Organizing Harvested Knowledge

Eduard Hovy
USC/ISI
(and soon, CMU)
Our basic premise

- NLP applications need [semantic] information
- Every semantic representation ever built can be represented as a network/graph/frame of units (terms/nodes) and relations (links/edges)
- By cleverly matching patterns to large corpora in various ways we can harvest [most?] units and relations
So it is natural to ask...

• What kinds of units and relations should be harvested [for a given domain]?
  – How many?  
  – How to integrate, organize, and store them?  

• How do you measure results?
  – Precision and Recall — against what?
  – Extrinsic tasks — which?  

• How do you make public and integrate results?
  – YAGO, NELL, Textrunner, others....  

Don’t know
Kinds of knowledge harvested

• **Concepts**: *car, jaguar, dreaming*
• **Instances**: *Albert Einstein, Montreal, WW II*
• **Relations** (perhaps with properties / role filler constraints): *color-of(entity,color), birthdate(living-being,date)*
  – Eventually: use partly instantiated relations as axioms: *insects fly-to flowers, people fly-to cities/parties/family*

• **Challenges**:
  – Sense disambiguation
  – Synonymy (entity linking)
Let’s talk about relations

• The semantics is not in the units/symbols, it’s in their relationships to one another
• Luckily, we can also harvest these relations
• The trouble is, we don’t know which relations there are, overall or even per concept type...
  – Entity relations:
    • is-a
    • Physical entities: relations for all the attributes (size, color, weight, age, name, etc.)
    • Non-physical ones: other attributes?
    • What about functions, sources, etc.?
  – Event relations:
Approaches toward event roles

1. Case roles:
   – Charles Fillmore: *The Case for Case*, 1968. Each verb has a set of ‘deep case’ roles, named
   – Usual approach: ~25, then ~150, taxonomized

2. Dependencies, some of them named:
   – Lucien Tesniere: *Éléments de syntaxe structurale*, 1959. (Father of dependency grammar.) Each verb has a small core set of idiosyncratic roles, plus additional standard ones
   – Like PropBank’s arg0, arg1, ..., argM-loc, argM-time

3. FrameNet/Hobbs solution:
   – Every entity has its own relation set; don’t bother to try to generalize
The FrameNet solution

• FrameNet:
  – *Why do you use frame-specific frame element names rather than thematic roles?* First of all, there are too many different semantic relations to fit into any of the so-called standard list of thematic roles or case roles. We are in the process of preparing a more complete answer to this question...
  – *Doesn't this frame-specific approach lead to multiple names for what is really the same frame?* Strictly speaking the frame element names proposed for one frame are relative to that frame, so decisions about choosing labels that are also used in other frames are always reparable. We want the cross-frame recycling of frame element names to be justified, ultimately, through establishing principles of frame inheritance. The picture is complicated ... because of the possibility of multiple inheritance: the same argument of a single predicate can be seen as an instance of one frame element by virtue of its membership in one frame, of another frame element through its participation in a different co-existing frame. For our purpose in the annotation phase of the work, the main function of the frame element labels is to be transparent to the annotators....
Approaches 2

4. The ‘subatomic particle’ solution: Try to develop elemental aspects of relations and combine them to form case roles:
   - Patient (entity changed by event): \(+\text{change}, -\text{uses-energy}, +\text{event-root-entity}\)
   - Theme (entity is unchanged by event, info is needed): \(-\text{change}, -\text{uses-energy}, +\text{event-root-entity}\)
   - Agent (uses ‘energy’ to initiate event, causes change): \(-\text{change}, +\text{uses-energy}, +\text{cause}, -\text{event-root-entity}\)
   - Instr (mediates transmission of ‘energy’ to patient, or of info to agent): \(-\text{uses-energy}, -\text{change}, +\text{cause}, -\text{event-root-entity}\)

5. Autoencoders:
   - Neural networks encode words into vector spaces that predict how likely given words are to appear in given role positions
   - This implicitly encodes role preferences
   - Collobert and Weston 08; Bengio 09; Turian et al. 10; Socher et al. 11
Taxonomy of top-level case relations

• Events: Case roles
  - Agent family: Agent, Experiencer
  - Patient family: Patient, Theme
  - Instr family: Tool, Prop
  - Spatio-Temporal family: Loc, Source Dest, Time
  - Benef

