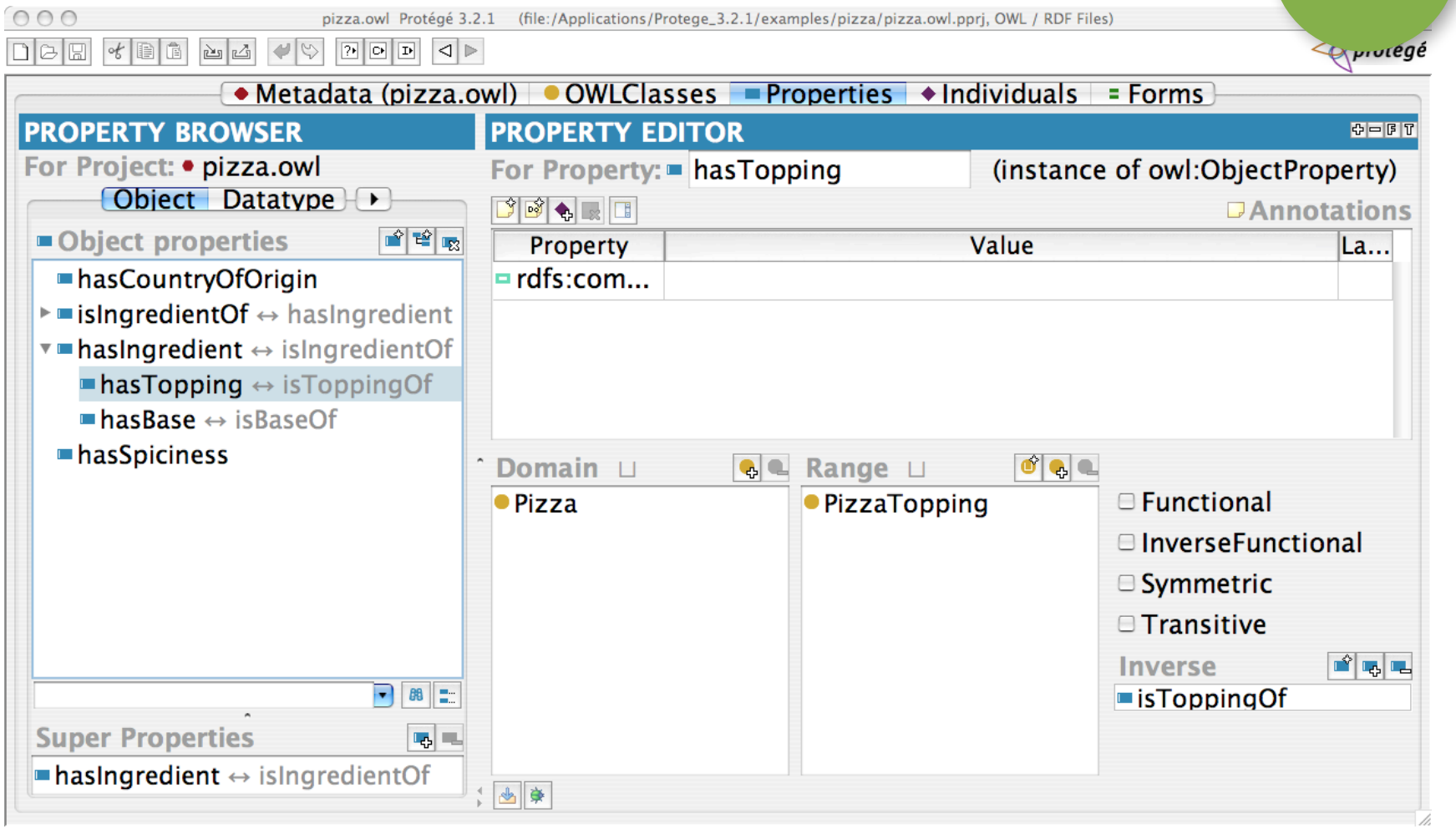
- Form of knowledge representation
- Represent knowledge about the world - or more specific domains
- Use concepts and relations to form triples:

- Possible: Inheritance relations between concepts (and relations)

