
Answering Elementary Science Questions by Constructing Coherent Scenes using Background Knowledge

Yang Li and Peter Clark



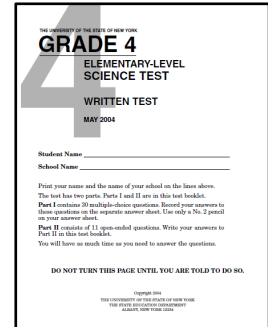
UC Santa Barbara



ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

Introduction: Project Aristo at AI2

- A first step towards “machine intelligence”
 - Gain knowledge of elementary science
 - Pass science exams
 - currently 4th grade multiple choice



28 When a baby shakes a rattle, it makes a noise. Which form of energy was changed to sound energy?

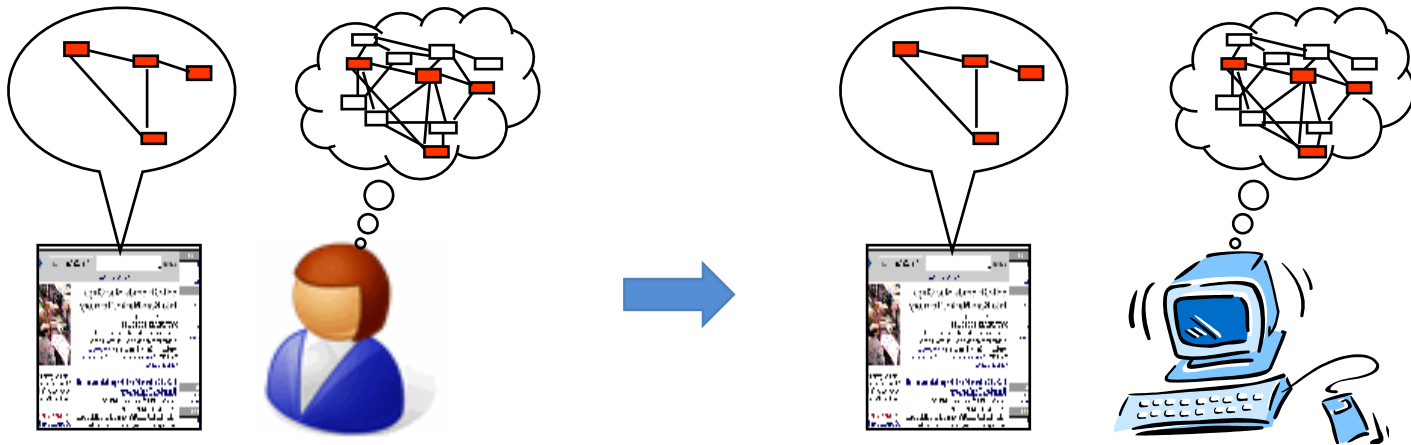
- A electrical
- B light
- C mechanical
- D heat



Information Retrieval fails for such questions!

Intuition

- Not all knowledge is explicitly stated in text
- People use background knowledge to fill in implicit info
- Can we make machines mimic the same process to better answer questions?



Our Insight

- Prior approaches have had limited success due to the lack of background knowledge to fill in implicit info
- Our insight: If we use a simple "knowledge graph" representation, then we can leverage existing large-scale linguistic resources to supply the implicit knowledge