  Prep senses: SemEval 2007; Hovy et al. 11

• Objects: Property relations
  - Structure family: Morphology, Material
  - Function family: Use, Purpose, Operation
  - Provenance family: Source, Maker

  Noun-noun relation senses: Tratz and Hovy 10
Entities: Noun-noun relations

- Created about 45 relations, taxonomized
  - Annotated 15k NN expressions
  - Compared to and absorbed data from previous studies in NLP and Linguistics
  - Relations overlap about 50% with SemEval prep senses
- Validation: Annotated test data on MTurk
  - 25 annotators, 8c/decision
  - Weighted them by overall group agreement
- Built automated NN classifier
  - 10x cross-validation: 79.3% agreement
  - Merged into Tratz parser
- See (Tratz and Hovy ACL-10)
Tratz’s noun-noun relations

- **Causal Group**
  - Communicator of Communication
  - Performer of Act/Activity
  - Creator/Provider/Cause Of

- **Purpose/Activity Group**
  - Perform/Engage_In
  - Create/Provide/Sell
  - Obtain/Access/Seek
  - Modify/Process/Change
  - Mitigate/Oppose/Destroy
  - Organize/Supervise/Authority
  - Propel
  - Protect/Conserve
  - Transport/Transfer/Trade
  - Traverse/Visit

- **Temporal Group**
  - Time [Span] + X
  - X + Time [Span]

- **Ownership, Experience, Employment, and Use Group**
  - Possessor + Owned/Possessed
  - Experiencer + Cognition/Mental
  - Employer + Employee/Volunteer
  - Consumer + Consumed
  - User/Recipient + Used/Received
  - Owned/Possessed + Possession
  - Experiencer + Experiencer
  - Thing Consumed + Consumer
  - Thing/MMeans Used + User

- **Location & Whole+Part/Member of**
  - Location/Geographic Scope of X
  - Whole + Part/Member Of

- **Composition & Containment Group**
  - Substance/Material/Ingredient + Whole
  - Part/Member + Collection/Configuration/Series
  - X + Spatial Container/Location/Bounds

- **Topic**
  - Topic of Communication/Imagery/Info
  - Topic of Plan/Deal/Arrangement/Rules
  - Topic of Observation/Study/Evaluation
  - Topic of Cognition/Emotion
  - Topic of Expert
  - Topic of Situation
  - Topic of Event/Process

- **Attribute Group**
  - Topic/Thing + Attribute
  - Topic/Thing + Attribute Value Characteristic Of

- **Attributive and Coreferential**
  - Coreferential
  - Partial Attribute Transfer
  - Measure + Whole

- **Other**
  - Highly Lexicalized / Fixed Pair
  - Other
Comparison to other studies

<table>
<thead>
<tr>
<th>Score</th>
<th>Type</th>
<th>Relations</th>
<th>Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>.57</td>
<td>Kappa</td>
<td>43</td>
<td>Tratz and Hovy, 2010</td>
</tr>
<tr>
<td>.61</td>
<td>Kappa</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>.68</td>
<td>Kappa</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>.52</td>
<td>Agree</td>
<td>20</td>
<td>Kim and Baldwin, 2005</td>
</tr>
<tr>
<td>.58</td>
<td>Kappa</td>
<td>35 (21)</td>
<td>Girju et al., 2005</td>
</tr>
<tr>
<td>.80</td>
<td>Kappa</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>.61</td>
<td>Kappa</td>
<td>22</td>
<td>Girju, 2007</td>
</tr>
<tr>
<td>.68</td>
<td>Kappa</td>
<td>6</td>
<td>Ó Séaghdha, 2007</td>
</tr>
</tbody>
</table>
Tratz: MTurk Kappa agreements (N > 15 only)
Tratz: Mturk weighted Kappa agreements (N > 15 only)
Summary for semantic relations

• **Event**-anchored relations: Various proposals
  – Roles: Proposals of ~20, ~150... say, 100s? 1000s?
  – Inter-event (discourse) relations: RST 25; Penn ~200...

• **Entity**-anchored relations:
  – Noun-noun: Proposal of about 43
  – Noun-modifier: 1,000s? 10,000s? Open????

• SemEval provides about 45 **preposition** senses, some of which signal event roles (“by” = :agent, etc.) and others NN relations

• The long tail is a serious problem
Basically, we just don’t know much about the appropriate/right set of relations....

