Multiple Choice Question Generation Utilizing An Ontology

Katherine Stasaski and **Marti Hearst**
UC Berkeley
{katie_stasaski, hearst}@berkeley.edu

Abstract

Ontologies provide a structured representation of concepts and the relationships which connect them. This work investigates how a pre-existing educational Biology ontology can be used to generate useful practice questions for students by using the connectivity structure in a novel way. It also introduces a novel way to generate multiple-choice distractors from the ontology, and compares this to a baseline of using embedding representations of nodes.

An assessment by an experienced science teacher shows a significant advantage over a baseline when using the ontology for distractor generation. A subsequent study with three science teachers on the results of a modified question generation algorithm finds significant improvements. An in-depth analysis of the teachers' comments yields useful insights for any researcher working on automated question generation for educational applications.

1 Introduction

An important educational application of NLP is the generation of study questions to help students practice and study a topic, as a step toward mastery learning (Polozov et al., 2015). Although much research exists in automated question generation the techniques needed for educational applications require a level of precision that is not always present in these approaches.

Ontologies have the potential to be uniquely beneficial for educational question generation because they allow concepts to be connected in non-traditional ways. Questions can be generated about different concepts' properties which span different areas of a textbook or even different educational resources.

However, ontologies are not commonly used in NLP approaches to generate complex, multi-part questions. This may be due to concern about ontology's incompleteness and the fact that they are usually structured for other purposes.

In this work, we describe a novel method for generating complex multiple choice questions using an ontology, with the aim of testing a student's understanding of the bigger picture of how concepts interact, beyond just a definition question. This technique generates questions that help achieve understanding at the second level of Bloom's taxonomy (Bloom et al., 1956). We also generate multiple choice distractors using several ontology- and embedding-based approaches.

We report on two different studies. The first assesses both the questions and the question distractors with one domain expert, a middle school science teacher. This finds evidence that the ontology-based approach generates novel and useful practice questions. Based on the findings from that study, we adjust the question generation algorithm and report on a subsequent evaluation in which three experts quantitatively rank and qualitatively comment on a larger selection of questions. The results are strong, with more than 60 questions out of 90 receiving positive ratings from two of the judges. Additionally, we categorize and provide in-depth analysis of qualitative feedback and use this to inform multiple future directions to improve educational practice question generation.

2 Related Work

Prior work has explored both automatically generating educational ontologies from text and utilizing expert-created ontologies for other tasks. For instance, Olney et al. (2011) explored extracting

Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, pages 303–312
Copenhagen, Denmark, September 8, 2017. ©2017 Association for Computational Linguistics

nodes and relationships from text to build a concept map ontology automatically from textbooks. Other work has also attempted to build ontologies from non-educational texts (Benafia et al., 2015; Szulman et al., 2010) and has explored utilizing crowd-sourcing to build an ontology from text (Getman and Karasiuk, 2014).

Prior approaches to question generation from ontologies have involved hand-crafted rules to transform a relationship into a question (Olney et al., 2012b; Papasalouros et al., 2008; Ou et al., 2008). However, these approaches mainly generate questions for a single fact and do not combine multiple pieces of information together to create more complex questions. There is the potential to explore other, more complex, types of question generation procedures from the ontology. Approaches have also utilized online questions for ontology-driven generation, but this is less generalizable (Abacha et al., 2016).

Prior work aimed at generating educational practice questions has generated questions directly from text using a series of manual translations and a ranking procedure to determine quality (Heilman and Smith, 2010, 2009; Heilman, 2011).

Other work has focused on question generation, independent of an educational context. A large-scale question generation task posed to the community prompted a focus on factual question generation from texts and knowledge bases (Rus et al., 2008; Graesser et al., 2012). Approaches have included factual generation directly from text (Brown et al., 2005; Mannem et al., 2010; Mazidi and Tarau, 2016; Yao et al., 2012) as well as generation from knowledge bases (Olney et al., 2012a).

Recent advances in text generation have used neural generative models to create interestingly worded questions (Serban et al., 2016; Indurthi et al., 2017). However, because we are using a human created ontology and lack specialized training data, we utilize hand-crafted rules for generation.

