Educational Question Answering Motivated by Question-Specific Concept Maps

Conference Paper · June 2015

DOI: 10.1007/978-3-319-19773-9_2

3 authors:

Thushari Atapattu
University of Adelaide

Katrina Falkner
University of Adelaide

Nickolas Falkner
University of Adelaide

Educational Question Answering Motivated by Question-Specific Concept Maps

Thushari Atapattu(✉), Katrina Falkner, and Nickolas Falkner

School of Computer Science, University of Adelaide, Adelaide, Australia
{thushari.atapattu,katrina.falkner,
nickolas.falkner}@adelaide.edu.au

Abstract. Question answering (QA) is the automated process of answering general questions submitted by humans in natural language. QA has previously been explored within the educational context to facilitate learning, however the majority of works have focused on text-based answering. As an alternative, this paper proposes an approach to return answers as a concept map, which further encourages meaningful learning and knowledge organisation. Additionally, this paper investigates whether adapting the returned concept map to the specific question context provides further learning benefit. A randomised experiment was conducted with a sample of 59 Computer Science undergraduates, obtaining statistically significant results on learning gain when students are provided with the question-specific concept maps. Further, time spent on studying the concept maps were positively correlated with the learning gain.

Keywords: Concept mapping · Educational question answering · NLP

1 Introduction

Question answering (QA) is a modern application of information retrieval where exact answers are returned as a result of questions submitted by humans in natural language. This is in contrast to the more typical approach in information retrieval systems or search engines, in returning a ranked list of relevant documents. QA systems have been developed for a range of contexts, including open-domain to answer general questions [1], and closed-domains such as medicine, sports (e.g. BASEBALL) and geology (e.g. LUNAR).

QA systems are well suited to the educational context [2], providing assistance where learners struggle to find answers or in addressing common misconceptions. According to Novak and Canas [3], text-based short answers have been shown to support short-term learning. There are concerns as to the effectiveness of text answers in achieving long-term learning goals, including meaningful learning. Concepts included in both the question expressed by the learner, and the text answer are not explicitly related to other concepts within the domain, meaning that 'obtaining text answers' through typical QA systems do not support effective construction of knowledge structures [3]. Additionally, the majority of the QA systems are supportive only of 'factoid' question types (e.g. list, definition) [1, 4] which aid lower levels of educational objectives [5].

C. Conati et al. (Eds.): AIED 2015, LNAI 9112, pp. 13–22, 2015.
DOI: 10.1007/978-3-319-19773-9_2

As an alternative, this paper focuses on returning knowledge organisation techniques, particularly concept maps as answers. To achieve this, a framework capable of automatically extracting concept maps from lecture slides was developed using natural language processing (NLP) techniques, enabling the use of auto-generated concept maps as a positive alternative to expert concept maps [6, 7]. Concept maps are effective educational tools which consists of concepts, connected by directed edges to form relations, employing a hierarchical organisation scheme with the most general concept at the top, and more specific concepts arranged below. This aids meaningful learning, as relevant prior knowledge is able to be integrated with new information [3].

Additionally, this paper investigates whether adapting the returned concept map to the specific question context provides further learning benefits. This research develops a framework enabling the automated extraction of question-specific concept maps to answer learner questions using NLP and graph theory techniques. These concept maps assist the learner to understand the interrelationships between concepts for parallel processing in contrast to the sequential nature of text-based answers.

2 Related Works

Question answering systems can be classified based on several factors including *question types* (e.g. factoid, opinion, casual) [8], *input type* (e.g. natural language text, spoken natural language), *expected answer type* (short text answers, paragraphs, semantic graphs) and *supporting context* (open-domain or closed-domain) [9]. Among them, this work focuses on *question type* and *expected answer type*, which we consider to be most useful within educational question answering (EQA).

Within the educational context, 'question type' is the category of the question to measure various skills of the learner, defined by the question stem (e.g. *what, why*) [8, 9]. Questions are commonly constructed according to learning objectives defined by taxonomies like Bloom's taxonomy [5]. A computer model for question answering called *Quest* was proposed by Graesser et al. [8] which simulates psychological aspects of human question answering. The validation of model using convergence score (close to 0) suggested that very few nodes of conceptual graphs are good answers to the given question.

AnswerArt [4] is among the few systems, apart from ours, which utilises knowledge organisation techniques to formulate answers. This provides text answers along with the lists of facts associated with the answer, a summarised paragraph and visual representation of the *knowledge source* using semantic graphs [4]. AnswerArt is not restricted to a specific domain, however is limited to a number of question types with *pre-defined question templates* such as Yes/No, list, and reason (why). This restricts the flexibility of formulating questions, however, their system not possess restrictions on vocabulary. Although semantic networks are effective in knowledge organisation, their familiarity among learners is relatively low. Our own study found that 82% (*n=56*) of participants have heard, previously used or currently using knowledge organisation techniques, however, among them, only 1 student (out of 56) has experience in semantic networks, while 91% have experience in either concept maps or mind maps.

