Towards Learning
World Knowledge
Suitable for Inference

Jonathan Gordon
University of Rochester
Motivation & Perspective
Commonsense Knowledge

Large-scale
High-quality
Suitable for Inference
Language & Knowledge

We should take language very seriously when thinking about how to represent commonsense knowledge.

*Montague’s Thesis:* Syntax mirrors semantics.

A language-like logic isn’t the only representation needed for common sense, but it’s an important one.
Semantic Resources of Language

**First-Order Logic:**

- Name entities and ascribe properties and relations to them.
- And, or, not, if-then
- Every, some, no
Semantic Resources of Language

**Episodic Logic** (Schubert & Hwang):

- **Generalized quantifiers** (‘most people who drive’)
- **Intensionality** (‘is planning a heist’)
- **Event Reference** (‘all the guests sampled each dessert; that led to a shortage.’)
- **Modification of predicates & sentences** (‘barely awake’)
- **Reification** (‘That there is poverty is unfortunate.’)
- **Uncertainty** (‘It will probably rain tomorrow.’)
Episodic Logic Example

Schubert & Hwang (2000):

A predatory animal attacks a nearby creature only when it is hungry or enraged.

\[
(\forall x: [x ((\text{attr predatory}) \text{ animal})] \ (\forall y: [y \text{ creature}]
\ (\forall e_1: [[y \text{ near } x] \leftrightarrow e_1]
\ (\forall e_2: [e_2 \text{ during } e_1]
\ [[x \text{ attack } y] \leftrightarrow e_2]
\rightarrow (\exists e_3: [e_3 \text{ same-time } e_2]
\ [[x \text{ hungry} \leftrightarrow e_3] \lor [[x \text{ enraged} \leftrightarrow e_3]])])
\]
Epilog

www.cs.rochester.edu/research/epilog
Schubert, Schaeffer, Morbini, et al., 1990 to present

The inference engine for Episodic Logic:
Supports forward and goal-directed inference.
EL is natural-logic-like;
Epilog can emulate the inferences of such systems.
Uses lexical and world knowledge
Approach
Knext is an initial approach to acquiring knowledge from text.

Open-domain Knowledge Extraction
Based on the interpretation and abstraction of logical forms created from semantic post-processing of syntactic parse trees.
Knext Architecture

- Adjust phrase structure for interpretation:
  - Identify temporal phrases, etc.

- Compute LFs:
  - Apply 80 interpretation rules.

- Extract & abstract propositions:
  - E.g., use gazetteers of names & entities

- Verbalize and filter propositions:
  - Abstract LFs & English output
‘Aristotle wrote many dialogues, only fragments of which survived.’
Aristotle wrote many dialogues, only fragments of which survived.

(S1 (S (NP (NNP Aristotle)))
   (VP (VBD wrote))
   (NP (NP (JJ many) (NNS dialogues)))
   (RB only)
   (NP (NP (NNS fragments)))
   (PP (IN of))
   (SBAR (WHNP (WDT which))
     (S (VP (VBD survived)))))
( . )))
Aristotle wrote many dialogues, only fragments of which survived.
‘A philosopher may write some dialogues’
Corpora & the Web

The Web offers a vast quantity of text with a breadth of topics, but it also presents the problem of dealing with casual, unedited writing, which can lead to low-quality knowledge.
Why not just use Wikipedia?

Diverse topics, written for the express purpose of conveying accurate information about the world.

But what we’re looking for is **implicit knowledge**.
Why implicit knowledge?

‘The emperor was succeeded by his son, Akihito.’

Intelligent behavior doesn’t require knowing this. We want:

   An emperor may be succeeded by a son.

   A male may have a son.
vs Weblogs

Looking for implicit knowledge, resources like Wikipedia aren’t necessarily more important.

There’s more writing on weblogs (and forums, etc.) than there is on Wikipedia, and they’re growing faster.
Knowledge Extraction on Web Corpora

**Knowledge-centric vs Unfocused Writing:**
General world knowledge learned from both is rated approx. the same.

**Disjoint Content:**
Weblogs and Wikipedia are not obvious replacements for each other – the overlap in resulting knowledge is quite small.
### Differences: Top Factoids

<table>
<thead>
<tr>
<th>Weblogs</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>A PERSON MAY THINK</td>
<td>A MALE MAY HAVE A WIFE</td>
</tr>
<tr>
<td>A PERSON MAY HAVE A LIFE</td>
<td>A MALE MAY HAVE A FATHER</td>
</tr>
<tr>
<td>A PERSON CAN BE SURE</td>
<td>A MALE MAY HAVE A CAREER</td>
</tr>
<tr>
<td>A PERSON MAY HAVE FRIENDS</td>
<td>A MALE MAY HAVE A DEATH</td>
</tr>
<tr>
<td>A PERSON MAY HAVE A FRIEND</td>
<td>A MALE MAY HAVE A WORK</td>
</tr>
<tr>
<td>A PERSON MAY HAVE A HEAD</td>
<td>A MALE MAY HAVE A SON</td>
</tr>
<tr>
<td>A PERSON MAY HAVE A MIND</td>
<td>A MALE MAY HAVE A LIFE</td>
</tr>
<tr>
<td>A PERSON MAY KNOW A PROPOSITION</td>
<td>A MALE MAY HAVE A MOTHER</td>
</tr>
<tr>
<td>A MALE MAY HAVE EYES</td>
<td>A MALE MAY HAVE A BROTHER</td>
</tr>
<tr>
<td>A MALE MAY HAVE A HEAD</td>
<td>A MALE MAY HAVE A FAMILY</td>
</tr>
</tbody>
</table>
Do we really need to read the whole web?

