Moving towards formalisation

COMP62342

Sean Bechhofer
sean.bechhofer@manchester.ac.uk

Uli Sattler
uli.sattler@manchester.ac.uk

(thanks to Bijan Parsia for slides)
Previously...

- We started the **Knowledge Acquisition** process...
  - to *elicit* tacit knowledge
    ... in a variety of ways
    ... about a set of terms or concepts
- Even there we can be more/less **explicit & precise**
  - normalising terms: e.g., “symmetry or symmetric”?
  - hierarchy - and other direct relations between terms
  - categorizing terms: e.g., as modifiers or self-standing
  - **constraining** and **defining** terms

**Next:** 2 important steps

1. getting even more explicit & precise
   - Refining our proto-representation
2. getting actionable
   - Building a representation
There are several sorts of domesticated animals, though by far the most are mammals (like us!). For example, our faithful pets, cats and dogs, are clearly domesticated (or we would not keep such dangerous carnivores in our homes), as is the delicious yet docile cow which is farmed in ever increasing numbers.
Step 1: Term extraction

• Highlight the **relevant, domain-dependent** terms in:

There are several sorts of *domesticated animals*, though by far the most are *mammals* (like *us!*). For example, our faithful *pets, cats* and *dogs*, are clearly *domesticated* (or we would not keep such *dangerous carnivores* in our *homes*), as is the *delicious* yet *docile cow* which is *farmed* in ever *increasing numbers*. 
Step 1: Term extraction

• Pull these terms out
  – domesticated
  – animals
  – mammals
  – us
  – pets
  – cats
  – dogs
  – dangerous
  – carnivores
  – homes
  – delicious
  – cow
  – farmed
  – increasing
  – numbers
Step 1: Term extraction

- Pull these out and **ponder**:
  - domesticated
  - animals
  - mammals
  - us
  - pets
  - cats
  - dogs
  - dangerous
  - carnivores
  - homes
  - delicious
  - cow
  - farmed
  - increasing
  - numbers

These are quite odd but in different ways
Step 1: Term extraction

- Pull these out and **ponder some more:**
  - domesticated
  - animals
  - mammals
  - us
  - pets
  - cats
  - dogs
  - dangerous
  - carnivores
  - homes
  - delicious
  - cow
  - farmed
  - increasing
  - numbers

These are similar but have different levels of generality, and non-uniform spelling
Step 2: Grouping

• Base animal categories (noun-y terms)
  – animals
  – cats
  – dogs
  – mammals
  – cow
  – us

• Ways an animal can be (adjective-y terms)
  – domesticated
  – pets
  – dangerous
  – carnivores
  – delicious
  – farmed

• Stuff
  – homes
  – increasing
  – numbers
Step 2: Grouping

- Base animal categories (noun-y terms)
  - animals
  - cats
  - dogs
  - mammals
  - cow
  - us

- Ways an animal can be (adjective-y terms)
  - domesticated
  - pets
  - dangerous
  - carnivores
  - delicious
  - farmed

- Stuff
  - homes
  - increasing
  - numbers

Scoping:
Should we care about these?
A Key Slogan

to determine which terms to care about:

Representations are context sensitive & interest relative

• Context sensitive?
  – for which (kind of) application do we build KR?

• Interests?
  – Application needs
    • Teaching, categorising, data acquisition
  – Audience
    • Children, lay people, different disciplines, clinicians vs. researchers

• Establish context and relevant interests
  – Here: context is “this course unit/exercise”
  – Here: interests is “to work up a reasonable example”
Step 2: Grouping

- Base animal categories (noun-y terms)
  - animals
  - cats
  - dogs
  - mammals
  - cow
  - us

- Ways an animal can be (adjective-y terms)
  - domesticated
  - pets
  - dangerous
  - carnivores
  - delicious
  - farmed

- Stuff
  - homes
  - increasing
  - numbers

Scoping:
Should we care about these?
No! (Why?)
### Step 3: Normalise Terms

- **Base animal categories (noun-y terms)**
  - animals
  - cats
  - dogs
  - mammals
  - cow
  - us

- **Ways an animal can be (adjective-y terms)**
  - domesticated
  - pets
  - dangerous
  - carnivores
  - delicious
  - farmed

### Unify
- number (singular/plural)
- spelling
  (incl. upper/lower case)

- Animal
- Cat
- Dog
- Mammal
- Cow
Step 3: Normalise Terms

- Base animal categories (noun-y terms)
  - Animal
  - Cat
  - Dog
  - Mammal
  - Cow
  - Us

- Ways an animal can be (adjective-y terms)
  - Domesticated
  - Pets
  - Dangerous
  - Carnivores
  - Delicious
  - Farmed

Give a good name

- Animal
- Cat
- Dog
- Mammal
- Cow
- Human
Step 3: Normalise Terms

- Base animal categories (noun-y terms)
  - Animal
  - Cat
  - Dog
  - Mammal
  - Cow
  - Human