relation tuples



knowledge graph

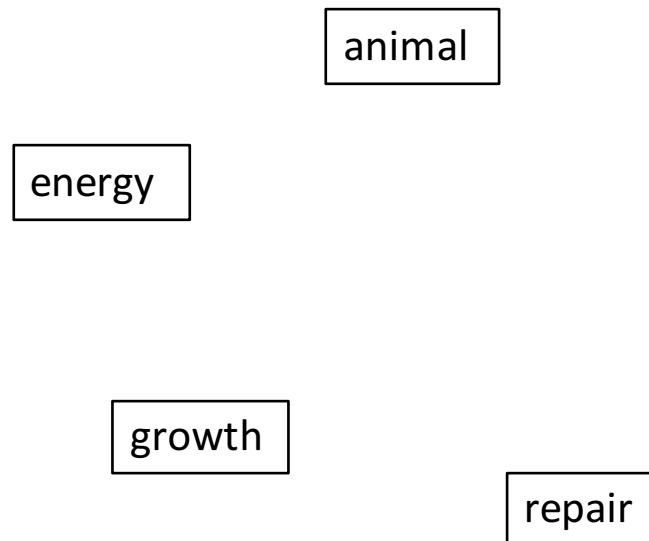
Problem Statement

- **Input:**
 - A multiple-choice question
 - Background knowledge base(s)
 - contain plausible relationships $R(x,y)$ between concepts x and y
 - e.g. WordNet, FreeAssociation, DART
 - A metric to measure the relatedness between concepts
- **Output:**
 - Main: the answer to the question
 - Side: the coherent scene
 - a dense subgraph containing keywords in question, the correct answer option and some necessary implicit concepts.
 - facilitate interpretation of the question/answer pair

Illustration of the Idea

- For example (idealized):

Question: Animals get energy for growth and repair from (A) food (B) air



Initial
representation
of question

Illustration of the Idea

- For example (idealized):

Question: Animals get energy for growth and repair from (A) food (B) air

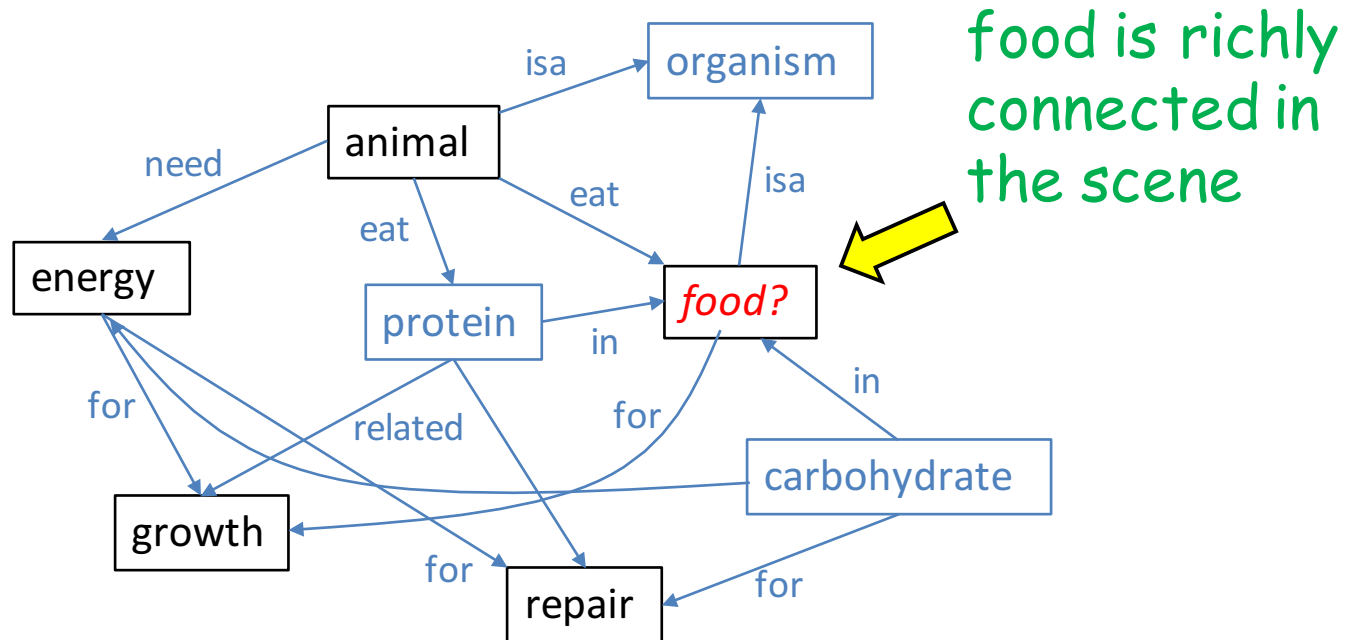


Illustration of the Idea

- For example (idealized):