The number of unique factoids learned from weblogs & Wikipedia shows *little sign of leveling off* after hundreds of millions of sentences.

However:
There’s a *declining utility* as new knowledge becomes more obscure, but as we seek to abstract from individual factoids to more general truths, more data is still useful.
Sharpening
Why ‘Sharpening’?

KNEXT has learned hundreds of millions of unique factoids from text.

We want to automatically sharpen the best ones into quantified general statements suitable for inference.

Previous work:
Van Durme, Michalak, & Schubert on using WordNet to abstract factoids into conditional knowledge, e.g., if a person writes something, it is probably a communication.
Weak Knowledge

KNEXT Input:
‘She had a cup of tea.’

KNEXT Output:
[<det female.n> have.v (:a <a{n} cup.n> (of.p (K tea.n)))]
‘A female may have a cup of tea.’
Sharpened Formula

**KNEXT Input:**
‘She had a cup of tea.’

**KNEXT Output:**
[<det female.n> have.v (:a <a{n} cup.n> (of.p (K tea.n)))]
‘A female may have a cup of tea.’

**Sharpened Output:**
(many x: [x female.n]
  (occasional e
   (some y: [y (cup-of.n tea.n)]
    [[x drink.v y] ** e]])))
Method

Apply handwritten rules to sharpen large classes of factoids that are likely to deserve a strong form.
Semantic Patterns

The unsharpened formula

\(((\text{det person.n}) \ \text{have.v} \ \text{det sister.n})\)

is matched by the pattern:

\(((\text{1}\_\text{det?} \ \text{2}\_\text{agent?}) \ \text{have.v} \ (\text{3}\_\text{det?} \ \text{4}\_\text{role?}))\)
Output Transformation

The unsharpened formula

\(((\text{det} \ \text{person.n}) \ \text{have.v} \ (\text{det} \ \text{sister.n}))\)

is matched by the pattern:

\(((1\_\text{det}? \ 2\_\text{agent}?) \ \text{have.v} \ (3\_\text{det}? \ 4\_\text{role}?)\))

With output form:

\((\text{many} \ x: [x \ 2\_] \\
\text{(some} \ e \\
\text{(some} \ y: [y \ (\text{role-cat!} \ 4\_)] \\
\text{[[x \ (has-as} \ 4\_) \ y] \ ** \ e]]))))\)
Output Transformation

The unsharpened formula
((det person.n) have.v (det sister.n))

Sharpened Form:
(all-or-most x: [x person.n]
  (some e
    (some y: [y female.n]
      [[x (have-as sister.n) y] ** e]))
Semantic Predicates

Sources of lexical semantic classes include:
  WordNet
  VerbNet
  Propbank
  Some hand enumeration

But the sharpened formulas could not be generated from those resources alone.
Individual-Level Predicates

**Individual-level** predicates endure over most of the existence of the individual.

Most **stative verbs** are individual-level though it can depend on the arguments:

- Books contain information *for the duration of their existence*.
- Baskets contain objects *for a short while*.
Individual-Level Example

**Input:**
‘She proved a quick, intelligent scholar...’

**Unsharpened:**
\[
<\text{det scholar.n}> \text{ intelligent.a}
\]

**Sharpened:**
\[
(\text{many-or-some } x: [x \text{ scholar.n}]
\quad (\text{some } e [[x \ | \ e] \text{ enduring}]
\quad [[x \text{ intelligent.a}] \ast \ast e]))
\]
Stage-Level Predicate

**Stage-level** predicates describe transient situations. If an entity has a capacity, such as being able to walk, we assume it is at least occasionally exercised.
Stage-Level Example

**Input:**
‘He said Mitchell is against the centralization...’

**Unsharpened:**

\[
\text{[<det male.n> say.v <det proposition.n>]}\]

**Sharpened:**

\[
\text{(all-or-most } x: [x \text{ male.n] (occasional e (some y: [y proposition.n] [[x say.v y] ** e]]))}
\]
Non-repeatable Stage-Level Predicates

Some predicates are repeatable, while others are one-time-only (either for the subject or the object):

A person does not occasionally *die*.  
A person can occasionally *kill*, but cannot **be** killed occasionally.

We use VerbNet classes to find **non-repeatable** or **pivotal** verbs.
Generic Sentences

‘Dogs are furry.’

[(K (plur dog.n)) furry.a]

(many-or-some x: [x dog.n]
  (some e: [[x | e] enduring]
    [[x furry.a] ** e]))

‘Dogs bark.’

[(K (plur dog.n)) bark.v]

(many x: [x dog.n]
  (at-least-occasional e [[x bark.v] ** e]))
Sharpening Summary

We can sharpen classes of factoids expressing commonsense knowledge from weak, possibilistic claims to stronger, quantified claims.

We use semantic categories and extraction frequencies in writing a small number of sharpening rules with broad coverage.

In sharpening, we implicitly disambiguate many predicates based on their arguments.
Reasoning
Inference

ALL GOLDFISH ARE PETS
ALL-OR-MOST PETS ARE HARMLESS

→

(probably ALL-OR-MOST GOLDFISH ARE HARMLESS)

(in the absence of other information)
Inference

Given

[John.name person.n]

We want to generate inferences like

(probably (some x: [x head.n] [John.name has-as-part.v x]))
Live

Knowledge Browser:  
www.cs.rochester.edu/research/knext/browse

Public release of basic KNEXT code and data:  
www.cs.rochester.edu/research/knext
Questions?

www.cs.rochester.edu/research/knext