3 Question Generation

We utilize an educational Biology ontology to generate multiple choice questions, which consist of the text of a question, the correct answer, and three distractor multiple choice candidates.

3.1 Dataset

We use an expert-curated ontology documenting K-12 Biology concepts (Fisher, 2010) designed for educational applications. While more resources could be used to accomplish this task, we only utilize the ontology to explore the efficacy of this question generation approach. By utilizing an expert-curated ontology instead of an automatically generated one, we operate under the assumption that the ontology is correct and complete. Future work can explore utilizing this method in conjunction with other educational resources and techniques.

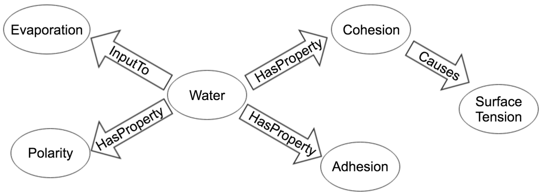


Figure 1: Selected part of the Biology ontology.

The ontology contains 1,260 unique concept nodes and 227 unique relationship types with a total of 3,873 node-relationship-node triples. The average outgoing degree is 7. Figure 1 shows a small sample.

3.2 Using The Structure of the Ontology

The novel aspect of our approach is the manner in which we use an ontology to go beyond simple factoid question generation. Rather than generating a question from a node-relation-node triple, this algorithm makes use of the graph structure of the ontology to create complex questions that link different concepts, with the aim of challenging the student to piece together different concepts.

The goal of this evaluation was to determine if this novel way of combining concepts would be judged as creating useful, coherent questions for testing students.

To create these novel structured questions, the algorithm chooses a node to act as the answer, and from three randomly-chosen outgoing links it generates a question. The relations of the outgoing links and the nodes on the other ends are used to form the question words. For instance, from the node “Water” emanates the links (DissolvesIn, “salt”), (HasProperty, “cohesion”), and (InputTo, “evaporation”) from which is generated the question “What dissolves salt, has cohesion, and is an input to evaporation? (Water)”

A total of 992,926 questions can be generated via this method from the ontology. These questions are distributed over 426 nodes, with the average number of questions that can be generated per node being 2,330. While 834 nodes do not have three outgoing links to generate a question from, these nodes can be chosen as properties to commpose other questions.

3.3 Generating Distractors

Good multiple choice questions should have distractors (alternative answers to distract the student from the correct answer). These should not be synonymous with the correct answer, but should be a plausible answer which should not be so far-fetched as to be obviously incorrect.

We experimented with several different ways of generating multiple choice distractors using the structure of the ontology, and compared these with two embedding based methods. In each case, if the text of a distractor overlaps with the correct answer, we do not use it.

3.4 Ontology Distractor Generation

We experimented with 5 different ontology-based distractor methods. For each distractor generation method, the correct answer node, n is connected to three property nodes $n1, n2$, and $n3$ via relationships $r1, r2$, and $r3$ respectively. In order to ensure that distractor node m does not correctly answer the question, we make sure at least one of $n1, n2$, or $n3$ does not connect to m. The following methods are illustrated in Figure 2.

Two Matching Relationships: This method chooses m such that m is connected to $n1$ via $r1$ and m is connected to $n2$ via $r2$.

One Matching and One New Relationship: This method chooses m such that m is connected to $n1$ via $r1$ and m is connected to $n2$ via a different relationship, $r4 \neq r2$.

Two New Relationships: This method chooses m such that m is connected to $n1$ via a different relationship type $r4 \neq r2$ and m is connected to $n2$ via a different relationship, $r5 \neq r3$.

One Matching Relationship: This method chooses m such that m is connected to $n1$ via $r1$.

We also examined an additional question-independent ontology-grounded approach.

Node Structure: This approach rates pairs of nodes by similarity, where similarity is determined by their tendency to link to similar relation types and to link to the same intermediate nodes. More formally, let c_n denote the set of nodes which are connected to any node n and let $l_{n,r}$ denote the number of connections that n has of type r. The similarity between n and m is computed as:

$$s_{n,m} = count(c_n \cap c_m) - \sum_r |l_{n,r} - l_{m,r}|$$

3.5 Embedding Distractors

We implemented two methods to generate distractors grounded in the embeddings of the nodes. Both utilize pre-trained word embeddings (Mikolov et al., 2013). A given node n consists of a series of words, $w_1, w_2, ..., w_n$. We create a multi-word embedding by distributing weight and placing more emphasis on the last word in a sequence, which we assume to be the head word. The similarity s between the two embedded nodes e_n and e_m is determined by cosine similarity.