Question: Animals get energy for growth and repair from (A) food (B) air

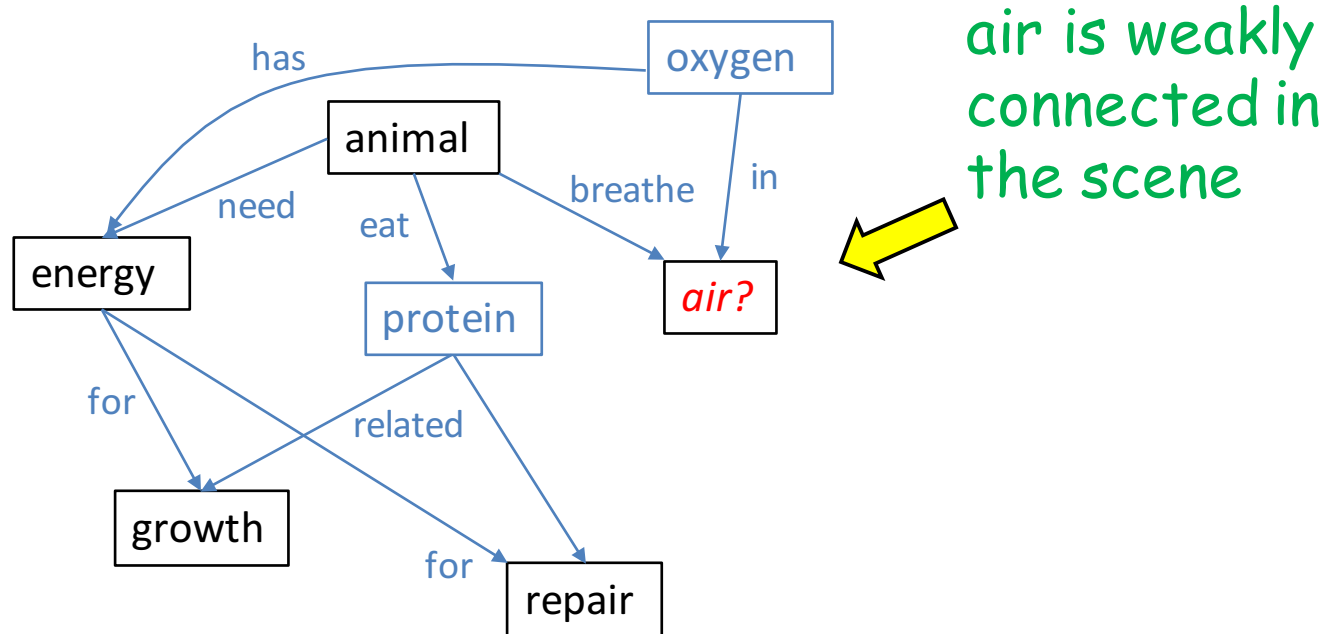
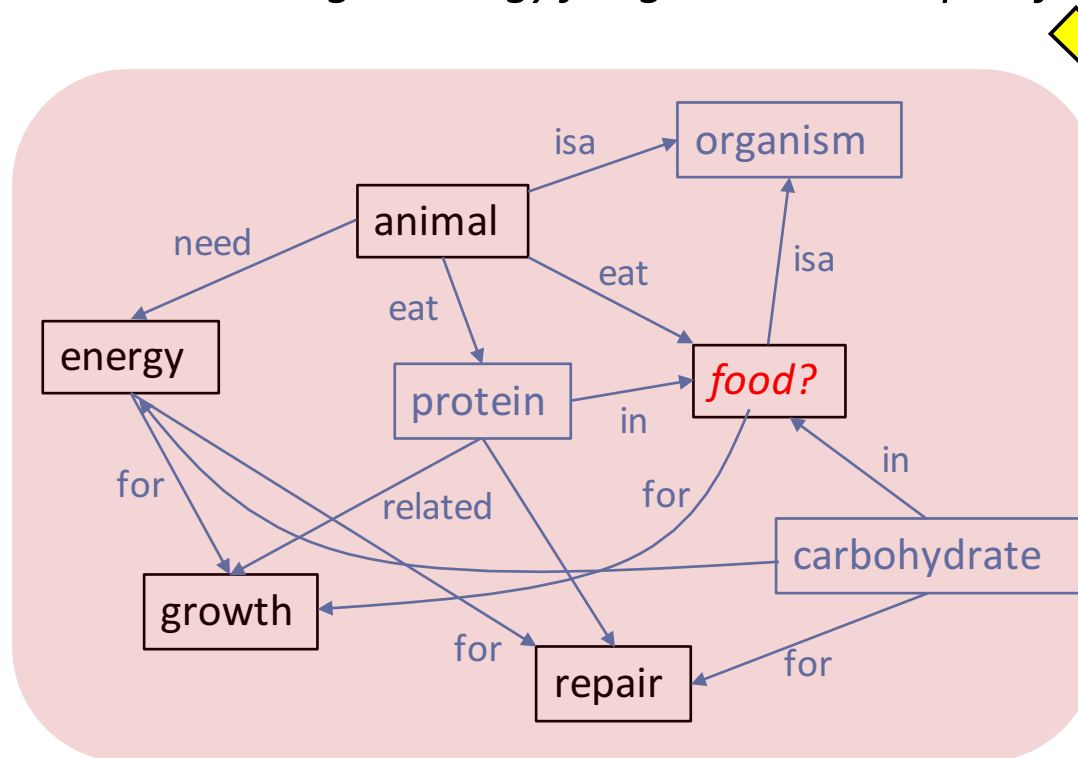