A Simple Example: The Pizza Ontology

- Aim: Represent knowledge about Pizzas
- **Concepts (Classes):** *PizzaBase*, *PizzaTopping* (with subclasses *CheeseTopping*, *MeatTopping*)
- **Relations (Properties):** *hasIngredient* (with subproperties *hasTopping*, *hasBase*)


pizza.owl Protégé 3.2.1 (file:/Applications/Protege_3.2.1/examples/pizza/pizza.owl.pprj, OWL / RDF Files)
protégé
Metadata (pizza.owl)
OWLClasses
Properties
Individuals
Forms
SUBCLASS EXPLORER
For Project: pizza.owl
Asserted Hierarchy
owl:Thing
DomainConcept
Country
IceCream
Pizza
PizzaBase
PizzaTopping
CheeseTopping
CheeseyVegetable
FourCheesesToppi
GoatsCheeseTopp
GorgonzolaToppin
MozzarellaTopping
ParmesanTopping
FishTopping
FruitTopping
HerbSpiceTopping
MeatTopping
CLASS EDITOR
For Class: MozzarellaTopping
(instance of owl:Class)
Inferred View
Annotations
Property
Value
Lang
rdfs:com...
rdfs:label
CoberturaDeMozzarella
pt
Asserted Conditions
NECESSARY & SUFFICIENT
NECESSARY
CheeseTopping
hasCountryOfOrigin has Italy
hasSpiciness some Mild
Disjoints
GoatsCheeseTopping
FourCheesesTopping
GorgonzolaTopping
ParmesanTopping
Logic View
Properties View


pizza.owl Protégé 3.2.1 (file:/Applications/Protege_3.2.1/examples/pizza/pizza.owl.pprj, OWL / RDF Files)
Metadata (pizza.owl)
OWLClasses
Properties
Individuals
Forms
PROPERTY BROWSER
For Project: pizza.owl
Object
Datatype
Object properties
hasCountryOfOrigin
isIngredientOf ↔ hasIngredient
hasIngredient ↔ isIngredientOf
hasTopping ↔ isToppingOf
hasBase ↔ isBaseOf
hasSpiciness
Super Properties
hasIngredient ↔ isIngredientOf
PROPERTY EDITOR
For Property: hasTopping
(instance of owl:ObjectProperty)
Annotations
Property
Value
La...
rdfs:com...
Domain
Pizza
Range
PizzaTopping
Functional
InverseFunctional
Symmetric
Transitive
Inverse
isToppingOf

The Pizza Ontology: Facts & Figures

- 98 classes, 8 properties, 5 **individuals** (countries of origin)
- Not complete (by far!), but easily expandable (more properties...!)
- Try to imagine a food ontology!

Ontology Building: Problems

„When I think of an ontology, I think of putting the universe in a bottle. It's a very ambitious thing to do. **If you have a proper ontology worked out, it means you know everything about everything.**
In general, the more useful an ontology would be, the closer to impossible it's going to be to make it.“

Found on the discussion boards on www.metafilter.com

Ontology Building: Problems - cont.

➔ Usefulness of ontologies depends strongly on the domains they are used in!

- Usual practice: Create reliable ontologies for small domains or domains that already make use of taxonomies (biology, medicine) rather than try to build a universal ontology.

Intermission

- ✔ What are Ontologies?
- ✔ What do they look like?
- ? How can they be used? ← now!
- 😐 Why should we want to use them?

A very popular application: WordNet

- A large* ontology for the English language
- Hierarchically organizes nouns, verbs, adjectives, and adverbs into synonym sets
- Also offers: antonyms, meronyms, etc. ...

* (147278 unique strings, 117659 synonym sets)

WordNet: Example

Noun

- S: (n) **historian**, historiographer (a person who is an authority on history and who studies it and writes about it)
  - *direct hyponym* / *full hyponym*
  - *domain category*
    - S: (n) history (the discipline that records and interprets past events involving human beings) *"he teaches Medieval history"; "history takes the long view"*
  - *has instance*
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) scholar, scholarly person, bookman, student (a learned person (especially in the humanities); someone who by long study has gained mastery in one or more disciplines)
      - S: (n) intellectual, intellect (a person who uses the mind creatively)
        - S: (n) person, individual, someone, somebody, mortal, soul (a human being) *"there was too much for one person to do"*
          - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
            - S: (n) living thing, animate thing (a living (or once living) entity)
              - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                  - S: (n) physical entity (an entity that has physical existence)
                    - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
          - S: (n) causal agent, cause, causal agency (any entity that produces an effect or is responsible for events or results)
            - S: (n) physical entity (an entity that has physical existence)
              - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
  - *derivationally related form*

Why we should be interested in Ontologies, Part Two

- Philosophical / epistemological interest
- Because they are a fairly intuitive, appealing way of storing and organizing knowledge

Uses of Ontologies

- Semantic Web: Take meaning into account
- Store world knowledge ➔ use to automatically draw conclusions based on logical rules
- Question Answering systems: Enable more complex / natural language queries

What do we want to use them for?