So let’s just take something simple...What do we know about *is-a*?
IS-A FOR TAXONOMIZING CONCEPTS
Aristotle, the father of it all

- The gold plated KR approach:
  - Find a primitive concept—undefined
  - Specialize it in various ways by adding various differentiae
  - Define these differentiae elsewhere in the ontology
  - Don’t confuse **definitional** aspects with mere **properties**!

  An *apple* is-a *fruit* with essential differentium XXX and properties
  
  :color=red, :size=tennis-ball-sized...

- **human**
  - :gender: man, woman
  - :gender, :animacy…

- **human**
  - :age
  - :gender
      - adult
      - child
      - man
      - woman
      - boy
      - girl
The main relation: Subsumption

(also called *a-kind-of, is-a...*)

Specialization enables property inheritance

Purposes:

• **Reduces storage** required: store only additional local differentiae at each node

• **Lends perspicacity**: easier to see relationships of nodes

• **Supports inference**: allows general rules to apply
The problem with Aristotle

• Problems:
  – What are the differentiae?
  – How do you order them?

• There’s no ‘correct’ order
• Automated taxonomizing ends up with tangled hierarchies
Tweety and the Nixon Diamond

• Simple is-a allows problems with values inherited from (multiple) parents
  – So people build systems of defeasible reasoning; various logics that operate over the ontology’s symbols and structure

• Just building naïve is-a hierarchies defeats the purpose of supporting inference
CYC has 2 subsumption relations

*genls* (subset): transitive

*isa* (element-of): not transitive
CYC has two subsumption relations

- **gensls (subset of):** transitive
- **Isa (element of):** not transitive
Five styles of truth

1. Abstraction and feature combination: the philosophers
2. Intuitive distinctions: the cognitive scientists
3. Inference-based organization: the computational people
4. Cross-linguistic phenomena: the linguists
5. Domain analysis: the domain specialists

• Taxonomic clarity: everyone
Example cognitive scientist: Rosch

- Functional purpose of classes: “provide maximum information with the least cognitive effort”
- Established experimental paradigms for determining subjects’ ratings of how good an example of a category a member is judged to be
- **Basic Level** categories:
  - A basic category is the largest class of which we can form a fairly concrete image, like *chair* or *ball*. These are the first classifications that children make
  - Superordinate categories are collections of basic categories: *furniture* includes chairs, lamps, desks, beds, etc.; *toys* include balls, dolls, furry animals. No one object clearly represents them
  - Subordinate categories represent divisions of basic classes (e.g., *deck chairs*, *bar stools*, *teddy bears*, *school desks*)
The problem of categories: The Prototype Theory view

• Traditional theory: people categorize using the common features of the members (differentiae)

• Rosch observations:
  – (1) When people categorize, they cannot tell you what features they are using — often don’t know the differentiae!
  – (2) When people categorize, they usually find some members of categories more “typical” (“better”) than others (e.g., a robin is a better member of the category Bird than an ostrich)
  – (3) When people categorize, they categorize more typical members more quickly than less typical ones

• Rosch suggestion:
  – Create ‘star structure’ of prototypes rather than (or in addition to?) a subsumption hierarchy with differentiae
Example computationalist: me

• Computer scientists write programs and created several data types, each one with a different function.
  - For example, for MT, typical termsets:
    • Part of speech tags
    • Syntactic categories
    • Named entity categories (Person, Organization, Numerical-expression, Location, Time-expression, etc.)

• These define a small set of categories
• The program does different things with the different categories
• Making the set of terms explicit → ontology
Example (cognitive) linguist: Lakoff

- Create classes according to the way one (or more) language(s) behave(s):
  - Classes of noun
  - Classes of verb
- Do they make conceptual categories? How do we judge?
- E.g., Dyirbal noun categorization:
  - Class I: human males + storms, rainbow (from myths) + fish (and so also fishing tools) + moon (husband of the sun) + ...
  - Class II: human females + birds (myth: because they have female spirits) + sun (wife of the moon) + fire (associated with sun) + hot things (experienced like fire) + ...
  - Class III: edible plants
  - Class IV: the rest
- E.g., Hopi time categorization
WordNet: Miller and Fellbaum