Illustration of the Idea

- For example (idealized):

Question: Animals get energy for growth and repair from (A) food (B) air

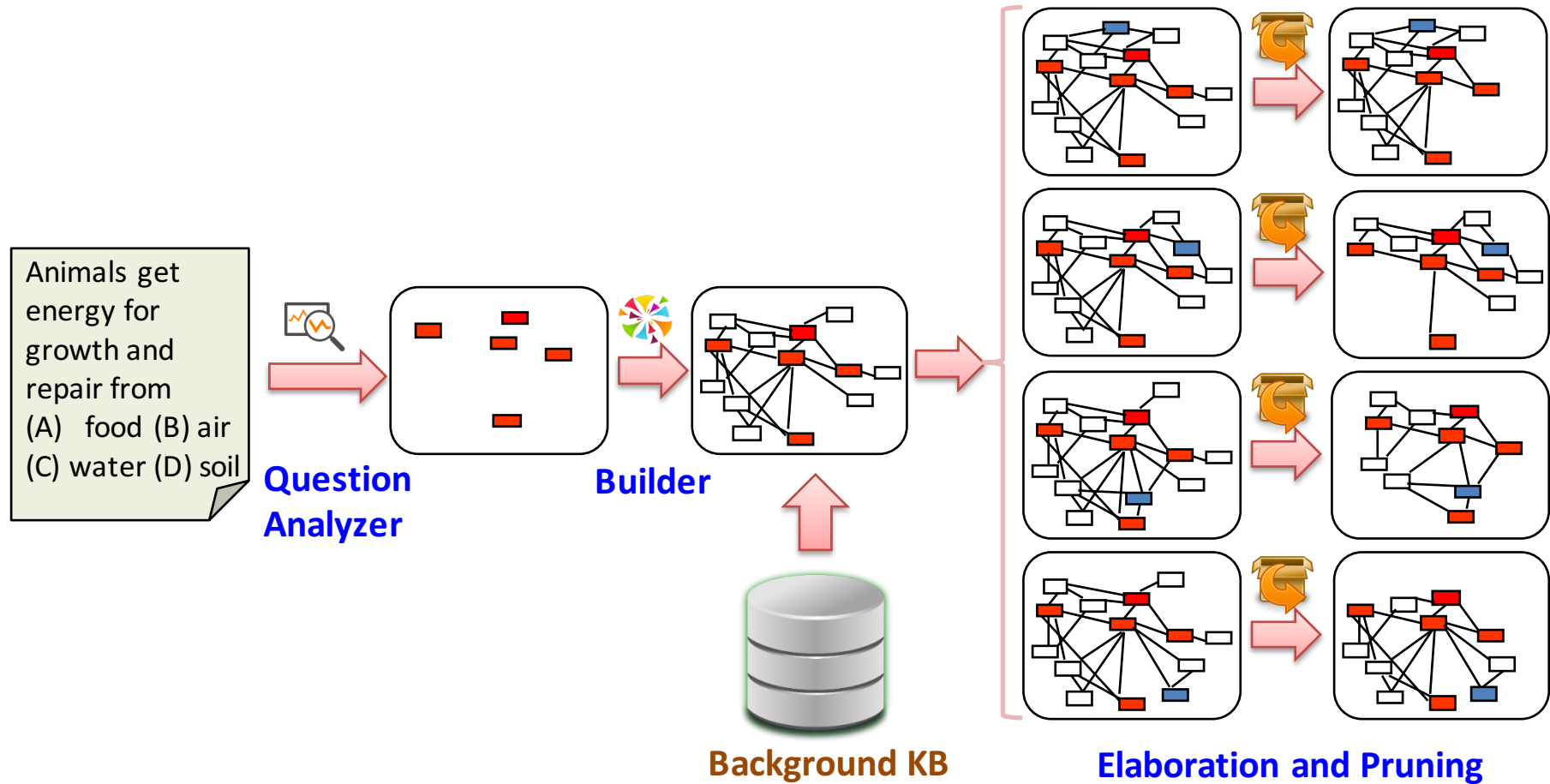


Answer:
(A) food

So food is the
most coherent
answer

coherent scene

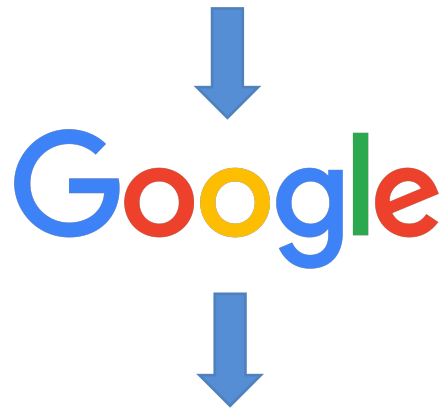
Our Approach



Question Analyzer

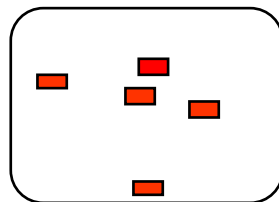
- Input:

Animals get
energy for growth
and repair from
(A) food (B) air
(C) water (D) soil



- Search question in Google
- Fetch top 20 snippets
- Calculate the TFIDF value for each non-stop word in question, treat TFIDF value as importance score.