- We want to be able to look for things like:
  - *people who are scientists and musicians*
  - *people affiliated with Catharism outside the Languedoc region*
  - *when did Elvis win the Grammy Award?*

➔ a queryable semantic knowledge base!

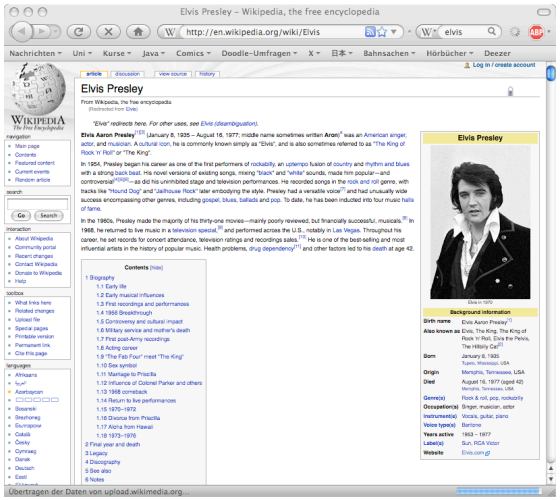
YAGO

- „Yet Another Great Ontology“
- Uses data from Wikipedia and WordNet
- Automatically assembled knowledge base
- Contains 2 million entities and 20 million facts about these entities

Where Wikipedia comes in

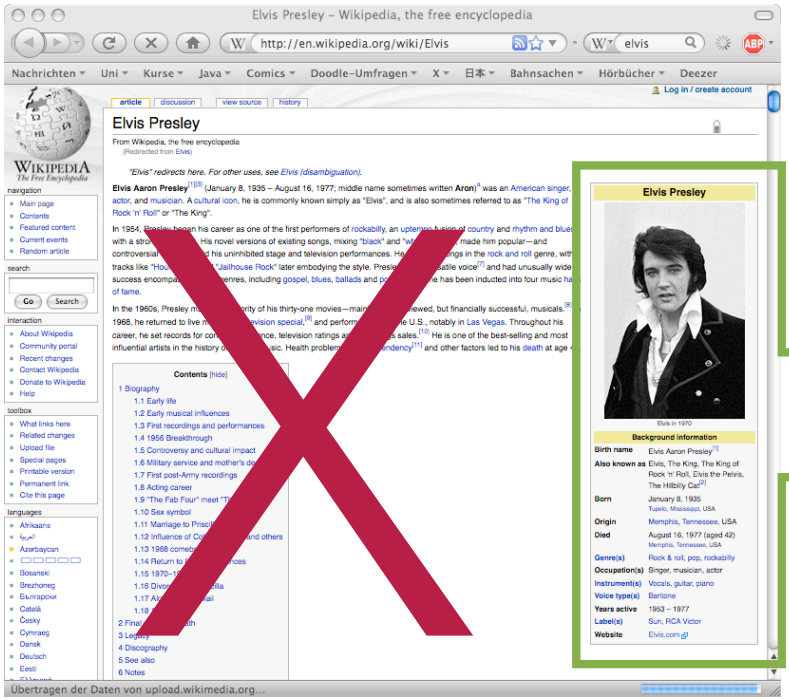
Or: How does the ontology learn who Elvis is?

➔ Answer: Just like everyone else!



?

