Tag the Tag!

Bottom-up collaborative ontology building using two-dimensional tagging

Maarten van Gompel, Steven Langerwerf, Wouter Bouvy

April 6, 2009

Ontologies

Model

Subject - Relation - Object

Properties

- Semantic expressiveness
- High quality
- Top-down
- Formal
- Rigid
- Difficult to construct

Folksonomies

Model

User - Tag - Instance

Properties

- Flexible
- Easy to construct
- Collaborative
- Bottom-up
- Informal
- No semantics
- No quality guarantees

Our approach: Two-dimensional Tagging

Model

Subject - **Relation** - Object - User

Properties

- Flexible
- Easy to construct
- Informal
- Collaborative
- Semantic expressiveness
- Bottom-up

Research Question

Finding the middle ground: Can we build another type of folksonomy/ontology using two-dimensional tagging?

1. more semantic expressiveness than 'flat' folksonomies
2. easy to build in bottom-up collaborative fashion

Objective:

1. What can emerge?
2. Is this a viable approach for building ontologies?

Related Work

Extreme Tagging

(Tanasescu and Streibel (2007))

- Similar idea
- Tags and resources are interchangeable
- Implicit tagging versus Explicit tagging
- Not really evaluated (only 5 test-subjects)

ConceptNet and Open Mind Common Sense

(Liu, H. and Singh, P. (2004)

- An attempt to capture common-sense knowledge in triples (huge domain)
- From this new common-sense knowledge is inferred and suggested to the user for approval
- Relations are more top-down, more formal
- Focusses also on expressing triples in natural language, and extracting from natural language

