A Study of Automatically Acquiring Explanatory Inference Patterns from Corpora of Explanations: Lessons from Elementary Science Exams

Peter Jansen, School of Information, University of Arizona
Our long term interest is in building **inference algorithms** capable of answering **questions** and producing **human-readable explanations** by aggregating **information from multiple sources** and knowledge bases.
Our long term interest is in building **inference algorithms** capable of **answering questions** and producing **human-readable explanations** by **aggregating information from multiple sources** and knowledge bases.

Standardized elementary science exam questions require an average of 4 to 6 facts (range 1-16) to answer and explain their reasoning. (Jansen et al., COLING 2016; Jansen et al., submitted to LREC 2018)

Assembling long chains of facts to answer questions is challenging. Inference algorithms have difficulty aggregating more than 2 pieces of information, and quickly drift off-topic due to “semantic drift”. (Fried et al., TACL 2015; Khashabi et al., IJCAI 2016; Jansen et al., CL 2017)

Some new method of controlling for semantic drift to assemble long inference chains is required for complex explainable QA

This paper investigates whether extracting large “common inference patterns” from corpora of explanations is a viable method.
Main Questions

Are there “explanatory patterns” in corpora of explanations that are commonly reused?
(i.e. 2 or more facts that are commonly seen together in different explanations)

If explanatory patterns exist, is it possible to abstract them in some way to make them more general?
(i.e. specific kinds of knowledge, instead of specific instances of knowledge, that tend to be seen together)

How much of explanations for unseen questions is it possible to reconstruct just by merging, adapting, or adding to these explanatory patterns?
(i.e. can we use common explanatory patterns to reduce the need for “raw” information aggregation, and start to build large inferences with 10 or more facts for QA)
Part 1:

Question answering as building explanation graphs
THE BEST AI STILL FLUNKS 8TH GRADE SCIENCE

In 2012, IBM Watson went to medical school. So said The New York Times, announcing that the tech giant’s artificially intelligent question-and-answer machine had begun a “stint as a medical student” at the Cleveland Clinic Lerner College of Medicine.

This was just a metaphor. Clinicians were helping IBM train Watson for use in medical research. But as metaphors go, it wasn’t a very good one. Three years later, our artificially intelligent machines can’t even pass an eighth-grade science test, much less go to medical school.

The top performers successfully answered about 60 percent of the questions. In other words, they flunked.
Q: Which of the following is an example of an organism taking in nutrients?
(A) A dog burying a bone  (C) An insect crawling on a leaf
(B) A girl eating an apple  (D) A boy planting tomatoes
### Science Exam Example Question

**Q:** Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone  
(B) A girl eating an apple  
(C) An insect crawling on a leaf  
(D) A boy planting tomatoes

*Rarely* will we be able to retrieve a single passage in a corpus that directly answers a given question:

“*A girl eating an apple is an example of an organism taking in nutrients...*”
Q: Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone  (C) An insect crawling on a leaf
(B) A girl eating an apple  (D) A boy planting tomatoes

- **Girl**
  - “a girl means a human girl”
  - “humans are living organisms”
  - Simple Wiktionary

- **Eating**
  - “eating is when an organism takes in nutrients in the form of food”
  - 4th Grade Study Guide

- **Apple**
  - “an apple is a kind of fruit”
  - “fruits are foods”
  - Simple Wiktionary
Q: Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone  
(B) A girl eating an apple  
(C) An insect crawling on a leaf  
(D) A boy planting tomatoes

- Girl: "a girl means a human girl"  
  "humans are living organisms"

- Eating: "eating is when an organism takes in nutrients in the form of food"

- Apple: "an apple is a kind of fruit"  
  "fruits are foods"

Simple Wiktionary  
4th Grade Study Guide  
Simple Wiktionary
WorldTree Corpus: 1,680 Standardized Science Questions paired with detailed, manually authored, and lexically-connected explanation graphs.

Question: Which characteristic would best help a tree survive the heat of a forest fire?


Domain Expert (e.g. teacher)
Bark is a protective covering around the trunk and branches of a tree.

Domain Novice (e.g. student)
As an object’s thickness increases, it’s resistance to damage will also increase.

Young Child (e.g. 5-year old)
Protecting something means preventing harm.
Fire causes harm to trees, forests, and other living things.
Thickness is a measure of how thick an object is.
A tree is a kind of living thing.