- Cognitive scientists at Princeton University
- Word hierarchy built by hand during 1980s, using dictionaries and manual insight
- Approx. 110,000 nodes at present:
  - Synonym, antonym, part-of links; examples; frequencies
  - WordNets for other languages: EuroWordNet (Vossen et al.): Dutch, Italian, Spanish, English
  - Global WordNet: see http://www.globalwordnet.org/
- Hierarchy info:
  - Noun hierarchy depth ~12
  - Verb hierarchy depth ~3
  - Adjective/adverb not in hierarchy, but in star structure
  - Almost no top-level structure
- Freely available: http://wordnet.princeton.edu/
- Extensively used in CompLing, but not very useful yet
  - Except: definitions converted to axioms and used for theorem proving in automated QA (Moldovan et al.)
Example linguist: Matthiessen

- Penman NL generation system (ISI, 1979–1997, with Bill Mann and others); KPML (various; 1995–; John Bateman):
  - Systemic-functional Linguistics grammar and system
  - Penman Upper Model: taxonomy (network) of approx. 300 terms
  - Input representation terms defined in Domain Model; connected to Upper Model
  - For NLG, many grammar decisions determined by very general categories capturing English structure and word behavior:
    - Nouns / verbs (of various types) / adjectives
    - Count nouns / mass nouns
    - Tenses etc.
- Upper Model nodes represent conceptual-grammatical categories: at interface of language and world
Domain specialist examples

• Computational / expert systems:
  – Protégé Ontologies Library: Stanford University’s collection of 18 influential ontologies (http://protege.stanford.edu/ontologies/ontologies.html)
  – OntoSelect: over 700 ontologies in various domains (http://views.dfki.de/Ontologies/)

• Medical:
  – UMLS: Metathesaurus (over 1 mill biomedical concepts and 5 mill concept names from over 100 controlled vocabularies and classifications (some in multiple languages) used in patient records, administrative health data, bibliographic and full-text databases, expert systems), the Semantic Network (isa for type hierarchy; physically related, spatially related, temporally related, functionally related, conceptually related), and the SPECIALIST lexicon (http://www.nlm.nih.gov/research/umls/)

• Industrial etc.:
  – NAICS (North American Industry Classification System): numerical classifications of construction, agriculture, technology, wholesale, retail, industry, etc. (http://www.census.gov/epcd/www/naics.html)
Domain specialists

• Is a *dolphin* a *mammal* or a *fish*?
• Is a *steelhead trout* a *salmon* or not?
• When is someone *Jewish*?

• Which features are the determinate ones? Why? Who decides?

There is no *authority*: it can be tradition, the law, social consensus, or simply ad hoc purpose-driven. The point is to know which you adopt and to be *careful and consistent*. 
SHISH KEBABS
Concept ‘facets’

• **Library** as Org and Loc and Building
  – 3 *is-a* relations to 3 parents?
  – 3 separate senses?

• Same for Hospital, FireStation, PoliceStation, School...

• Other examples:
  – Building & Organization & Location: *library, school, museum*...
  – Company & Product & Stock: *most public companies*
  – LanguageObject & Event & Document: *agreement, vow*...
  – StudyDomain & Activity & Creation: *architecture, science, music*...
Shish kebabs
Why is this interesting?

• Without knowing about shish kebabs, your system can’t get a concept ironed out:
  “the library hired me for the summer”
  “the library burned down in April”
  “I’ll meet you at the library”
  “I drink Coke all the time”
  “Coke fired me last week”
  “Coke has dropped 5c”

• This is going to mess up any taxonomization algorithm you have
TAXONOMIZING USING DAP
Kozareva and Hovy

• **DAP: Double-anchored patterns:**
  - DAP: [? such as A and ?] [? flies-to A and ?] ...
  - Double anchoring -> recursivity -> few seeds but higher precision
  - ‘Forward’ and ‘backward’ application

• **Experiments exploring the power of DAP**
  - Concepts: ACL-08 (with Riloff), EMNLP-09
  - Taxonomizing: EMNLP-10, ACL-10
  - Relations: ICSC-11
  - Determining what makes a ‘good’ seed: HLT-10
DAP for taxonomy learning

• **Downward**: “*animals such as lions and ?*” gives:

  ...alligators ants bears bees camels cats cheetahs chickens crocodiles dachshunds dogs eagles lions llamas ... peacocks rats snails snakes spaniels sparrows spiders tigers turkeys varmints wasps wolves worms ...