Methodology

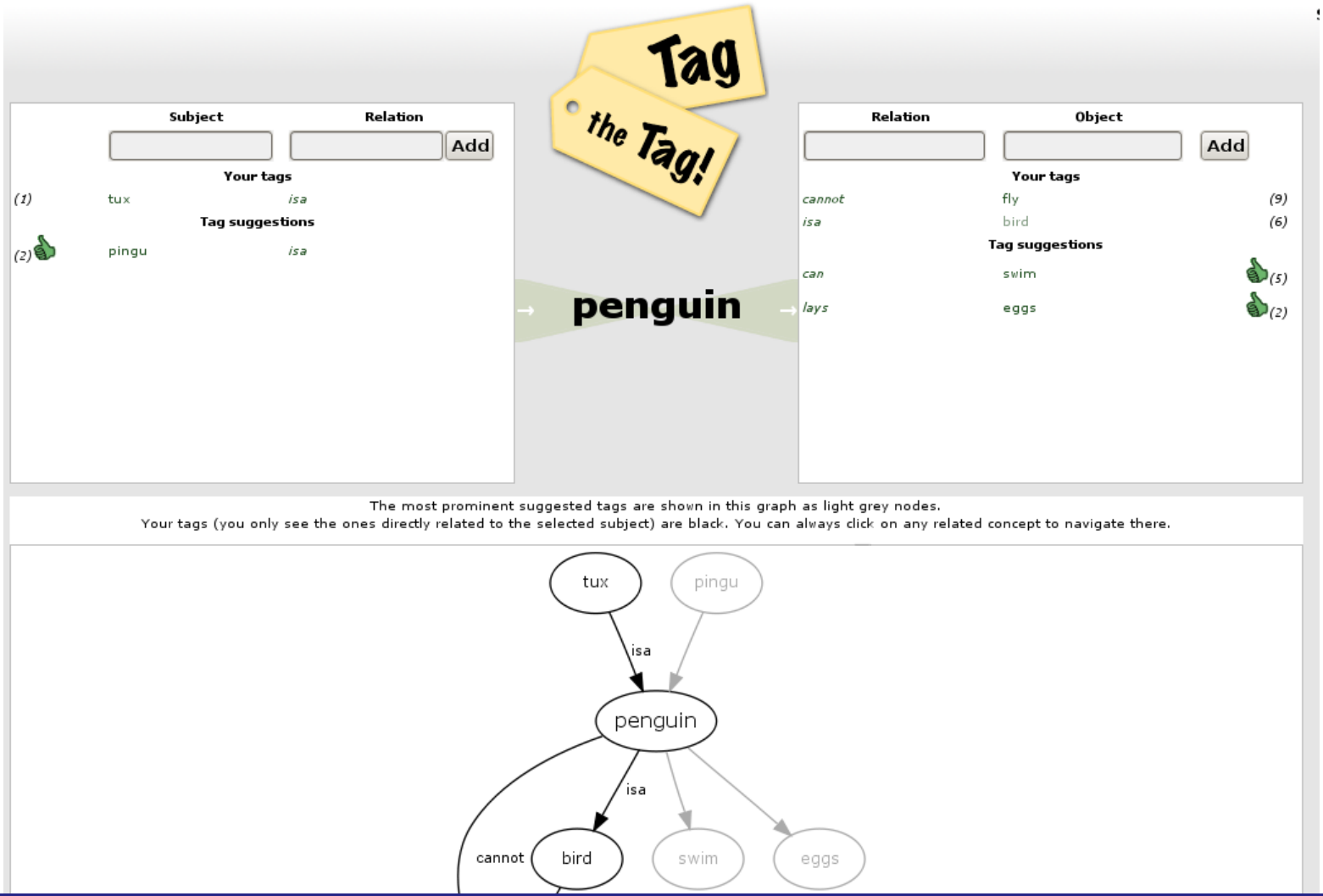
Type of research

Implementation & Experiment

Methodology

1. Build a web-based two-dimensional tagging system
2. Define domain and some initial tags (animals)
3. Volunteers tagged these resources using our new system
4. Analyse results
    1. What knowledge emerges from the data obtained?
    2. What problems can we identify?
    3. Does our approach improve traditional folksonomies?
    4. Can our approach help in constructing semantic web ontologies?

System


Tag
the Tag!
Subject
Relation
Add
Your tags
(1) tux isa
Tag suggestions
(2) pingu isa
penguin
Relation
Object
Add
Your tags
cannot fly (9)
isa bird (6)
Tag suggestions
can swim (5)
lays eggs (2)
The most prominent suggested tags are shown in this graph as light grey nodes.
Your tags (you only see the ones directly related to the selected subject) are black. You can always click on any related concept to navigate there.
tux
pingu
isa
penguin
isa
cannot
bird
swim
eggs

