Knowledge Elicitation Exercise
COMP34512

Sebastian Brandt
brandt@cs.man.ac.uk
Knowledge Acquisition (KA)

• Operational definition
  – Given
    • a source of (propositional) knowledge
    • a sink
  – KA is the transfer of propositions from source to sink

• Elicitation (for terminological knowledge)
  – Initial Capture:
    • Source: People, “experts”, “domain experts” (DE)
    • Sink: “Protocol” (record of behavior)
  – Term Extraction:
    • Source: Text (e.g., transcript, textbook, Wikipedia article)
    • Sink: List of terms (perhaps on cards)
  – Initial Regimentation:
    • Source: List of terms (on cards!)
    • Sink: Proto-representation
      – Hierarchy of categorized, normalized terms (with notes!)
Modelling Techniques

• (Often characterized by aspects of the target)
• Being picky
  – Pedantic refinement
• Hierarchy-generation techniques
  – such as *laddering* are used to build taxonomies or other hierarchical structures such as goal trees and decision networks.
• Matrix-based techniques
  – involve the construction of grids indicating such things as problems encountered against possible solutions.
• Sorting techniques
  – are used for capturing the way people compare and order concepts, and can lead to the revelation of knowledge about classes, properties and priorities
• Limited-information and constrained-processing tasks
  – are techniques that either limit the time and/or information available to the expert when performing tasks. For instance, the twenty-questions technique provides an efficient way of accessing the key information in a domain in a prioritised order.
Reminder: An Animals Taxonomy

• Task:
  – generate a controlled vocabulary for an index of a children’s book

• Domain:
  – Animals including
    • Where they live
    • What they eat
      – Carnivores, herbivores and omnivores
    • How dangerous they are
    • How big they are
      – A bit of basic anatomy
        » legs, wings, fins? skin, feathers, fur?
    • ...
      – (read the book!)

• Representation aspects
  – Hierarchical list with priorities
We’ve sorted

Horse
Sheep
Cat
Wolf
Bear
Cow
Dog
Animal

Plant
Wheat
Grass
Tree
Fish
Herring
Goldfish
Shark
Trout
Triadic Elicitation: The 3 card trick

• Select 3 cards at random
  – Identify which 2 cards are the most similar?
    • Write down why (a similarity)
      – As a new term!
    • Write down why not like 3rd (a difference)
      – Another new term!

• Helps to determine the characteristics of our classes
  – Prompts us into identifying differences & similarities
    • There will always be two that are “closer” together
    • Although which two cards that is may differ
      – From person to person
      – From perspective to perspective
      – From round to round
Example

- Cat
- Sheep
- Bear
- Cat
- Bear
- Sheep
- Fur
- Wool
Same(?) Example
Protocol Analysis

• From interviews/behaviour to analysable items
  – Text! Text is good!

• From a text,
  – find key terms
  – normalize them
    • capitalisation, pluralization (or not), orthography, etc.

• Keep track of
  – Significance
    • Core or peripheral terms
    • Illustrative? Defining?
  – Situation
    • Sentences or sections

• Output: List of Terms
Key Goal: Laddering

• Terms *vary* in generality
  – Tree *vs.* Plant
  – Dog *vs.* Rover

• Each sort may be *implicit*!
  – Goal: *Flesh out* the generality hierarchy
    • Get more specific (if too general)
    • Get more general (if mostly specific)

• How?

1. Take a group and ask *what they have in common*
   • During sorting or 3-card or directly

2. Then *investigate relations* of new term
   • Siblings, missing children, and (eventually) parents (back to 1)
A (Partial) Hierarchy

- Living Thing
  - Animal
    - Mammal
      - Cat
      - Dog
      - Cow
      - Person
    - Fish
      - Trout
      - Goldfish
      - Shark
  - Plant
    - Tree
    - Grass
    - Wheat
Categorisation: “Grammatical”

• **Types\Classes\Categories**
  - **Self standing entities**
    • Things that can exist on their own
    • People, animals, houses,
      - actions, processes, …
    • Roughly nouns

• **Modifiers**
  - Things that modify (“inhere”) in other things
  - Roughly adjectives and adverbs

• **Relations\Properties**
  - Things which relate two individuals
  - Roughly verbs, and (variable) attributes
  - (Perhaps defer to later)
Categorisation: Modelling

• In general, given a set of terms:
  – We describe the world using them
  – We describe terms using other terms
    • Paradigmatically, we define terms

• Assumable
  – Terms which have no or minimal modelling
    • Too hard to model or not needed to model or don’t know
      – For “Living thing” we might just have a list of subclasses
    – Sometimes known as the “primitive vocabulary”

• Definable
  – Terms for which we can give a full definition
    • Or reasonably full definition
  – “Carnivore is an animal that eats only meat.”
Result!

- Living Thing
  - Animal
    - Mammal
      - Cat
      - Dog
      - Cow
      - Person
    - Fish
      - Trout
      - Goldfish
      - Shark
    - Plant
      - Tree
      - Grass
      - Wheat

- Modifiers
  - Domestic
    - Pet
    - Farmed
      - Draft
      - Food
  - Wild
  - Health
    - Healthy
    - Sick
  - Sex
    - Male
    - Female
  - Age
    - Adult
    - Child

- Relations
  - eats
  - owns
  - parent-of
  - ...

- Definable
  - Carnivore
  - Herbivore
  - Child
  - Parent
  - Mother
  - Father
  - Food Animal
  - Draft Animal
So! The Task

• Capture
  – See printouts

• Extract
  – List of terms; put them on cards!

• Extend
  – Esp. laddering

• Categorise
  – As modifier vs. self-standing
  – As definable
    • Sketch definition on (back of) card
So! The Task

- **Capture**
  - See printouts

- **Extract**
  - List of terms; put them on cards!

- **Extend**
  - Esp. laddering

- **Categorise**
  - As modifier vs. self-standing
  - As definable
    - Sketch definition on (back of) card

---

**Explicit stuff**

**Implicit stuff**
So! The Task

- **Capture**
  - See printouts

- **Extract**
  - List of terms; put them on cards!

- **Extend**
  - Esp. laddering

- **Categorise**
  - As modifier vs. self-standing
  - As definable
    - Sketch definition on (back of) card

- **Encode!**

Explicit stuff

Implicit stuff
Coursework

• Take the KE done in class
  – Feel free to refine it further

• Encode it using Protege 4
  – Each category term becomes a class
    • Capture your hierarchy using subsumption/subclassing
  – Each relation becomes a property
  – For each class
    • Add a comment saying “Modifier” or “Self-Standing”
      – Depending on which it is!
    • Add a comment saying “Definable”
      – If it is so according to your elicitation
      – If so, add a comment given your (English) definition

• Submit a zipped version of your RDF/XML file

• Full description on Blackboard!
• Deadline 13. February 2014