Notes for Tagging Inference Knowledge Type
in Elementary Science Exams

Peter Jansen

March 28th, 2016

1 Introduction

This is a quick reference for annotating knowledge base requirements for explanably answering elementary science exams, based on the short paper *"A study of the Knowledge Base Requirements for Passing an Elementary Science Test"* by Peter Clark, Phil Harrison, and Niranjan Balasubramanian (AKBC 2013). This paper contains many detailed explanations and examples for the annotation process.

Below is a brief summary of the knowledge types, and common ambiguities or difficulties in the annotation process.

1.1 Framework and Knowledge Types

Clark et. al, *"A study of the Knowledge Base Requirements for Passing an Elementary Science Test" (AKBC 2013)* examined 50 questions from the 4th grade New York Regents science exam, and identified seven different categories of questions and supporting knowledge required to answer them correctly.

1. **Retrieval-based methods:** Taxonomic (is-a), Property (part-of, has, etc.), Definition

2. **General Inference methods:** Causality, Processes, Examples of situations

3. **Domain-specific models:** Specific domain representations and processes.

1.2 Examples

1. **Taxonomic (is-a)**

   ```
   X, Y, Z are forms of ...
   X is a kind of ...
   ```

2. **Property (part-of, has, etc...)**

   ```
   Which part of a plant produces the seeds?
   ```

3. **Definition**

   ```
   <text> is called ...
   Movement of soil by wind and water is called ...
   What instrument would be used to atmospheric pressure? (barometer)
   ```

4. **Causal**

   ```
   One way to change water from a solid to a liquid is ... (adding heat)
   ```

5. **Simple Processes**

   ```
   One way animals respond to a sudden change in temperature is... (shivering)
   ```

6. **Examples of Situations**

   ```
   Which object is the best conductor of electricity?
   Which of the following describes an organism taking in nutrients?
   ```

7. **Domain Models**

```
Does X go up when Y goes down?   (Relationships)
If a baby shakes a rattle, it makes a nosie.   Which form of energy was changed
   to sound energy?
What is the ROLE of ENTITY during PROCESS?
What performs ROLE during EVENT?
During X, what happens after Y?
```

2   Data

The notes in this annotation guide were devloped while extending the AKBC'13 procedure to annotate the 432 3rd, 4th, and 5th grade elementary science questions in the training set of the "AI2 Elementary School Science Questions (No Diagrams" dataset available from `http://www.allenai.org/data.html` .

The distribution of questions with a given label is included in the figure below. Approximately 23% of questions fall into the retrieval-based categories, while the remaining 77% are questions requiring complex inference. The most-frequent category of questions (39%) appear to require model-based reasoning to explainably answer.
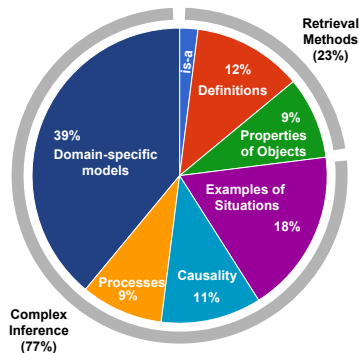
3   Annotation Challenges

Some of the questions clearly fall into one of the 7 categories, but often challenging questions can convincingly span more than one category. Some of these issues are outlined in the "Ambiguities" section, below. The general summary is:

- **Causal and process:** There are clear cases of primarily causal or primarily process questions, but often the line is very blurry between these.

- **Model vs causal and process:** Model questions often contain causal and process elements.

- **Multistep problems:** Many (but not all) questions are multi-step, and require integrating (for example) causal knowledge with taxonomic and definitional knowledge, or other such examples that span categories.

- **Property vs other:** Sometimes it's not clear whether a question is testing a commonly-known property (e.g. an inherited characteristic), or whether this knowledge might be more complicated and said to be a domain model. This is discussed further in the ambiguities section.