Correct Answer Embeddings: are generated by comparing the correct answer, n with the most similar node in the graph G:

$$\text{distractor} = \arg\max_{m \in G} s_{e_n, e_m}$$

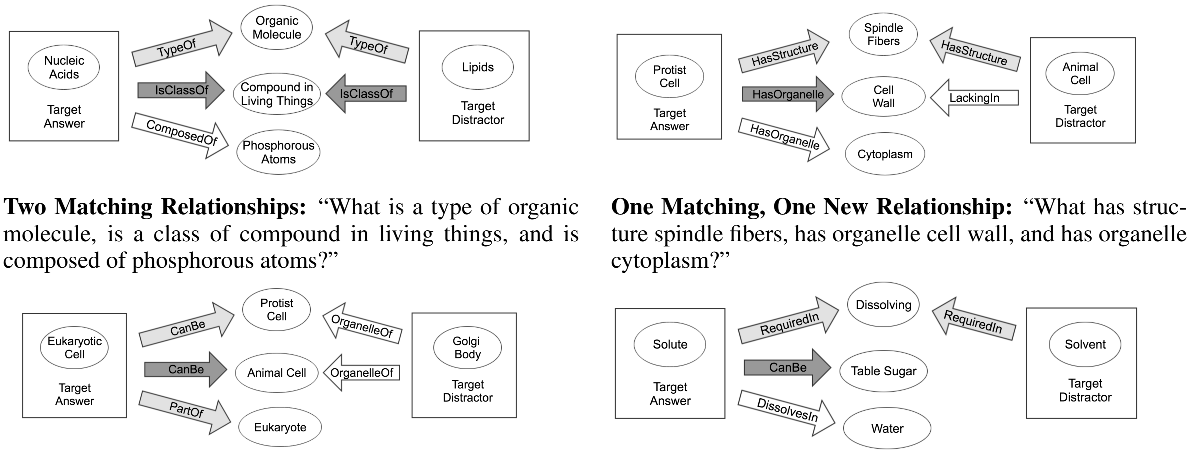
Question Component Embeddings: are generated by finding the most similar node to the question components n_1, n_2, and n_3. The above equation is computed for each component.

3.6 Ontology Coverage

Each of these methods is applicable to a subset of nodes in the ontology. From a randomly sampled selection of 10,000 questions, 15.6% met requirements for Two Matching Relationship Distractors, 16.2% met requirements for One Matching, One New Distractors, 29.1% met requirements for Two New Relationship Distractors, and 25.6% met requirements for One Matching Relationship Distractors. Node Structure, Correct Answer Embeddings, and Question Component Embeddings all had complete coverage due to the nature of the methods.

3.7 Pedagogical Motivation

This question generated method is similar to an inverse of the "Feature Specification" questions described by Graessner et al (1992) in which students are asked to describe properties of a concept. An example format of this type of question is "Which qualitative attributes does entity X have?" (Graesser et al., 1992). Instead of prompting students to list properties of a concept, we provide



Two Matching Relationships: "What is a type of organic molecule, is a class of compound in living things, and is composed of phosphorous atoms?"

One Matching, One New Relationship: "What has structure spindle fibers, has organelle cell wall, and has organelle cytoplasm?"

Two New Relationships: "What can be protist cell, can be animal cell, and is a part of a Eukaryote?"

One Matching Relationship: "What is required in dissolving, can be table sugar, and dissolves in water?"

Figure 2: Question-specific distractor generation methods. Correct answer nodes are leftmost in each graph, and chosen distractor nodes are rightmost.

three features of concepts and ask the students to choose the correct concept given the features.