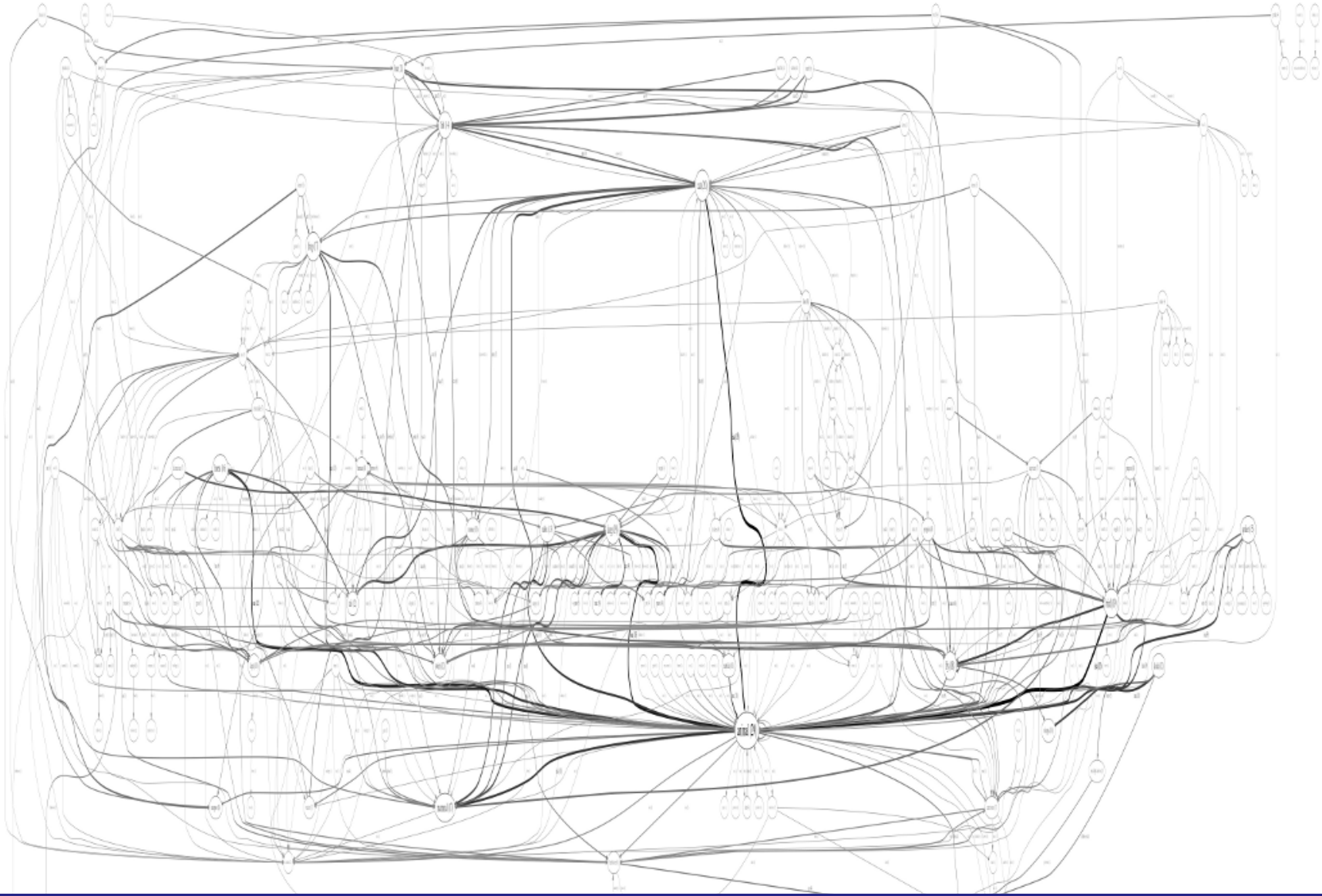
Data Analysis

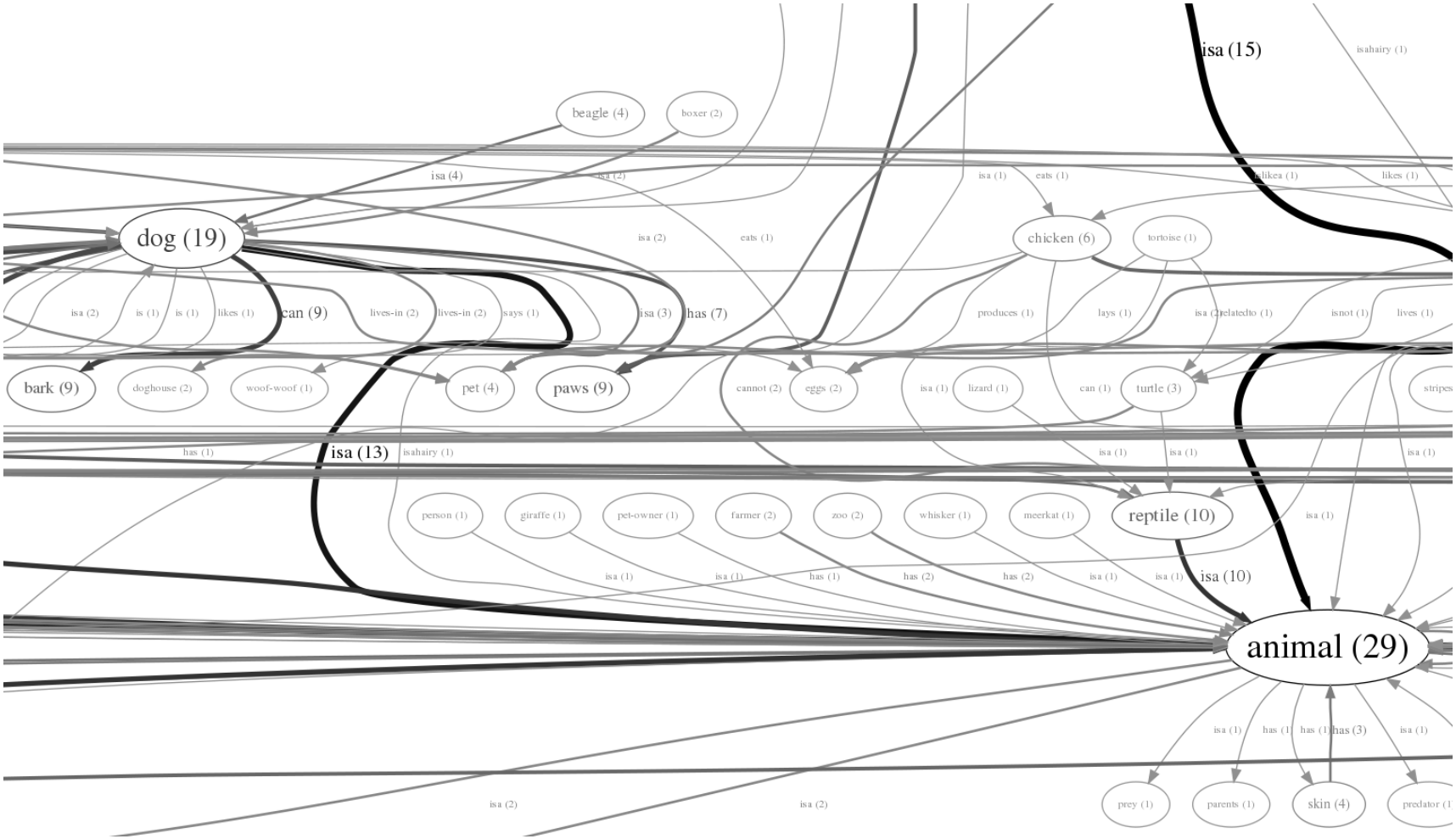
Data gathered

- 33 people
- 1043 triples, 464 unique triples
- 193 unique concepts
- 52 unique relations

What knowledge can we extract from the data obtained?

1. Overlap between individual semantic networks, popular concepts and relations




isa (15)
isahairy (1)
beagle (4)
boxer (2)
isa (4)
isa (2)
isa (1)
eats (1)
likea (1)
likes (1)
dog (19)
isa (2)
eats (1)
chicken (6)
tortoise (1)
isa (2)
is (1)
is (1)
likes (1)
can (9)
lives-in (2)
lives-in (2)
says (1)
isa (3)
has (7)
produces (1)
lays (1)
isa (2)
relatedto (1)
isnot (1)
lives (1)
bark (9)
doghouse (2)
woof-woof (1)
pet (4)
paws (9)
cannot (2)
eggs (2)
isa (1)
lizard (1)
can (1)
turtle (3)
stripes
has (1)
isa (13)
isahairy (1)
isa (1)
isa (1)
isa (1)
person (1)
giraffe (1)
pet-owner (1)
farmer (2)
zoo (2)
whisker (1)
meerkat (1)
reptile (10)
isa (1)
isa (1)
isa (1)
has (1)
has (2)
has (2)
isa (1)
isa (1)
isa (10)
animal (29)
isa (1)
has (1)
has (1)
has (3)
isa (1)
isa (2)
isa (2)
prey (1)
parents (1)
skin (4)
predator (1)

What knowledge can we extract from the data obtained?

If folksonomies are a *mess*, then we are *mess*^{2}! **Both concept-tags and relation-tags can be ambiguous!** However, our data doesn't really reflect this!