• **Upward**: “*? such as lions and tigers*” gives:

  amphibians apes ... felines fish fishes food fowl game game_animals grazers grazing_animals grazing_mammals herbivores herd_animals household_pests household_pets house_pets humans hunters insectivores insects invertebrates laboratory_animals ... monogastrics non-ruminants pets pollinators poultry predators prey ... vertebrates water_animals wetlands zoo_animals
Experiment

• Seeds: *Animals + lions* and *People + Madonna* (seed term determines Basic Level or instance)

• Procedure:
  – Sent DAP and DAP\(^{-1}\) queries to Google
  – Collected 1000 snippets per query, kept only unique answers (counting freqs)
    (for DAP\(^{-1}\), extracted 2 words in target position)
  – Algorithm ran for 10 iterations

• Results: 1.1 GB of snippets for Animals and 1.5 GB for People:
  – 913 Animal basic-level concepts and 1,344 People instances with Out-Degree > 0
Results

• Found staggering variety of terms
• [Surprisingly,] many more classes than instances:

• Much more diverse than expected:
  – Probably useful: laboratory animals, forest dwellers, endangered species ... 
  – Useful?: bait, allergens, seafood, vectors, protein, pests ... 
  – What to do?: native animals, large mammals ... 

• Problem: How to evaluate this?
Evaluation woes: Precision

• Would like to evaluate against WordNet or Wikipedia (international standards, available, large, etc.)

• BUT:
  – They do not contain many of our learned terms (even though many are sensible and potentially valuable)
  – Point of our work is to learn more/new concepts than currently available

• Other projects use measures that don’t fit our work:
  – E.g.: Ritter et al. learn \{jaguar ISA: animal, mammal, toy, sports-team, car-make, operating-system\} and count all correct — even if not Animal

• Our strategy:
  – Count only correct classes
  – Compare against WordNet and do manual evaluation (if possible)
Evaluation woes: Recall

• Cannot easily compare to WordNet:
  – Doesn’t indicate Basic Level
  – Doesn’t include Instances (very few proper names)

• So, need to ask people ... this is expensive
Evaluation measures

• **Precision:**
  
  \[ \text{Pr}_{\text{WN}} = \frac{\text{#terms found in WordNet}}{\text{#terms harvested by system}} \]
  
  \[ \text{Pr}_{\text{HUM}} = \frac{\text{#terms judged correct by human}}{\text{#terms harvested by system}} \]

• **Recall substitute:**
  
  \[ \text{NotInWN} = \text{#terms judged correct by human but not in WordNet} \]
Evaluation #1: Basic terms and Instances

<table>
<thead>
<tr>
<th></th>
<th># harvested</th>
<th>$\text{Pr}_{\text{WN}}$</th>
<th>$\text{Pr}_{\text{HUM}}$</th>
<th>Not In WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>913</td>
<td>.79</td>
<td>.71</td>
<td>48</td>
</tr>
<tr>
<td>People</td>
<td>1344</td>
<td>.23</td>
<td>.95</td>
<td>986</td>
</tr>
</tbody>
</table>

Animals: Precision at Rank N

People: Precision at Rank N
Evaluation #2: ISA links

• Accuracy of algorithm on taxonomy links?

• Very expensive to consider all links
  – Need concept disambiguation in Wordnet
  – Need manual inspection of each term

• Consider only links from instance/basic level to immediate parent:

<table>
<thead>
<tr>
<th></th>
<th># harvested</th>
<th>Pr_{WN}</th>
<th>Pr_{HUM}</th>
<th>NotInWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>1940</td>
<td>.47</td>
<td>.88</td>
<td>804</td>
</tr>
<tr>
<td>People</td>
<td>908</td>
<td>.23</td>
<td>.94</td>
<td>539</td>
</tr>
</tbody>
</table>

WordNet lacks nearly half of the ISA links!
Is A above B?

• Categories:
  – 3 human judges; used web to check
  – *Category* = good answer; *Member* = inverse ISA; *Discard* = bad harvest
  – Very high pairwise Cohen kappas

• ISAs:
  – Randomly selected 120 each (animal and people) relations
    (100 from harvesting; 20 made at random to include some *False* answers)
  – 3 humans judges; asked if subcategory *always* / *sometimes* / *never* under supercategory
  – Averge pairwise Cohen kappa = 0.71 (animals) and 0.84 (people)
ISA relationship tests

• Concept Positioning Test:
  (apply DAP twice, inverting terms)
  Count freqs of terms generated by each term pair

• Concept Children Test:
  – Count intersections of terms generated by each term pair
Evaluation #3: Intermediate concepts