Through these questions, we aim to challenge students to connect different features of a concept while working within the constraints that the questions and corresponding answers be able to be generated via the ontology. The type of questions that arise are intended for mastery learning, in which students learn simpler facts about a concept before tackling more difficult conceptual problems (Polozov et al., 2015). While the questions are not critical thinking ones, they are designed to be more complex than a simple definition question and to be a gateway to more difficult questions.

Another potential application of these questions is preparation for extracurricular trivia competitions, such as Quiz Bowl[1]. One type of question asked at these competitions is one which lists many characteristics of a concept and challenges students to quickly identify the concept. Connecting multiple facets of a concept are essential to answer these questions.

3.8 Generating the Text

Because the purpose of this study was to examine the feasibility of the ontology structure for question generation, we use hand-crafted rules to produce the question text. The human-generated ontology has nodes and relationships that are worded somewhat naturally, which helped this process. We devised simple relationship-to-text translations rules; examples are shown in Table 1.

Relationship	Rule
Has - Characteristic	If n = verb → "n." If n = noun → "has n. " If n = adjective → "is n."
HasProcess	"has a process called n."
CanBe	"can be a/an n."

Table 1: Common relationship types and rule-methods used to generate the question segment.

4 Study 1

4.1 Method

We conducted a study to assess the quality of the questions and distractors. We asked a middle school science teacher with 20 years of experience to rate the quality of 20 complex questions on a scale of 1-7. The scale was explained such that 1 was "Poor," 4 was "OK," and 7 was "Excellent." To create the test set, we randomly selected nodes and generated questions about them [2].

We also asked the teacher to rate distractors on a scale of 1-5 (a narrower scale was chosen as it was thought these would be more difficult to differen-

[1] https://www.naqt.com/about-quiz-bowl.html

[2] Full text of questions evaluated in both studies can be found in an appendix in the supplementary materials.

tiate than the questions). Each distractor method was tested with 10 different questions (each with 3 distractors). Questions were randomly chosen from all possible questions that could be generated from the ontology, and were disregarded if a given distractor method was not able to generate three valid distractors. For both processes described above, the teacher was prompted to enter optional comments about the question or distractors.

4.2 Results

Quantitative distractor generation results can be seen in Table 2. The difference in ontology-generated distractors compared to embedding-generated distractors was significant using a t-test with p value < 0.001. Comparing the highest-performing ontology and embedding methods (One Matching, One New Relationship and Correct Answer Embedding) is also significant under the t-test with $p < 0.05$. This indicates that while the overall feedback was critical, there is promise in using ontology over embedding distractors.

The questions' ratings averaged 2.25 out of 7. After analyzing the qualitative comments, this can be attributed primarily to the unnatural wording of the questions. Qualitative comments about the questions are categorized in Table 3.

Distractor Type	Avg
Two Matching Relationships	2.37
One Matching, One New Relationship	2.78
Two New Relationships	2.03
One Matching Relationship	2.07
Node Structure	2.63
Correct Answer Embedding	2.10
Question Component Embedding	1.60

Table 2: Averaged distractor scores.

Type of Comment	Count
Unnatural Wording of Question	31
Good question	24
OK question	17
Unnatural Grouping of Characteristics	2
Text of Node was Confusing	2
Imprecise Relationship	1
Not Middle School Level	1

Table 3: Categorization of qualitative feedback for questions. Feedback was included for all questions, including those in the distractor section.

4.3 Discussion

All ontology distractor methods except for One Matching Relationship received explicit comments pointing out that the distractors were "good," while no embedding approach received these comments. This indicates that the combination of different methods have strengths that contribute to a good set of distractors. The Node Structure method provides broad distractors, while the other methods provide question-specific ones. For example, the distractors "cell wall," "chloroplast," and "central vacuole" for the question "What is an organelle of eukaryotic cell, is an organelle of animal cell, and is an organelle of fungal cell? (Golgi body)" are plausible but incorrect.

However, the teacher also commented that some distractors of both embedding and ontology methods were "poor." Examining the "poor" distractors for embedding questions shows that the distractors can come from concept areas unrelated to the question. For instance, for the question "What contains chromosome, is an organelle of eukaryotic cell, and is a type of organelle? (nucleus)," the Question Component Embeddings generated "new genetic recombinations" as a distractor.