4   Ambiguities

Clark et al. note the difficulties associated with knowledge type annotation for the 50 New York Regents Questions. I have endeavored to use the following guidelines to achieve consistency in ambiguous situations when annotating

the 432 training questions from 12 states in the AI2 Elementary Science Questions dataset:

4.1 Property or Definition?

For retrieval questions, sometimes the line between property knowledge and definitional knowledge is blurry. **Part-of** relations are the cannonical property relations – e.g. `Pollen is a part of a flower`, and seemingly easier questions sometimes include explicit lexical cues for this (e.g. `What part of a plant contains pollen?`).

For properties that are not `Part-of`, but are common simple categories that are part of the curriculum (e.g. inherited/learned trait, renewable/non-renewable energy source), I tend to included these as `PROPERTIES`. For example `Which of the following is a renewable resource?  (wind)`.

For anything deeper, especially quesitons that include or require either (i) multiple pieces of information, or (ii) functions/purposes, I usually annotate these as `DEFINITION`. For example:

```
Which units should be used to measure time/distance/temperature/etc
Which instrument should be used to measure volume/temperature/distance/etc...
A simple machine that helps a flag move up a flagpole is: (a pulley)
```

4.1.1 Definition, Taxomonic, or Example?

Similarly, one might argue that the last example (flagpole) may tend to mix a `DEFINITION` with an `EXAMPLE`. Unfortunately the knowledge category choice here is even more complex – a *simple machine* is a kind of coarse category including machines such as levers, inclined planes, wheels, screws, and pulleys, potentially making this a `TAXONOMIC` question, depending on the other answer choices.

Many definitions include taxonomic and example knowledge as part of their definition. Questions with the `EXAMPLE` knowledge type tend to be specifically crafted to test (and require) non-trivial inference, rather than being able to use look-up terms common in definitional knowledge. In addition,

- The words "simple machine" and the flagpole example are commonly included in definitions targetted at elementary students (e.g. from knowledge resources such as study guides), and

- Using the "is called" test (i.e. adding the phrase "is called" at the end of the question) doesn't change the questions meaning

So this appears to be best suited as a definition question – that, like many definitions, includes taxonomic and example knowledge as part of the definition.

4.2 Causal or Process?

While one can imagine cannonical examples of primarily causal knowledge, or primarily process knowledge, often times these appear to blend together, and the categorization is less clear.

4.2.1 Causal Knowledge

Cannonical causal knowledge might appear as a single relational link, e.g.:

$$Friction --causes--> Heat$$

A somewhat more complicated example might be a generic change from one state to another through the introduction of some vehicle or gatekeeper, such as:

```
            A solid can change to a liquid by increasing it's temperature,
```

which might be simply described as:

```
                    Solid --heated--> liquid.
```

These gatekeepers or conditions might be more complex or further specified, such as:

```
            Solid --(heated above melting point)--> liquid).
```

4.2.2 Processes

Cannonical processes are things that occur in a stage-like fashion. The life cycle, water cycle, recycling process, erosion, and so forth are additional examples (some of which are continuous systems, repeating systems).

Middle ground: Questions that explicitly test the outcome from a change (causality), as well as testing particular, easily-separable aspects of processes, stages, or cycles (processes) are common. Here are a few somewhat more ambiguous cases:

- `Which statement can explain why the Sun can warm up water in a glass?`
  `(Water absorbs light energy).`
  I tag this as a `CAUSAL` question – Objects absorb light (a change of state). Water is a kind of object (taxonomic). Absorbing energy causes objects to increase in temperature (a change in state). The question seems to be more testing adding two causal relations together, rather than a knowledge of a broader process – though one could argue that "objects heating up from absorbing light" is itself an abstract multistep process.

- `Which of the following is produced when sugar is digested in an animal cell?`
  `(Carbon Dioxiode).`
  I tag this as a `PROCESS` question – Cells convert oxygen+sugar into energy+waste, which one could argue is a direct causal relation, but it's two steps of the larger process of cell intake and outtake, which is a process.