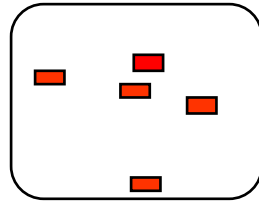
- Output:



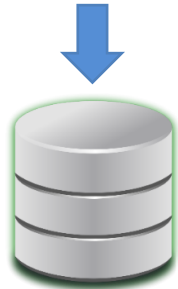
repair (1.19)
growth (0.97)
animal (0.82)
energy (0.80)

Builder

- Input:



repair (1.19)
growth (0.97)
animal (0.82)
energy (0.80)

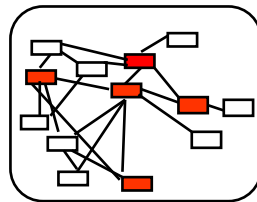


Background KB

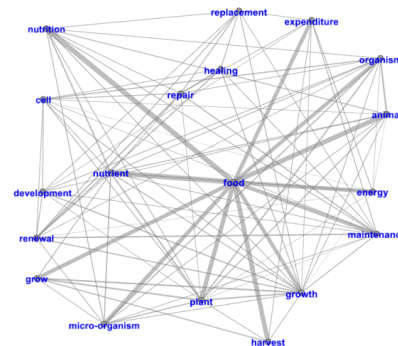
$$\sum_i \omega_i \text{Association}(q_i, w)$$

- DART (e.g. animal <eat> protein)
- WordNet (e.g. animal <isA> organism)
- FreeAssociation (e.g. animal <related> nature)

- Output:

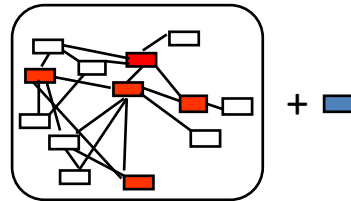


renewal; nutrient; organism; plant; maintenance;
nutrition; food; replacement; grow; harvest; cell;
micro-organism; healing; expenditure; development



Elaboration

- Input:



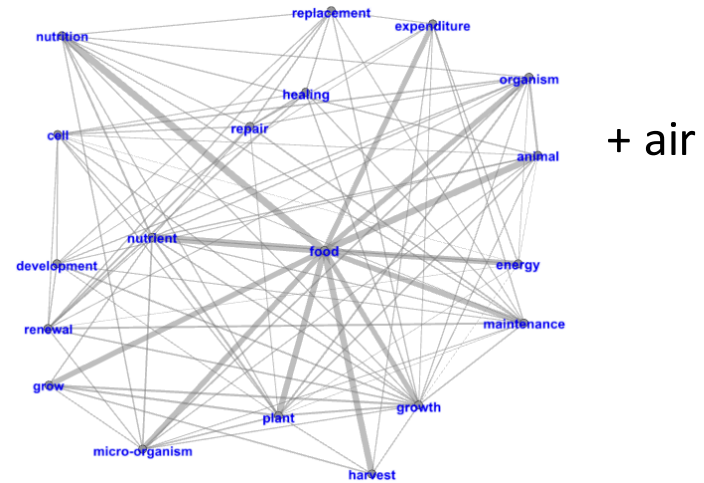
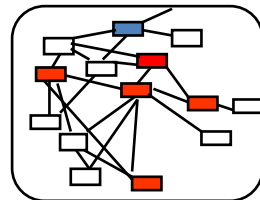
+ 



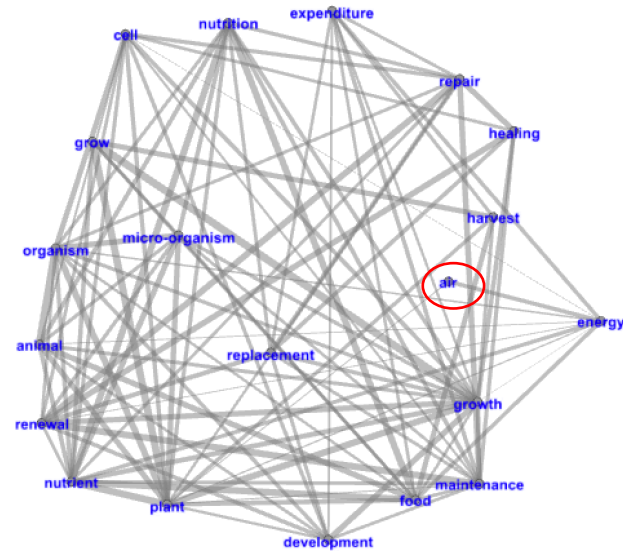
Background KB



- Output:

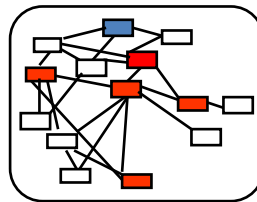


+ air



Pruning

- Input:

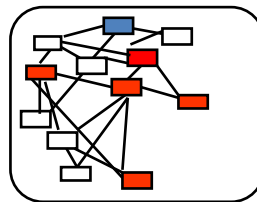


```
while (answer option node is still in graph)
{
    remove the non-input node with minimum weighted degree;
    update the graph;
    record the resulting graph;
}
return the subgraph with maximum "minimum weighted degree"
```

summed weight of
all incident edges



- Output:

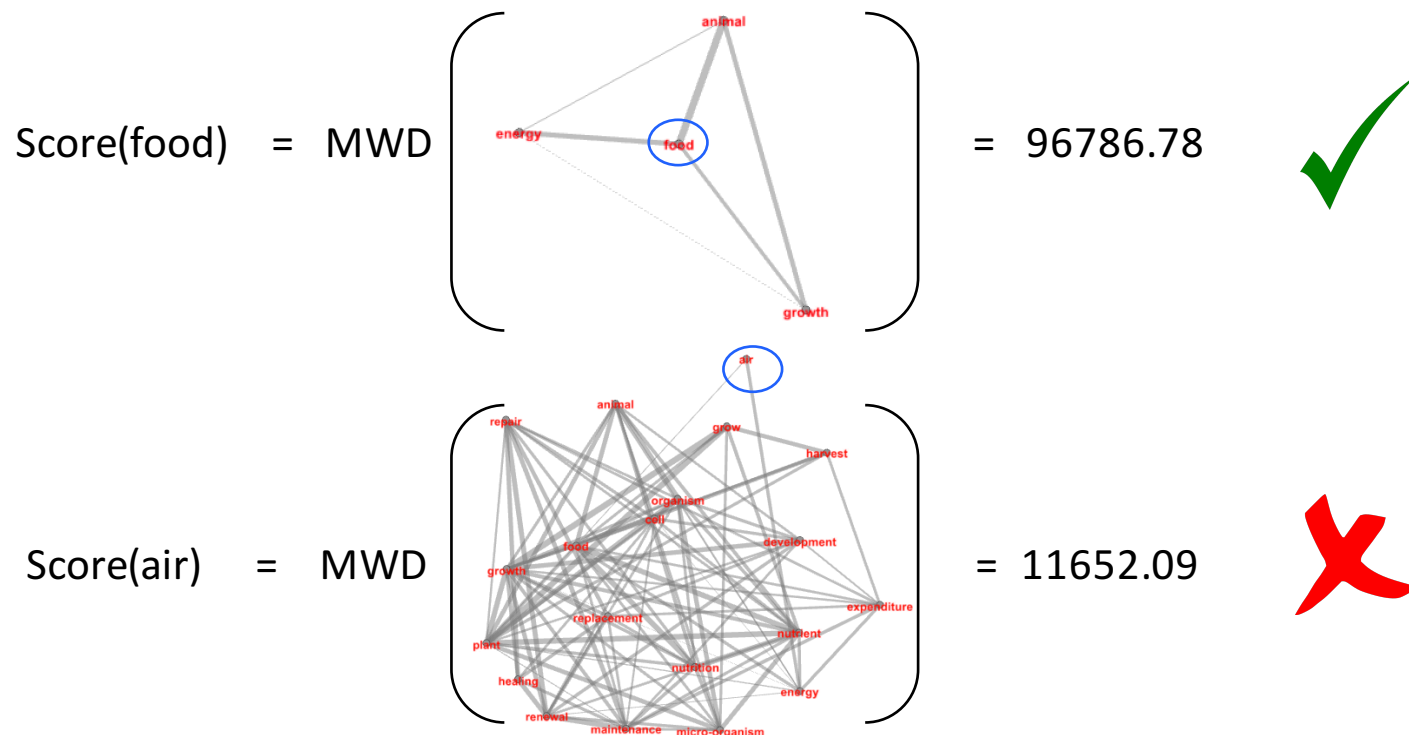


Only the necessary nodes
bridging the question and the
answer option should be kept!