- Human Evaluation with four annotators

<table>
<thead>
<tr>
<th></th>
<th>Animals after CPT</th>
<th>People after CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correct</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>146</td>
<td>126</td>
</tr>
<tr>
<td>A2</td>
<td>133</td>
<td>126</td>
</tr>
<tr>
<td>A3</td>
<td>144</td>
<td>114</td>
</tr>
<tr>
<td>A4</td>
<td>141</td>
<td>116</td>
</tr>
<tr>
<td><strong>Borderline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>A2</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>A3</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>A4</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td><strong>BasicConcept</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>A2</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>A3</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>A4</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td><strong>NotConcept</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>28</td>
<td>7</td>
</tr>
<tr>
<td>A2</td>
<td>31</td>
<td>10</td>
</tr>
<tr>
<td>A3</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td>A4</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td><strong>Acc1</strong></td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Acc2</strong></td>
<td>0.84</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Acc1</strong></td>
<td>0.77</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Acc2</strong></td>
<td>0.82</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Acc1 = percentage Correct
Acc2 = percentage Correct or Borderline

- Comparison with WordNet

<table>
<thead>
<tr>
<th></th>
<th># harvested</th>
<th>Pr_{WN}</th>
<th>Pr_{HUM}</th>
<th>NotInWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>437</td>
<td>.20</td>
<td>.57</td>
<td>204</td>
</tr>
<tr>
<td>People</td>
<td>296</td>
<td>.51</td>
<td>.85</td>
<td>108</td>
</tr>
</tbody>
</table>
Effect of In-degree concept ranking

- **In-degree** measures popularity of concept
- Precision drops as In-degree drops:

![Graph of Animal Intermediate Concepts](image)

![Graph of People Intermediate Concepts](image)
Still...results are a bit of a mess

The problem?
Too many different kinds of categories
How to taxonomize?

• Goal: Group terms into small sets, then taxonomize

• Need to find groups / families of classes

  [predators prey]
  [carnivores herbivores omnivores]
  [pets wild_animals lab_animals ...]
  [water_animals land_animals ...]

• Consult online dictionaries, encyclopedias:
  – Some classes are defined by behaviors (such as eating), some by body structure, some by function ...
  – Try to define search patterns that capture salient aspects:
    “[carnivores/ herbivores/ omnivores] are animals that eat...”
    “[water_animals/ land_animals] are animals that live...”
    “[pets/ lab_animals/ zoo_animals] are animals that ?”
Parallel \textit{is-a} hierarchies

- First, created a small Upper Model manually:
  - BasicAnimal
  - GeneticAnimalClass
  - BehaviorClasses
  - MorphologicalTypeAnimal
  - RoleOrFunctionOfAnimal
  - BehaviorByFeeding
  - BehaviorByHabitat
  - BehaviorBySocialization

- Then, had 4 independent annotators choose appropriate Upper Model class(es) for several hundred harvested classes

- Kappa agreement for some classes ok, for others not so good
  - Sometimes quite difficult to determine what an animal term means
Evaluating concepts

• First checked whether learned intermediate concepts are correct
  – Manually created small taxonomy to begin to group terms
  – Also included categories for wrong and dubious terms

• Then checked for ISA taxonomization using CPT
1. **BasicAnimal**  
The **basic individual** animal. Can be visualized mentally. Examples: Dog, Snake, Hummingbird.

2. **GeneticAnimalClass**  
A **group** of basic animals, defined by **genetic similarity**. Cannot be visualized as a specific type. Examples: Reptile, Mammal. Note that sometimes a genetic class is also characterized by distinctive behavior, and so should be coded twice, as in Sea-mammal being both GeneticAnimalClass and BehavioralByHabitat. (Since genetic identity is so often expressed as body structure—it’s a rare case that two genetically distant things look the same structurally—it will be easy to confuse this class with MorphologicalTypeAnimal. If the term refers to just a portion of the animal, it’s probably a MorphologicalTypeAnimal. If you really see the meaning of the term as both genetic and structural, please code both.)

3. **NonRealAnimal**  
**Imaginary** animals. Examples: Dragon, Unicorn. (Does not include ‘normal’ animals in literature or films.)

4. **BehavioralByFeeding**  
A type of animal whose essential defining characteristic relates to a **feeding pattern** (either feeding itself, as for Predator or Grazer, or of another feeding on it, as for Prey). Cannot be visualized as an individual animal. Note that since a term like Hunter can refer to a human as well as an animal, it should not be classified as GeneralTerm.