By contrast, the ontology-generated distractors were marked "poor" when the question included one unique property. For example, for the question "What eats mice, eats deer, and is a type of predator? (mountain lion)," the improbable distractor "vole" was chosen because it eats mice and is a predator. This suggests the necessity of more formal reasoning and real world knowledge coupled with the ontology information.

There were also instances in which the proposed distractors were unintentionally correct answers. For the embedding-generated distractors, this happened because the method favors distractors that are more similar to the correct answer. For the ontology-generated distractors, this occurred where the ontology was incomplete.

While the distractor evaluation is quite preliminary as it only involves one expert evaluating 10 sets of distractors per generation method, these results suggest the potential to explore an ontology method of distractor generation in future work.

5 Study 2

Based on these initial results, we extended the work in several ways. First, based on the results of Study 1, we found that the teacher was sensitive to any flaws of the wording of the questions, so we modified the assessment with questions that were manually touched up to remove grammatical errors. Second, although we had evidence that the ontology-based method was producing high-quality questions, we noticed that the target answers of many of the questions were quite general (e.g. "solids", "water"). Therefore, we modified the algorithm to take out-degree and relation commonality into account. We also decided to investigate questions composed of only two relations as well as three relations. Finally, we wanted to improve the evaluation in two ways: (i) by assessing more questions, and (ii) by having more independent judges per question. We accomplished this by finding assessors with appropriate backgrounds on an expert-oriented crowdwork site. Each of these modifications is described in detail below, along with results of this second evaluation.

5.1 Modifications to Generation Algorithm

We wanted to assess if the ontology generation method worked well, but were wondering if perhaps including three relations made the questions too complex or unusual. For this reason, we decided to include questions with only two relations in Study 2.

After adapting the question generation method to generate questions with two properties as opposed to three, given that the node "Evaporation" is connected to the two relations (Outputs, "water vapor") and (OppositeOf, "condensation"), the question generated from these two properties is: "What yields water vapor and is the opposite of condensation? (evaporation)."

Improving diversity of questions and coverage of the ontology were priorities in this study. We modified our question generation algorithm to prioritize these goals. We placed restrictions on the number of outgoing connections of a node $conn_n$ such that $5 < conn_n < 30$.

In addition, for two-property questions, we imposed the constraint that the collection of chosen properties yields exactly one unique correct answer. This added an additional check that the question generated was not about general properties that multiple nodes in the ontology fulfill.

For a random selection of 45 questions, we do not allow the algorithm to generate more than 2 questions about the same concept, to evaluate a more diverse set of questions. We also do not allow a node to be asked as a property involved in a question more than 5 times. This procedure was used for the selection of questions for the study below.

5.2 Manual Adjustments of Question Expression

The first experiment showed that the grammatical errors were distracting and affected the evaluation of the content of the questions. Therefore, we adjusted the grammatical correction rules as well as made minor edits by hand to ensure grammatical correctness. Some minor grammatical errors still exist, but major ones which obscure the meaning of the question were manually corrected. So, for instance, we fixed errors in the specification of articles, as seen in the removal of *a* and addition of *s* when transforming the question "What yields a RNA and is contained in chromosome? (gene)" to "What yields RNA and is contained in chromosomes? (gene)." These changes were made to a total of 16 questions. When a question was modified, the average number of changed characters was 3.3.

5.3 Evaluation of Question Quality

Three middle school science teachers, each with at least 10 years of experience, were recruited to evaluate the generated questions via Upwork[3], an expert-oriented freelance work matching site. Each teacher evaluated 90 questions (45 two-property and 45 three-property) both quantitatively on a scale from 1 to 7 and qualitatively via open-ended comments.

The title shown to the assessors was "Middle School Science Practice Question Evaluation" and the instructions for the assessment were:

> Below are shown some automatically generated biology questions, intended for practice studying. Please rate the quality of these questions for these purposes. Use the rating 1 to 7 where 1 = Poor, 4 = OK, and 7 = Excellent.
>
> Please ignore grammatical errors. For each question, please briefly explain your rating in one sentence.