Other things to note: Ultimately, we may end up blending these together, or at least using similar problem solving methods to solve the questions. The distinction between a single causal relation and multiple relations chained into an interacting process may be artificial, as one can certainly construct a process using a number of interconnected causal relations, likely with some additional structural information. Similarly, a single stage of some processes could be seen as a direct cause-effect relation. The causal/process knowledge categories may represent a continuum of the number of stages, or simply convention (i.e. we naturally group some collections of causal relations into processes, but not others).

Best judgement: Ultimately, for ambiguous cause/process questions, use your best judgement. In cases where there are clearly aspects of both, I usually assign two labels (e.g. `CAUSAL,PROCESS`), where the first (primary) label is the one that I believe the question most strongly represents.

4.3 Model or Causal/Process?

From Clark et al., *"Some questions require a specific modeling paradigm (i.e. way of thinking about the world) to answer, beyond just a sea of rules. One of the roles of teachers is to teach students such modeling techniques. In other words, certain classes of questions are answerable by a computation over a certain style of representation."*

Model questions almost always involve causality and process relations. Examples of model questions include:

```
Relations (does X go up when Y goes down?)
What is the ROLE of ENTITY during PROCESS?
What performs ROLE during EVENT?
During X, what happens after EVENT?
Spatial or Mathematical Reasoning
```

```
Energy conversions
```

Examples from the AI2 Elementary set:

- `Which best describes the role of the esophagus in digestion?`
  `(A: It caries food from the mouth to the stomach).`
  Fits the ROLE of ENTITY during PROCESS schema for a model-based question. Here, rather than simply detailing the stages of digestion (eat -> digest -> absorb -> excrete waste), we're testing the student's mental model of the digestion process, which actors, entities, and roles are at a given starting condition, and what interactions they complete to move to the next stage in the process.

- `What will most likely result when the jet stream moves south of North Carolina?`
  `(A: North Carolina will experience cold weather)`
  i.e. Given a model of weather movements, what would happen if we change the starting conditions to X?

- `In a grassland ecosystem, if the population of eagles suddenly decreased, what would the most likely effect on the rest of the ecosystem be?`
  `(A: The ecosystem will become overpopulated with snakes).`
  A relational question (does X go up when Y goes down?), instantiated in an ecosystem model where one must know that an eagle is a kind of predator, and that snakes are a kind of prey for eagles.

- `Which of these statements best explains why fossils of clam shells were found in the Appalacian Mountains?`
  `(A: The mountain range was once under water)`
  Requires a model with knowledge that fossils are dead animals, that if fossils were found in a place, that animal used to live in that place, and also that if an animal used to live in that place, then it's environment/conditions for living also used to exist in that area.

5 Ambiguities (smaller)

Other, smaller ambiguities:

5.1 Lists

Ocassionally a question or answer includes more than one item that the test taker has to evaluate. For example:

`Sleet, rain, hail, and snow are all forms of: (Precipitation)`

In these cases, I generally treat lists as orthogonal to knowledge type, and tag knowledge type were this to be four separeate questions (Sleet is a form of... rain is form of... etc). In the above case, I would annotate the above question as requiring/testing taxonomic (is-a) knowledge. In uncommon cases, the items in the list are integrally linked, describing the stages of a process. For example:

`The stages of a bird's life cycle are: (egg -> hatchling -> adult bird)`

In the above case, I would tag this example as a process question.

5.2 Other misc.

Other annotation notes:

1. There are many questions about **inherited characteristics**. (e.g. `Which of the following is an inherited characteristic or trait? (blue eyes)`). For questions testing whether a characteristic is inherited, I've generally considered whether a characteristic (e.g. eye color) is inherited to be a property of that concept

to an elementary-aged student, and tagged these as `PROPERTY` questions. There may be a case here to call these `EXAMPLE` questions. Conversely, more complex questions that ask how traits get passed from parent to offspring often appear to test more complex knowledge types depending on the specific question (e.g. `CAUSAL, PROCESS, MODEL...`).

6   Comments, questions, additions

This document is maintained by Peter Jansen (`pajansen@email.arizona.edu`). Please feel free to get in touch with comments or questions.