5. **BehavioralByHabitat**  
A type of animal whose essential defining characteristic relates to its habitual or otherwise noteworthy **spatial location**. Cannot be visualized as an individual animal. (When a basic type also is characterized by its spatial home, as in South African gazelle, treat it just as a type of gazelle, i.e., a BasicAnimal. But a class, like South African mammals, belongs here.) Examples: Saltwater mammal, Desert animal. And since a creature’s structure is sometimes determined by its habitat, animals can appear as both; for example, South African ruminant is both a BehavioralByHabitat and a MorphologicalTypeAnimal.

6. **BehavioralBySocializationIndividual**  
A type of animal whose essential defining characteristic relates to its patterns of **interaction with other animals**, of the same or a different kind. Excludes patterns of feeding. May be visualized as an individual animal. Examples: Herding animal, Lone wolf. (Note that most animals have some characteristic behavior pattern. So use this category only if the term explicitly focuses on behavior.)
7. BehavioralBySocializationGroup
A natural **group of basic** animals, defined by interaction with other animals. Cannot be visualized as an individual animal. Examples: Herd, Pack.

8. MorphologicalTypeAnimal
A type of animal whose essential defining characteristic relates to its internal or external **physical structure** or appearance. Cannot be visualized as an individual animal. (When a basic type also is characterized by its structure, as in Duck-billed platypus, treat it just as a type of platypus, i.e., a BasicAnimal. But a class, like Armored dinosaurs, belongs here.) Examples: Cloven-hoofed animal, Short-hair breed. And since a creature’s structure is sometimes determined by its habitat, animals can appear as both; for example, South African ruminant is both a MorphologicalTypeAnimal and a BehavioralByHabitat. Finally, since genetic identity is so often expressed as structure—it’s a rare case that two genetically distant things look the same structurally—it will be easy to confuse this class with MorphologicalTypeAnimal. If the term refers to just a portion of the animal, it’s probably a MorphologicalTypeAnimal. But if you really see both meanings, please code both.

9. RoleOrFunctionOfAnimal
A type of animal whose essential defining characteristic relates to the **role or function** it plays with respect to others, typically humans. Cannot be visualized as an individual animal. Examples: Zoo animal, Pet, Parasite, Host.

G. GeneralTerm
A term that includes animals (or humans) but refers **also** to things that are neither animal nor human. Typically either a very general word such as Individual or Living being, or a general role or function such as Model or Catalyst. Note that in rare cases a term that refers mostly to animals also includes something else, such as the Venus Fly Trap plant, which is a carnivore. Please ignore such exceptional cases. But when a large proportion of the instances of a class are non-animal, then code it as GeneralTerm.

E. EvaluativeAnimalTerm
A term for an animal that carries an opinion judgment, such as “varmint”. Sometimes a term has two senses, one of which is just the animal, and the other is a human plus a connotation. For example, “snake” or “weasel” is either the animal proper or a human who is sneaky; “lamb” the animal proper or a person who is gentle, etc. Since the term can potentially carry a judgment connotation, please code it here as well as where it belongs.