We recognize that by informing teachers that the questions were generated automatically, they

[3] http://www.upwork.com

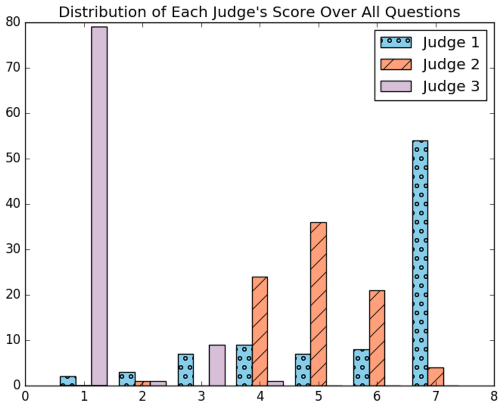


Figure 3: Distribution of judges' quantitative score over the 90 evaluated questions.

Evaluator	**2R Questions**	**3R Questions**
Evaluator 1	5.76	5.94
Evaluator 2	5.33	4.63
Evaluator 3	1.36	1.13
Average Score	**4.15**	**3.89**

Table 4: Quantitative feedback of question quality from the second study, scored on a scale from 1 to 7. 2R Questions are created from two relationships, 3R from three relationships.

could potentially be biased (either positively or negatively) when evaluating. However, because the generated questions are a new type of multiple-choice question, there is no naturally-arising human-generated baseline.

5.4 Results

Quantitative results from the second study can be viewed in Table 4. Both 2R and 3R questions achieved similar rankings, with no significant differences between the two (t-test, p=0.37). A histogram with judges' score distribution can be seen in Figure 3.

The qualitative results were analyzed and categorized in Table 5 and are discussed in the next subsection.

5.5 Discussion

Compared to the previous study, the fixing of grammatical errors allowed us to better determine the quality of the content of questions. While the evaluators did comment on the grammar of the questions and suggested corrections a total of 37 times, the quality of the content was able to be evaluated.

On the positive side, 77 questions were praised as being clear and easy to understand (see Table 5). We believe this is due to a combination of the two changes we made in response to the first study–improving the syntax and orienting the questions towards more specific answers.

Additionally, one teacher commented that 10 questions had particularly good properties. Two examples are: "What has a deoxyribose and occurs at the mitochondria? (DNA)" and "What includes carbon, is a part of organic molecule, and includes oxygen? (CHNOPS)." The updated algorithm ensured that the chosen properties were less vague and also not too specific.

In some cases, the evaluators had an explicit positive response to questions' logically grouping properties within questions. One teacher specifically pointed out that certain questions contained valuable properties which guide the students to the correct answer, as in: "What contrasts with plant cell and has an organelle called rough ER? (animal cell)". This provides positive support for the goal of this style of generating questions by include multiple pieces of information from the ontology.

The descriptive vocabulary of 8 questions was also pointed out by one evaluator, such as "What includes glucose, includes deoxyribose, and is a type of sugar? (monosaccharide)." Since we utilize a human-created ontology to generate questions, the nodes and relationships are often descriptive. Our generation method leverages this to create questions. Given an ontology with descriptive, precisely-worded relationships and nodes, questions generate via our method will reflect the diverse vocabulary.

On the negative side, one teacher was particularly skeptical of the ability of this method of questions to prepare students for standardized testing. She pointed out that this factual question style did not challenge students to think critically. While this is true, this is not the main focus of this work. We aim to over-generate simple practice questions to ensure students have adequate materials to practice and study with, before they have mastered a concept.

Two teachers specifically mentioned that the style of questions were repetitive. The lack of diversity of questions is something which can be addressed in future work. However, these stud-