First Principles
Protecting a living thing has a positive impact on it’s survival and health.
Nodes represent questions
Edges represent 2+ shared sentences/facts in explanations
Explanations are represented as one or more rows in 62 semi-structured tables, providing both coarse (sentence-level) and fine-grained (table column) explanation graph structure.

**Process Roles Table**

<table>
<thead>
<tr>
<th>PROCESS NAME</th>
<th>ACTOR</th>
<th>ROLE</th>
<th>ACTION</th>
<th>PATIENT</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the food chain</td>
<td>process, a</td>
<td>green plant</td>
<td>has the role of</td>
<td>producer</td>
<td>which creates</td>
</tr>
<tr>
<td>In the food chain</td>
<td>process, an</td>
<td>animal</td>
<td>has the role of</td>
<td>consumer</td>
<td>which eats</td>
</tr>
<tr>
<td>In the food chain</td>
<td>process, a</td>
<td>bacteria</td>
<td>has the role of</td>
<td>decomposer</td>
<td>which recycles</td>
</tr>
<tr>
<td>In the tree reproduction</td>
<td>process, a</td>
<td>squirrel</td>
<td>has the role of</td>
<td>seed disperser</td>
<td>which relocates</td>
</tr>
</tbody>
</table>

**Taxonomy Table**

<table>
<thead>
<tr>
<th>HYPONYM</th>
<th>SCOPE</th>
<th>HYPERNYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>deer</td>
<td>is a kind of</td>
</tr>
<tr>
<td>A</td>
<td>green</td>
<td>is a kind of</td>
</tr>
<tr>
<td>A</td>
<td>shelter</td>
<td>is a kind of</td>
</tr>
<tr>
<td>An</td>
<td>electromagnet</td>
<td>is a kind of</td>
</tr>
</tbody>
</table>

**PartOf Table**

<table>
<thead>
<tr>
<th>PART</th>
<th>WHOLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A leaf</td>
<td>green plant</td>
</tr>
<tr>
<td>roots are a part of</td>
<td>plant</td>
</tr>
<tr>
<td>pedals are a part of</td>
<td>bicycle</td>
</tr>
<tr>
<td>A cell wall</td>
<td>plant cell</td>
</tr>
</tbody>
</table>

**Answer Candidates:**

- [ ] A cat eats a mouse
- [ ] a deer eats a leaf
- [ ] a hawk eats a mouse
- [ ] a snake eats a rat

[ *B ] a deer eats a leaf
Part 2:

Identifying and reusing patterns in corpora of explanations
**Approach**

**Question**
What property of **water** can make a **rock break apart**?

**Answer**
freezing and thawing

**Step 1**
Search for explanations in the training set that are similar to the question.

**Step 2**
Merge training explanations.

**Step 3**
Adapt knowledge to new question.

**Step 4**
Add new knowledge.

**Step 5**
Evaluate accuracy of assembled explanation with known good explanation in test set.

Gold Explanation

one node = one explanation sentence
What property of water can make a rock break apart?

Answer: freezing and thawing.
# Abstracting Patterns by Abstracting Edges

<table>
<thead>
<tr>
<th>Gold Edge</th>
<th>water is a kind of <strong>liquid</strong> (KINDOF TABLE)</th>
<th>freezing means changing from a <strong>liquid</strong> to a solid (CHANGE TABLE)</th>
</tr>
</thead>
</table>

## Edge Overlap Rating

### HIGH

#### Identical Edge

<table>
<thead>
<tr>
<th>water is a kind of <strong>liquid</strong> (KINDOF TABLE)</th>
<th>freezing means changing from a <strong>liquid</strong> to a solid (CHANGE TABLE)</th>
</tr>
</thead>
</table>

#### Table, Column, & Lexical Overlap

<table>
<thead>
<tr>
<th>soup is a kind of <strong>liquid</strong> (KINDOF TABLE)</th>
<th>boiling means changing from a <strong>liquid</strong> to a gas (CHANGE TABLE)</th>
</tr>
</thead>
</table>

#### Table & Column

<table>
<thead>
<tr>
<th>a chick is a kind of <strong>bird</strong> (KINDOF TABLE)</th>
<th>the life cycle is when animals grow from <strong>babies</strong> into adults (CHANGE TABLE)</th>
</tr>
</thead>
</table>

#### Table Only

<table>
<thead>
<tr>
<th>a toaster is a kind of <strong>electrical device</strong> (KINDOF TABLE)</th>
<th><strong>electrical devices</strong> convert electricity into other forms of energy (CHANGE TABLE)</th>
</tr>
</thead>
</table>