- Ways an animal can be (adjective-y terms)
  - domesticated
  - pets
  - dangerous
  - carnivores
  - delicious
  - farmed

Unify grammatical form & spelling

- Domesticated
- Pet
- Dangerous
- Carnivorous
- Delicious
- Farmed
Step 3: Normalise Terms

• Base animal categories (noun-y terms)
  – Animal
  – Cat
  – Dog
  – Mammal
  – Cow
  – Human

• Ways an animal can be (adjective-y terms)
  – Domesticated
  – Pet
  – Dangerous
  – Carnivorous
  – Delicious
  – Farmed

We have some background knowledge we can use to “round out” these terms
Step 3: Normalise Terms

• Base animal categories (noun-y terms)
  – Animal
  – Cat
  – Dog
  – Mammal
  – Cow
  – Human

• Ways an animal can be (adjective-y terms)
  – Domesticated
  – Pet
  – Dangerous
  – Carnivorous
  – Omnivorous
  – Herbivorous
  – Delicious
  – Wild
  – Farmed

...so we add some terms
Step 4: Organise Terms

- Base animal categories (noun-y terms)
  - Animal
  - Mammal
  - Cat
  - Dog
  - Cow
  - Human

- Ways an animal can be (adjective-y terms)
  - Domesticated
  - Wild
  - Dangerous
  - Carnivorous
  - Omnivorous
  - Herbivorous
  - Delicious
  - Pet
  - Farmed
Step 4: Organise Terms

- Base animal categories (noun-y terms)
  - General:
    - Animal
    - Mammal
  - Specific:
    - Cat
    - Dog
    - Cow
    - Human

- Ways an animal can be (adjective-y terms)
  - General:
    - Domesticated
    - Wild
    - Dangerous
    - Carnivorous
    - Omnivorous
    - Herbivorous
    - Delicious
  - Specific:
    - Pet
    - Farmed

Next:
What terms are definable?
Interlude: what is a definition?

- Mini-exercise:
  - agree with your neighbour on a definition for
    - pet
    - person
    - table (furniture)
Interlude: Definitions?

A definition

- is a statement that fixes the meaning of a term
- can be
  - extensional: enumerate all elements a term describes e.g., “PrimaryColour = {Red, Yellow, Blue}”
  - intensional: often using genus–differentia pattern i.e., giving the next more general term (genus) plus differentiating features for this term and its siblings e.g., “An endotherm is an organism that maintains its body at a metabolically favourable temperature.”

Two consequences:
if Bob is an endotherm, then I know that…
if I find an organism that maintains its temperature…, then …
Step 4: Organise Terms

- Base animal categories (noun-y terms)
  - General:
    - Animal
    - Mammal
  - Specific:
    - Cat
    - Dog
    - Cow
    - Human

- Ways an animal can be (adjective-y terms)
  - General:
    - Domesticated
    - Wild
    - Dangerous
    - Carnivorous
    - Omnivorous
    - Herbivorous
    - Delicious
  - Specific:
    - Pet
    - Farmed

Red terms are easily definable (?)
Step 5: Define Terms

- Base animal categories (noun-y terms)
  - General:
    - Animal = eats some Stuff
    - Mammal = has MammGlands
  - Specific:
    - Cat
    - Dog
    - Cow = eats only Grass
    - Human = Omnivore

- Ways an animal can be (adjective-y terms)
  - General:
    - Domesticated
    - Wild
    - Dangerous
    - Carnivorous = eats only Meat
    - Omnivorous = eats Meat & Plants
    - Herbivorous = eats only Plants
    - Delicious = tastes good
  - Specific:
    - Pet = lives with Humans
    - Farmed = is eaten/used

New Terms:
eats, lives, tastes…
= , only, &
Stuff
Plants, Meat,…
An interlude/orientation
Capturing knowledge in an *actionable* form

- We can capture what we’ve done
  - in a *text document*
    - nice to read for humans
    - not easily understandable/processable by a computer:
      “which animals are there?” involves tricky string hackery!
  - in a *structured way*
    ...i.e., some form of knowledge base
  ⇒ and get some benefits!
Capturing our knowledge

• is an iterative process
• so far, representation is **informative**
  – Definitions (will) elicit new terms
  – Interests and Context tell us when we’re done, i.e., when a fix point is reached
    • Fatigue! Fatigue works...

• Until now, entirely informal, human process
  – Having a structured form helps a little
    • Generic versus specific
    • Self-standing (noun-y) versus Modifiers (adjectiv-y)
    • Contraries
    • Definitions
    • …could be used for easier search/browsing
  – But no “content” feedback
  – For this, we need to understand what we want to/can represent
So far...

• We are well into KA
  – Term extraction
  – Initial regimentation
    • Normalisation
    • Organise
      – Hierarchical organisation
      – Categorisation
  – Started additional capture
    • Adding definitions

• Ready to consider the next step
  – Proto-Formalisation!