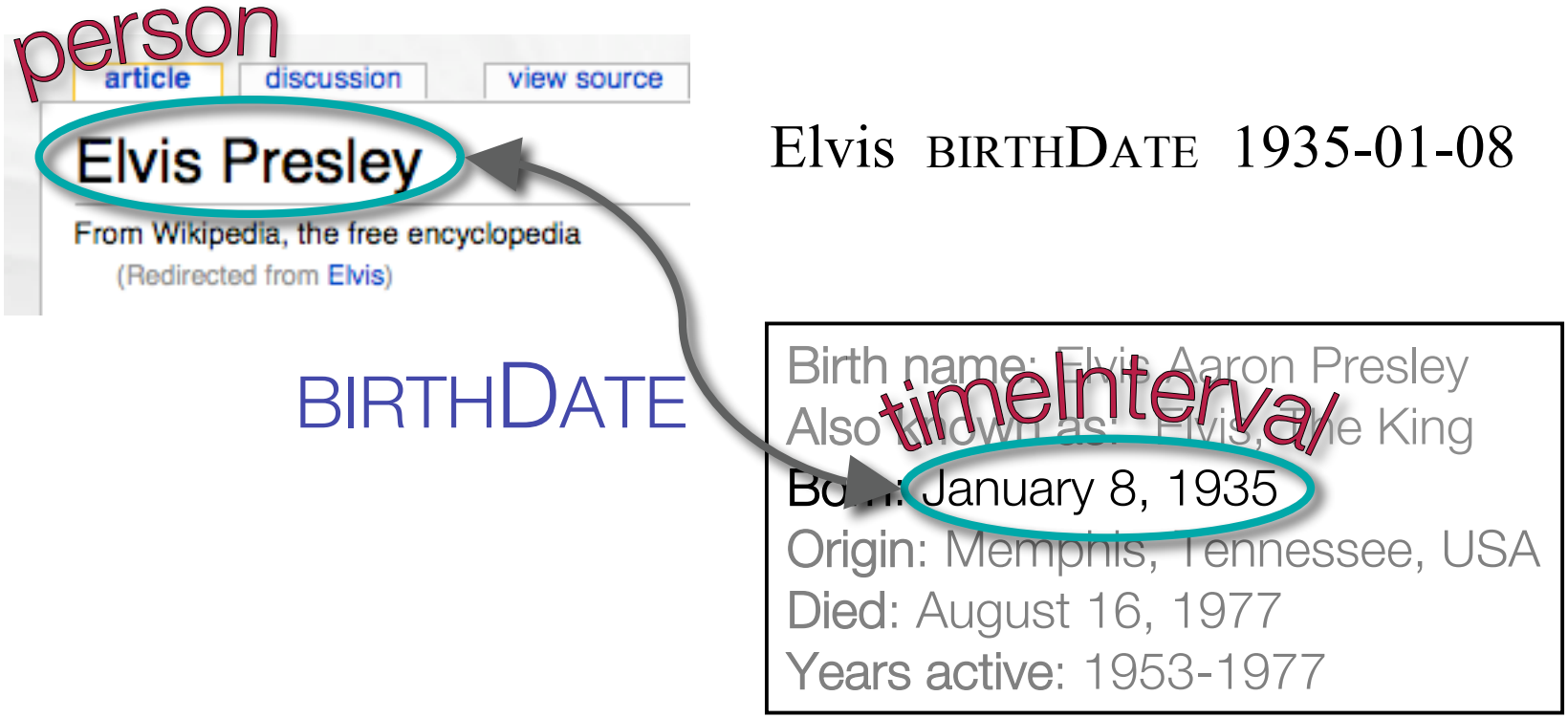
...and what comes next?



InfoBoxes!

Birth name: Elvis Aaron Presley
Also known as: Elvis, The King
Born: January 8, 1935
Origin: Memphis, Tennessee, USA
Died: August 16, 1977
Years active: 1953-1977

Create triples



Elvis BIRTHDATE 1935-01-08

Where Mozzarella comes back

=

?

Wikipedia Categories

v · d · e	**Elvis Presley**	[show]

Categories: 1935 births | 1977 deaths | Deaths from myocardial infarction | Elvis Presley | Actors portrayed posthumously | American actor-singers | American baritones | English-language singers | American country singers | American crooners | American expatriates in Germany | American film actors | American gospel singers | American male singers | American Pentecostals | American rock singers | English Americans | German-Americans | Americans of Native American descent | Scottish-Americans | Scottish-American musicians | Blues musicians from Mississippi | Burials in Tennessee | Gospel Music Hall of Fame inductees | Grammy Award winners | Grammy Lifetime Achievement Award winners | Country Music Hall of Fame inductees | Identical twins | Las Vegas musicians | Mississippi Blues Trail | People from Las Vegas, Nevada | People from Memphis, Tennessee | People from Tupelo, Mississippi | Performers of Christian music | Sun Records artists | RCA Victor Records artists | Rock and Roll Hall of Fame inductees | Rockabilly Hall of Fame inductees | Southern gospel performers | Actors from Tennessee | Musicians from Tennessee | UK Music Hall of Fame inductees | United States Army soldiers

Wikipedia Categories - cont.