A. OtherAnimal
Almost certainly an animal or human, but none of the above applies, or: “I simply don’t know enough about it”.
<table>
<thead>
<tr>
<th>Code</th>
<th>An1</th>
<th>An2</th>
<th>An3</th>
<th>An4</th>
<th>Ex.M</th>
<th>Par.M</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>BasicAnimal</td>
<td>29</td>
<td>24</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>12</td>
<td>0.51</td>
</tr>
<tr>
<td>BehavioralByFeeding</td>
<td>48</td>
<td>33</td>
<td>45</td>
<td>49</td>
<td>27</td>
<td>17</td>
<td>0.68</td>
</tr>
<tr>
<td>BehavioralByHabitat</td>
<td>85</td>
<td>58</td>
<td>56</td>
<td>54</td>
<td>36</td>
<td>36</td>
<td>0.66</td>
</tr>
<tr>
<td>BehavioralBySocializationGroup</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0.47</td>
</tr>
<tr>
<td>BehavioralBySocializationIndividual</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.46</td>
</tr>
<tr>
<td>EvaluativeTerm</td>
<td>41</td>
<td>14</td>
<td>10</td>
<td>29</td>
<td>6</td>
<td>19</td>
<td>0.51</td>
</tr>
<tr>
<td>GarbageTerm</td>
<td>21</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>12</td>
<td>3</td>
<td>0.74</td>
</tr>
<tr>
<td>GeneralTerm</td>
<td>83</td>
<td>72</td>
<td>64</td>
<td>79</td>
<td>19</td>
<td>72</td>
<td>0.52</td>
</tr>
<tr>
<td>GeneticAnimalClass</td>
<td>95</td>
<td>113</td>
<td>81</td>
<td>73</td>
<td>42</td>
<td>65</td>
<td>0.61</td>
</tr>
<tr>
<td>MorphologicalTypeAnimal</td>
<td>29</td>
<td>33</td>
<td>42</td>
<td>39</td>
<td>13</td>
<td>26</td>
<td>0.58</td>
</tr>
<tr>
<td>NonRealAnimal</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td>NotAnimal</td>
<td>81</td>
<td>97</td>
<td>82</td>
<td>85</td>
<td>53</td>
<td>40</td>
<td>0.68</td>
</tr>
<tr>
<td>OtherAnimal</td>
<td>34</td>
<td>41</td>
<td>20</td>
<td>6</td>
<td>1</td>
<td>24</td>
<td>0.47</td>
</tr>
<tr>
<td>RoleOrFunctionOfAnimal</td>
<td>89</td>
<td>74</td>
<td>76</td>
<td>47</td>
<td>28</td>
<td>56</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>641</td>
<td>578</td>
<td>511</td>
<td>488</td>
<td>239</td>
<td>375</td>
<td>0.57</td>
</tr>
</tbody>
</table>

**Human category judgments**

<table>
<thead>
<tr>
<th>Code</th>
<th>An1</th>
<th>An2</th>
<th>An3</th>
<th>An4</th>
<th>Ex.M</th>
<th>Par.M</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>BasicPerson</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0.55</td>
</tr>
<tr>
<td>FamilyRelation</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>0.86</td>
</tr>
<tr>
<td>GeneralTerm</td>
<td>38</td>
<td>12</td>
<td>21</td>
<td>12</td>
<td>4</td>
<td>18</td>
<td>0.50</td>
</tr>
<tr>
<td>GeneticPersonClass</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.44</td>
</tr>
<tr>
<td>ImaginaryPeople</td>
<td>14</td>
<td>16</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>0.47</td>
</tr>
<tr>
<td>NationOrTribe</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.78</td>
</tr>
<tr>
<td>NonTransientEventParticipant</td>
<td>29</td>
<td>63</td>
<td>41</td>
<td>32</td>
<td>16</td>
<td>33</td>
<td>0.57</td>
</tr>
<tr>
<td>NotPerson</td>
<td>31</td>
<td>31</td>
<td>28</td>
<td>38</td>
<td>24</td>
<td>9</td>
<td>0.80</td>
</tr>
<tr>
<td>OtherHuman</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td>PersonState</td>
<td>23</td>
<td>1</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>0.47</td>
</tr>
<tr>
<td>RealPeople</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.50</td>
</tr>
<tr>
<td>ReligiousAffiliation</td>
<td>10</td>
<td>16</td>
<td>12</td>
<td>15</td>
<td>5</td>
<td>11</td>
<td>0.61</td>
</tr>
<tr>
<td>SocialRole</td>
<td>62</td>
<td>61</td>
<td>39</td>
<td>44</td>
<td>25</td>
<td>36</td>
<td>0.61</td>
</tr>
<tr>
<td>TransientEventParticipant</td>
<td>30</td>
<td>27</td>
<td>13</td>
<td>7</td>
<td>2</td>
<td>17</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>257</td>
<td>256</td>
<td>197</td>
<td>164</td>
<td>85</td>
<td>150</td>
<td>0.58</td>
</tr>
</tbody>
</table>
More taxonomies... not so great

Another animal taxonomy:

Emotions—a disaster!
CONCLUSION
Today

Basically, we don’t today know
• How many concepts
• How many instances
• Which relations (or even how many)
• ...we can’t even properly agree about *is-a*

• Despite this we happily go and harvest terms and measure things like sampled precision and relative recall

• We publish massive collections like YAGO and NELL and TextRunner, but we don’t really use them well
Addendum: Other aspects we need to harvest

• Time of applicability of statement
  – “Bill Clinton is the President of the USA”

• Expected truth of fact
  – “They believe many things. For example, the world is flat”

• Trustworthiness of sources
THANK YOU