Qualitative Feedback	**Evaluator 1**	**Evaluator 2**	**Evaluator 3**	**Total**
Clear and easy to understand	13	64	–	**77**
Simple to answer	20	–	–	**20**
Poor wording	–	3	16	**19**
Does not prepare for standardized test	–	–	19	**19**
Too many “What“ questions	17	–	–	**17**
Vague/Broad	–	2	15	**17**
Confusing/Poor properties chosen	1	4	10	**15**
Too many options in question	9	1	1	**11**
Good chosen properties	10	–	–	**10**
Small rewording suggestion	–	8	2	**10**
Ontology flaw pointed out	–	5	4	**9**
More precise rewording suggestion	–	3	6	**9**
Descriptive vocab	8	–	–	**8**
Detailed question	7	–	–	**7**
“Trivia” question	–	–	6	**6**
Simple vocab	4	–	–	**4**
Too specific	–	–	4	**4**
Not a critical thinking question	–	–	4	**4**
Visual diagram needed	–	–	4	**4**
Good intermediate steps to guide to correct answer	4	–	–	**4**
Alternate phrasing for ontology node	–	2	1	**3**
Academic language	2	–	–	**2**
Redundant concepts chosen	2	–	–	**2**
Poor format of multiple choice	–	–	2	**2**
Too many similar questions	1	–	–	**1**
OK questions	1	–	–	**1**
Confusing property represented in the ontology	–	1	–	**1**
Should be a higher Bloom’s Taxonomy question	–	–	1	**1**

Table 5: Qualitative feedback from evaluators, categorized by type of comment.

ies show promise for using an ontology to inform factual question generation, and future work can extend this method to other question types.

Two of the teachers also pointed out parts of the generated questions that were scientifically incorrect. Examining these questions shows that the ontology contains incorrect information. This points to the necessity of validating and updating created ontologies. Our method, while generalizable to other ontologies, assumes the correctness of the information represented. Future work can examine verifying ontologies, via methods such as dialogue, parsing educational materials, or direct validation from experts.

In nine of the comments, the wording of questions was stated as being imprecise. For instance, *solid* is linked to (CharacteristicOf, ”Solvent”). It was pointed out that while most solvents are solid, some can be liquids or gases. This underscores the finding of the first study that errors in the ontology can lead to errors in the questions.

It also seems that the selection of nodes to form questions can be improved. Certain questions were pointed out to have poor groupings of properties. For instance, “What is produced by an ovary and via fertilization creates an embryo? (egg)”was thought not to be a good question, perhaps because to know either portions of the question one must know what an egg is.

6 Conclusion

We presented a novel way of generating complex multiple choice questions and distractors from an educational ontology. We showed significant improvement when the ontology was used to generate distractors compared to an embedding approach. Insights gained from evaluation indicate a necessity of ontology augmentation and a more advanced reasoning model.

We also showed in a subsequent study that question content, when adapted to account for potentially vague questions, has promising results. Future work may benefit from incorporating more knowledge rich approaches such as Berant et al.'s (2014) work on deep analysis of biology texts.

Our algorithm for choosing properties to include in questions and generating distractors is generalizable to other ontologies, although our method assumes a near-complete ontology, as distractors are generated via assumptions that the absence of a link implies the absence of a relationship. Changing the text-generating rules may be necessary to generalize our approach as these are tied to the specific relationships of our ontology. Our ontology contains many naturally-worded relationships, which aided this process. Other text generation methods can be explored in future work, as well, to rectify this.

Insights gained from the teachers' qualitative feedback are applicable to other question generation methods as well. Future work should focus on generating increasingly more complex questions which focus on higher levels of Bloom's taxonomy. External knowledge not represented in the ontology can be used to both increase the difficulty of the questions as well as improve simpler methods of question generation. Finally, verifying the completeness and accuracy of ontologies as well as wording questions diversely and precisely should be a focus going forward.

Acknowledgments

We thank Ann Peterson and the three Upwork science teachers for assessing the results of this analysis, Yiyi Chen for coding support, as well as Jonathan Gillick, David Bamman, Julian Park, and the anonymous reviewers for helpful comments. This work was supported by a UC Berkeley Chancellor's Fellowship.

References

Asma Ben Abacha, Julio Cesar Dos Reis, Yassine Mrabet, Cédric Pruski, and Marcos Da Silveira. 2016. Towards natural language question generation for the validation of ontologies and mappings. *Journal of biomedical semantics* 7(1):48.

Ali Benafia, Smaine Mazouzi, and Sara Benafia. 2015. Building ontologies from text corpora. In *ICEMIS '15 Proceedings of the The International Conference on Engineering MIS 2015*.

Jonathan Berant, Vivek Srikumar, Pei-Chun Chen, Abby Vander Linden, Brittany Harding, Brad Huang, Peter Clark, and Christopher D Manning. 2014. Modeling biological processes for reading comprehension. In *EMNLP*.