1. Split **homonymous** tags into two concepts -- `dog isa animal , animal isa muppet`
2. Merge **synonymous** tags into one concept -- `banana eaten-by monkey , banana consumed-by ape`
   - Syntactic similarity, semantic similarity (Wordnet?), Network analysis

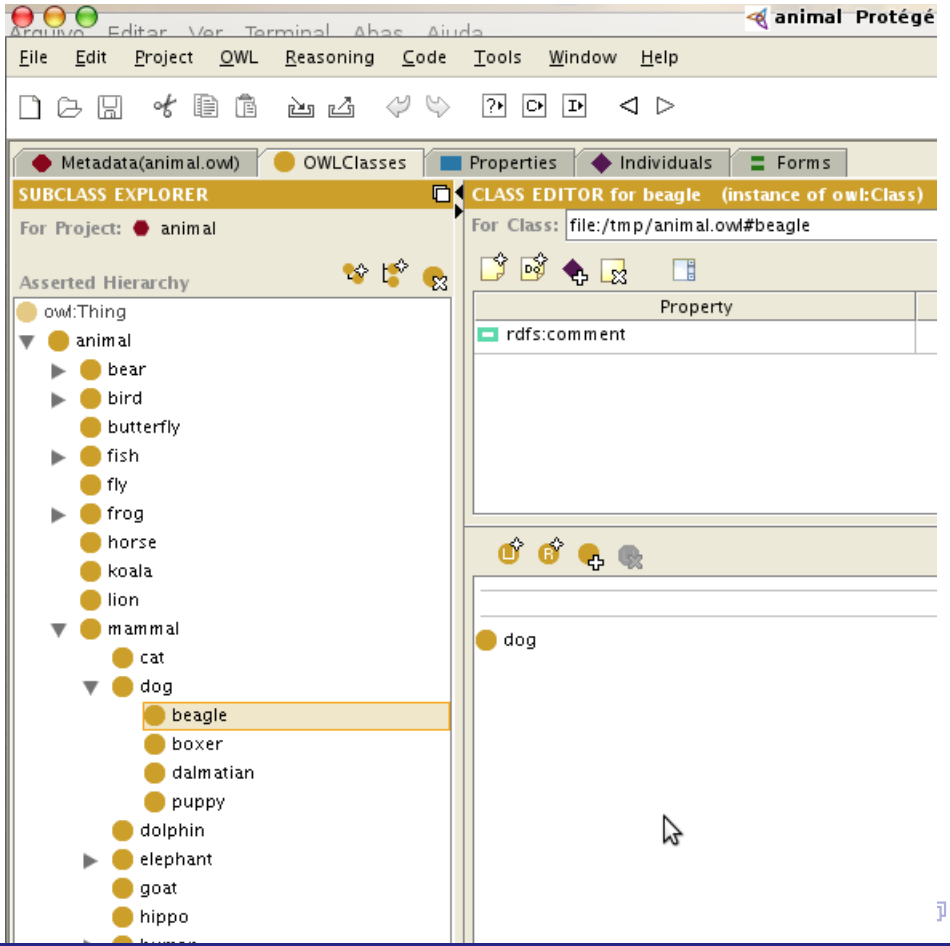
What knowledge can we extract from the data obtained?

Can we extract properties of relations? Such as transitivity, symmetry, reflexivity?

Average Reflexivity	1	breeds
	0.85714	marries
Average Transitivity	0.55556	breeds
	0.33333	turnsinto
Average Symmetry	1	family-of
	0.55556	relatedto
	0.42857	is
	0.4	isnot

What knowledge can we extract from the data obtained?

After semi-automatically identifying `isa` relations. We can attempt to extract a class-hierarchy as a basis for an ontology:



What problems can we identify?

1. Disambiguating synonyms and homonyms is hard
2. Lots of users needed to tag
3. Intransitive properties?

Does our approach improve traditional folksonomies?

1. Yes, in domains where more semantic expressiveness is needed

Can our approach help in constructing semantic web ontologies?

1. It can give a head-start and help explore the domain
2. It may aid in detecting conceptual problems in ontology-building
3. It can provide an initial ontology-skeleton from which to continue

Questions?