Example: Musicians from Tennessee

- is in categories: *American musicians by state, People from Tennessee by occupation, Tennessee culture*
- has subcategories: *Nashville bands, Rappers from Memphis*

➔ useful??

Where WordNet comes in

- Remember: WordNet already has a very good taxonomy!
- Take advantage of that: Integrate Synsets!
- How: Each Synset becomes a class of YAGO
- Fit Wikipedia categories into WordNet classes

- S: (n) **musician**, instrumentalist, player (someone who plays a musical instrument (as a profession))
- S: (n) **musician** (artist who composes or conducts music as a profession)
  - *direct hyponym* / *full hyponym*
    - S: (n) arranger, adapter, transcriber (a musician who adapts a composition for particular voices or instruments or for another style of performance)
    - S: (n) choirmaster, precentor, cantor (the musical director of a choir)
    - S: (n) composer (someone who composes music as a profession)
    - S: (n) conductor, music director, director (the person who leads a musical group)
    - S: (n) virtuoso (a musician who is a consummate master of technique and artistry)
    - S: (n) Herbert, Victor Herbert (United States musician and composer and conductor noted for his comic operas (1859-1924))
    - S: (n) Ono, Yoko Ono (United States musician (born in Japan) who married John Lennon and collaborated with him on recordings (born in 1933))
    - S: (n) Orff, Carl Orff (German musician who developed a widely used system for teaching music to children (1895-1982))
  - *domain category*
  - *has instance*
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) artist, creative person (a person whose creative work shows sensitivity and imagination)
      - S: (n) creator (a person who grows or makes or invents things)
        - S: (n) person, individual, someone, somebody, mortal, soul (a human being) "*there was too much for one person to do*"
          - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
            - S: (n) living thing, animate thing (a living (or once living) entity)
              - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) "*how big is that part compared to the whole?*"; "*the team is a unit*"
                - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) "*it was full of rackets, balls and other objects*"
                  - S: (n) physical entity (an entity that has physical existence)
                    - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Example - cont.

- Wiki category: Musicians from Tennessee
- „Musicians from Tennessee“ can be regarded as a hyponym of „musicians“
- „musicians“ is in WordNet class *musician*: has hypernyms *artist* ➔ *creator* ➔ *person* ➔ *organism*
- Musicians from Tennessee becomes a subclass of *musician*

Intermission

What have we seen so far:

- The transformation of Elvis Presley into a YAGO entity
- The creation of one „entity-relation-entity“ triple

Elvis	BIRTHDATE	1935-01-08

- The integration of WordNet classes into the ontology

What else is there?

- Apart from Elvis Presley: 2,733,684 Wikipedia articles - just in English! (as of Feb. 9th, 2009)
- Every Wikipedia page title (i.e. the subject of every article) is a candidate to become an individual in YAGO!

One more relation

Apparently, „Elvis“ usually refers to the person Elvis Presley!

Let‘s use this as a relation:

„Elvis“ MEANS *Elvis Presley*

How can we use MEANS?

- Express (and store in the ontology) that various different names may refer to the same concept - more freedom for queries!

„King Of Rock & Roll“ MEANS *Elvis Presley*

„King Of Pop“ MEANS *Michael Jackson*

Did you say Queries?

Query Form

YAGO-query:

?id0:	?x	isA	scientist
?id1:	?x	isA	musician
?id2:			

Daten absenden

?musician = musician
?scientist = scientist
?x = Alexander_Borodin

?musician = musician
?scientist = scientist
?x = Neil_J._Gunther

?musician = musician
?scientist = scientist
?x = Albert_Schweitzer

?musician = musician
?scientist = scientist
?x = William_Herschel

people who are scientists and musicians

Summary

- ✔ What are Ontologies?
- ✔ What do they look like?
- ✔ How can they be used?
- ✔ Why should we want to use them?

Thank you for your attention!

Sources

- F. M. Suchanek, G. Kasneci and G. Weikum. "Yago - A Large Ontology from Wikipedia and WordNet" Elsevier Journal of Web Semantics. 2008.
- http://www.philolex.de/ontologi.htm
- http://www.formalontology.it/
- http://en.wikipedia.org/wiki/Ontology
- http://de.wikipedia.org/wiki/Ontologie
- http://wordnet.princeton.edu/