• Remember:
  – Interest sensitive and context relative
  – We’re looking for benefits (to way against costs)

• But first...
Remember our passage

- With highlighting!

There are several sorts of domesticated animals, though by far the most are mammals (like us!). For example, our faithful pets, cats and dogs, are clearly domesticated (or we would not keep such dangerous carnivores in our homes), as is the delicious* yet docile cow which is farmed in ever increasing numbers.

- Why not?

There are several sorts of domesticated animals, though by far the most are mammals (like us!). For example, our faithful pets, cats and dogs, are clearly domesticated (or we would not keep such dangerous carnivores in our homes), as is the delicious* yet docile cow which is farmed in ever increasing numbers.
So, what terms should go in?

- It depends!
  - Interests and context
  - Resources, including
    - Time
    - Energy
    - Representational capabilities
    - Skill, etc.

- Fewer than all
  - A generally good rule of thumb

- Other than what’s there
  - Another good rule of thumb!
  - “Fleshing out”
    - Organisational needs (e.g., “LivingThing”)
    - Representational needs (e.g., “eats”)
    - Coverage, “completeness” (e.g., “omnivore”)

Scoping: use Competency Question to decide!
Back to our Term Definitions
Step 5: Define Terms

- Base animal categories (noun-y terms)
  - General:
    1. Animal = eats some Stuff
    2. Mammal = has MammGlands
  - Specific:

- Ways an animal can be (adjective-y terms)
  - General:
    - Domesticated
    - Wild
    - Dangerous
  - Specific:
    3. Cow = eats only Grass
    4. Human = Omnivore
    5. Carnivorous = eats only Meat
    6. Omnivorous = eats Meat & Plants
    7. Herbivorous = eats only Plants
    8. Delicious = tastes good

Discuss:
Which of these definitions is really good?
I.e., is really a definition?
What about these new terms?

- eats, lives, tastes...
- =, only, &
- Stuff
- Plants, Meat, …

Domain dependent, but verb-y
i.e., of a new kind!

Logic-y

???

New domain dependent Noun-y terms
Let’s try to formalise: towards actionable form!

Use Protégé & OWL rather than Word!
Underlying OWL Language

Class: Cow
Annotations:
  rdfs:comment "eats only Plants",
  rdfs:comment "Definable",
  rdfs:comment "SelfStanding"
SubClassOf:
  Mammal

OWL has many syntaxes; this is one of them called Manchester Syntax
Recall our first knowledge base:

- **Base animal categories (noun-y terms)**
  - General:
    1. Animal = eats some Stuff
    2. Mammal = has MammGlands
  - Specific:
    - Cat
    - Dog
    3. Cow = eats only Grass
    4. Human = Omnivore

- **Ways an animal can be (adjective-y terms)**
  - General:
    - Domesticated
    - Wild
    - Dangerous
  - Specific:
    5. Carnivorous = eats only Meat
    6. Omnivorous = eats Meat & Plants
    7. Herbivorous = eats only Plants
    8. Delicious = tastes good

Which of these definitions is really good? I.e., is really a definition?
Our mini-formalisation in OWL

Class: Cow

Annotations:
- rdfs:comment "eats only Plants",
- rdfs:comment "Definable",
- rdfs:comment "SelfStanding"

SubClassOf: Mammal

(Hierarchical) Relation to other term

Another named term
Meaning? Semantics?

Class: Cow

Annotations:
- rdfs:comment "eats only Plants"
- rdfs:comment "Definable"
- rdfs:comment "SelfStanding"

SubClassOf:
- Mammal

Stands for a set

Subsumption: Every Cow is a Mammal

More later today!
**Benefits of this formalisation?**

**Class:** Cow  
**Annotations:**  
rdfs:comment "eats only Plants",  
rdfs:comment "Definable",  
rdfs:comment "SelfStanding"

**SubClassOf:**  
Mammal

- Gives some structure to our set of terms:  
  - a **hierarchy** that we can browse  
  - we can retrieve classes  
  - we can search for comments
Side note: A “Computer View”

**Class**: Blah

**Annotations**:
- `rdfs:comment` "b123 623 7y3",
- `rdfs:comment"mch345",
- `rdfs:comment"lkjherhjhhhh"

**SubClassOf**: Foo
Better Annotations

**Class:** Cow

**Annotations:**
- `rdfs:comment` "eats only Plants",
- `isDefinable` True
- `hasGrammaticalType` SelfStanding

**SubClassOf:** Mammal

For less string-hackery and easier data-entry

Use good annotation properties
A Better Definition

Class: Cow
Annotations:
isDefinable True
hasGrammaticalType SelfStanding
EquivalentTo:
eats only Plant
SubClassOf:
Mammal

But why?
...we need to learn more about OWL!
...see next Section!