Benjamin S. Bloom, David Krathwohl, and Bertram B. Masia. 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.*. Green.

Jonathan C. Brown, Gwen A. Frishkoff, and Maxine Eskenazi. 2005. Automatic question generation for vocabulary assessment. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. HLT '05, pages 819–826.

Kathleen Fisher. 2010. Biology lessons at sdsu. http://naturalsciences.sdsu.edu/.

Anatoly P. Getman and Volodymyr V. Karasiuk. 2014. A crowdsourcing approach to building a legal ontology from text. *Artificial Intelligence and Law* 22:313–335.

Arthur C. Graesser, Natalie Person, and John Huber. 1992. Mechanisms that generate questions. In Thomas W. Lauer, Eileen Peacock, and Arthur C. Graesser, editors, *Questions and Information Systems*, Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, chapter 9, pages 167–187.

Arthur C. Graesser, Vasile Rus, and Zhiqiang Cai. 2012. Question answering and generation question answering and generation. In *Applied Natural Language Processing: Identification, Investigation and Resolution*.

Michael Heilman. 2011. *Automatic Factual Question Generation from Text*. Ph.D. thesis, Carnegie Mellon University.

Michael Heilman and Noah A. Smith. 2009. Question generation via overgenerating transformations and ranking. Technical Report CMU-LTI-09-013, Carnegie Mellon University.

Michael Heilman and Noah A Smith. 2010. Good question! statistical ranking for question generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 609–617.

Sathish Indurthi, Dinesh Raghu, Mitesh M. Khapra, and Sachindra Joshi. 2017. Generating natural language question-answer pairs from a knowledge graph using a rnn based question generation model. In *ACL 2016*.

Prashanth Mannem, Rashmi Prasad, and Aravind Joshi. 2010. Question generation from paragraphs at upenn: Qgstec system description. In *Proceedings of QG2010: The Third Workshop on Question Generation*.

Karen Mazidi and Paul Tarau. 2016. Infusing nlu into automatic question generation. In *INLG*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. *CoRR* abs/1310.4546.

Andrew Olney, Arthur C. Graesser, and Natalie K. Person. 2012a. Question generation from concept maps. *DD* 3:75–99.

Andrew M Olney, Whitney L Cade, and Claire Williams. 2011. Generating concept map exercises from textbooks. In *Proceedings of the 6th workshop on innovative use of NLP for building educational applications*. Association for Computational Linguistics, pages 111–119.

Andrew McGregor Olney, Arthur C Graesser, and Natalie K Person. 2012b. Question generation from concept maps. *Dialogue & Discourse* 3(2):75–99.

Shiyan Ou, Constantin Orasan, Dalila Mekhaldi, and Laura Hasler. 2008. Automatic question pattern generation for ontology-based question answering. In *FLAIRS Conference*. pages 183–188.

Andreas Papasalouros, Konstantinos Kanaris, and Konstantinos Kotis. 2008. Automatic generation of multiple choice questions from domain ontologies. In *e-Learning*. pages 427–434.

Oleksandr Polozov, Eleanor O'Rourke, Adam M. Smith, Luke S. Zettlemoyer, Sumit Gulwani, and Zoran Popovic. 2015. Personalized mathematical word problem generation. In *IJCAI*.

Vasile Rus, Zhiqiang Cai, and Art Graesser. 2008. Question generation: Example of a multi-year evaluation campaign. *Proc WS on the QGSTEC* .

Iulian Vlad Serban, Alberto García-Durán, Caglar Gulcehre, Sungjin Ahn, Sarath Chandar, Aaron Courville, and Yoshua Bengio. 2016. Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. pages 588–598.

Sylvie Szulman, Nathalie Aussenac-Gilles, Adeline Nazarenko, Henry Valéry Teguiak, Eric Sardet, and Jean Charlet. 2010. Dafoe: A platform for building ontologies from texts. In *EKAW 2010 Knowledge Engineering and Knowledge Management by the Masses*.

Xuchen Yao, Gosse Bouma, and Yi Zhang. 2012. Semantics-based question generation and implementation. *DD* 3:11–42.