Answer Ranking

- Intuition/Assumption: *answer* being more densely connected is more likely to be the correct answer.
- Rank answer options with respect to:

Score(option_i) = Minimum Weighted Degree of option_i's resulting scene



Experiments

- Background KBs: WordNet+FreeAssociation+DART
- Relatedness Measure: cos similarity between word vectors
 - Word2vec trained on 4th grade textbook (w/ Web enrichment)
- Datasets:

Dataset	# questions
Regents-Dev	47
Regents-Test	23
Viceroy	26
Grade5	197

real 4th grade multiple choice science questions

Experiments

- Ablations:

Configuration	Adding extra nodes?	Pruning?
Scene	Yes	Yes
-NewNodes	No	Yes
-Prune	Yes	No
-Both	No	No

- Baselines:

- LSModel: combines language models and IR scores using SVM.
- A*Rules: prove answer by applying lexical inference rules.

Performance Comparison

Baselines			Ablations			
	LSModel	A*Rules	SceneQA	-NewNodes	-Prune	-Both
Regents-Dev	65.96	65.96	83.51	65.96	70.74	59.57
Regents-Test	58.70	67.00	66.30	69.57	57.61	65.22
Viceroy	28.85	47.00	65.38	42.31	47.12	42.31
Grade5	30.08	29.22	55.20	51.78	50.13	50.25

Adding new nodes is helpful!

Performance Comparison

Baselines			Ablations			
	LSModel	A*Rules	SceneQA	-NewNodes	-Prune	-Both
Regents-Dev	65.96	65.96	83.51	65.96	70.74	59.57
Regents-Test	58.70	67.00	66.30	69.57	57.61	65.22
Viceroy	28.85	47.00	65.38	42.31	47.12	42.31
Grade5	30.08	29.22	55.20	51.78	50.13	50.25

Pruning is helpful!

Error Analysis

- **Insufficient relatedness measure!**
- Examples:
 - The process that changes a gas to liquid is called
(A) *condensation* (B) melting (C) *evaporation* (D) vaporization
 - An animal that has a backbone is called a(n)
(A) *invertebrate* (B) *vertebrate* (C) exoskeleton (D) sponge
- Analysis: the relatedness measure (i.e. word2vec) cannot distinguish words with opposite meanings since they usually share the similar distributional semantics

Error Analysis

- **Word order matters for some questions!**
- Examples:
 - A pot is heated on a stove. Which *process* causes the metal handle of the pot to also become hot?
(A) combustion (B) convection (C) radiation (D) conduction
 - Baby chicks peck their way out of their shells *when* they hatch. This activity is an example of which of the following types of behavior?
(A) instinctive (B) learn (C) plan (D) social
- Analysis: the representation of the scene is still based on BOW, it is not capable of representing a “process”

Conclusions

- Implicit background knowledge is critical for answering elementary science questions
- Linguistic resources can supply some of such knowledge
- Coherent scene extraction from knowledge graph is promising for adding implicit knowledge and answering elementary science questions
 - constructing a coherent "scene" to represent the question
 - the answer is then the one that best fits the "scene"

Thanks!

A hand-drawn illustration in black ink. It features a simple, round face with a wide, curved smile and two small dots for eyes. A hand with five fingers is reaching up from the bottom left towards the word 'Thanks!'. The word is written in a large, cursive, handwritten style. A horizontal line is drawn across the middle of the word, and the hand is positioned just below this line. The overall style is casual and friendly.