#### Zero Overlap

| a nucleus is a part of a **cell** (PARTOF TABLE) | the **cell** nucleus controls many functions of a cell (ACTIONS TABLE) |
## Example Common Edges

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Abs. Level</th>
<th>Pattern</th>
</tr>
</thead>
</table>
| 11 / 800 | Row | **water is a kind of liquid** ←→ **boiling means changing from a liquid to a gas**  
(KINDOF TABLE)  
(CHANGE TABLE) |
| 5 / 800 | Table, Column, & Lexical Overlap | **KINDOF TABLE**  
**“HYponym” COLUMN**  
**“PLANT”** ←→ **PARTOF TABLE**  
**“WHOLE” COLUMN**  
**a tree is a kind of plant** ←→ **a leaf is a part of most plants** |
| 199 / 800 | Table & Column | **KINDOF TABLE**  
**“HYponym” COLUMN** ←→ **KINDOF TABLE**  
**“HYpernym” COLUMN**  
**a dog is a kind of animal** ←→ **an animal is a kind of organism** |
| 80 / 800 | Table Only | **KINDOF TABLE** ←→ **USEDFOR TABLE**  
**a magnifying glass is a kind of tool** ←→ **a magnifying glass is used to see small objects by making them appear larger** |
Reconstruction Performance

If we use only “explanatory patterns” from other questions, how close can we come to reconstructing unseen explanations for novel questions?

<table>
<thead>
<tr>
<th>1 Subgraph</th>
<th>Additions (Missing Edges)</th>
<th>Adaptations (Edges similar to gold edge)</th>
<th>Matching Facts (Identical Edges)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29%</td>
<td>20% 16% 12%</td>
<td>23%</td>
</tr>
<tr>
<td>2 Merged Subgraphs</td>
<td>16%</td>
<td>18% 21% 16%</td>
<td>30%</td>
</tr>
<tr>
<td>3 Merged Subgraphs</td>
<td>12%</td>
<td>14% 23% 19%</td>
<td>31%</td>
</tr>
</tbody>
</table>

**Interpretation:** for a hypothetical explanation graph with 9 edges:
- 3 edges would be identical
- 4.5 edges would require varying degrees of adaptation
- 1.5 edges would require “raw” information aggregation

(Avg. of graphs w/1-9 edges)
Example Pattern (Short and Highly Abstracted)

1 Subgraph

<table>
<thead>
<tr>
<th>Additions (Missing Edges)</th>
<th>Adaptations (Edges similar to gold edge)</th>
<th>Matching Facts (Identical Edges)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29%</td>
<td>20% 16% 12%</td>
<td>23% (Avg. of graphs w/1-9 edges)</td>
</tr>
</tbody>
</table>

Question: Wendy wanted to find out if faster wind speeds increased the amount of wind erosion. Which instrument should she use to measure wind speed?


Gold Explanation:
wind speed is a property of weather (PROPERTIES table)
an anemometer is used to measure wind speed (USEDFOR table)
an anemometer is a kind of instrument (KINDOF table)

Similar Explanatory Pattern (from a different, known question):
air pressure is a property of air/the atmosphere (PROPERTIES table)
a barometer is used to measure air pressure (USEDFOR table)
a barometer is a kind of instrument (KINDOF table)

Example common 2-edge pattern at “Table & Column” level of abstraction
Little Data (Aggregated) is Bigger than the Biggest Data

We mapped out the knowledge required to answer and explain 2,200 3rd to 5th grade standardized elementary science questions.

We found that only 4,000 facts are required to answer and explain these exam questions, as long as they can be aggregated/combined into short explanations with an average of 6 facts/sentences. (Jansen et al., submitted to LREC 2018)

In contrast, systems with very large knowledge bases (~1,000,000 sentences) are answering ~60% of standardized questions.

Why can these questions be completely answered and explained with 200X less data? Information Aggregation vs Passage Retrieval

**Information Aggregation:** 4,000 facts combined into 6-fact explanations allows for ~4,096,000,000,000,000,000,000 possibilities

Empirically estimate that 2x10-8, or 133 trillion, would make good explanations (Jansen, submitted to LREC 2018)

133 trillion (little data aggregated) >> 1 million (big data)
Thank You!

Download common explanatory patterns at:
cognitiveai.org/explanations