3 Question-Specific Concept Map Extraction

This section presents the framework for extracting question-specific concept maps (QSCMap). Figure 1 illustrates an example of the process.

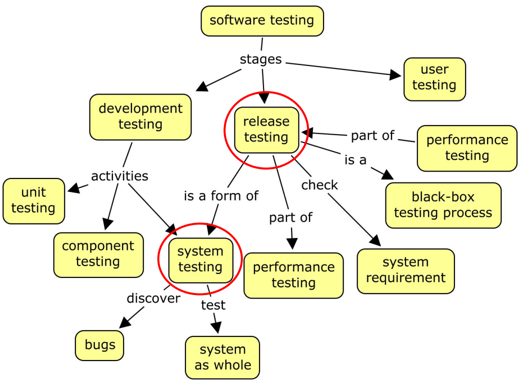
Question: Compare and contrast system testing and release testing

Sample text answers:

1. Release testing is a form of system testing

2. System testing focuses on discovering bugs while release testing checks that the system meets its requirement

(a)



(b)

Fig. 1. Example question a) text-based answer b) question-specific concept map

QSCMap (Figure 1(b)) provides additional information for the learner to effectively comprehend the relations between the concepts under assessment, and further, illustrates how these concepts are connected to the topic and other concepts.

Figure 2 illustrates the architecture of the framework including four main components: question processor, domain model, QSCMap extractor and visualisation.

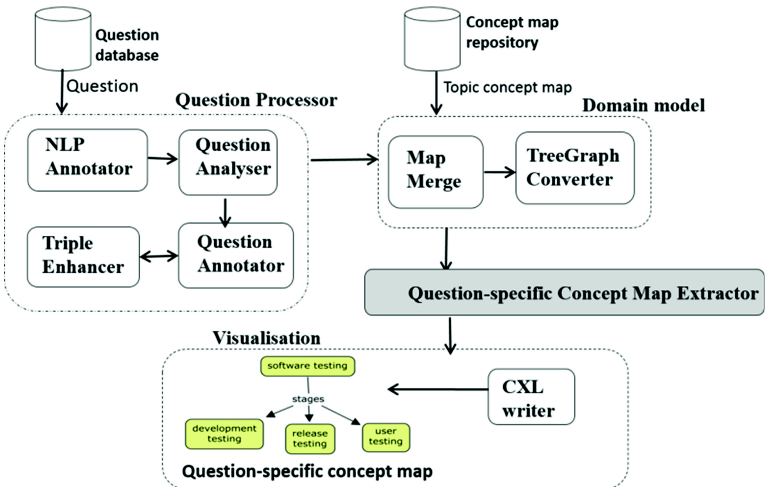


Fig. 2. Architecture of the question-specific concept map framework

Question Processor. identifies the 'question type' and converts the input question into 'triple' form. Within this work, we defined two question types: 'descriptive' and 'comparison', supported by a background study which analysed Software Engineering examination questions from year 2000 to 2012. The examinations consisted of 60 broad questions, the majority of which consisted of approximately 100 sub-questions. The study considered only the lecture material-based questions, with other types such as scenario-based eliminated.

NLP Annotator. parses the input question using Stanford NLP tools [10] and obtains part-of-speech tags, lemma annotations and parser tree (Table 1 and Figure 3).

Table 1. NLP annotations

Question	Part-of-speech	Lemma
What	WP	What
is	VBZ	be
the	DT	the
purpose	NN	purpose
of	IN	of
regrssion	NN	regression
testing	NN	testing

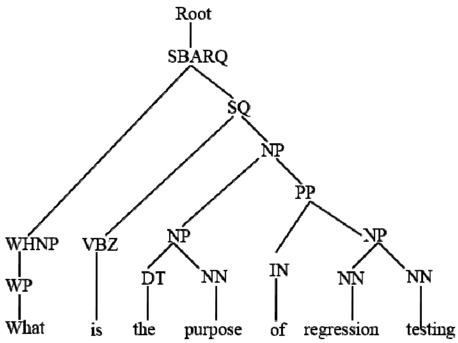


Fig. 3. Parser tree

Question Annotator. converts the question into 'triple' form (e.g. *what is white-box testing?* => (*white-box testing, is, ?*)) by considering the grammatical structure (Figure 3). This work reused some of the algorithms proposed to extract concept-relation-concept triples from English sentence in our own work [6] and other heuristics proposed by Dali et al. [4]. Unlike converting 'sentences' into triples which contains all the three elements (subject-verb-object), question triples can have one or more known elements, while other elements can be returned as the answer to the question.

Triple Enhancer. utilises the Synset library of WordNet [11] to find synonyms of non-terminological words (e.g. '*stage*' and '*phase*'). This arises when vocabulary used to write the questions is different from the terminology used in knowledge source of domain model (e.g. lecture slides).

Domain Model. consists of a repository of concept maps extracted from lecture slides [6, 7]. **Map merge** component automatically combines concept maps of related topics to create a richer domain model by reading the CXL (Concept Map Extensible Language) files [12] of each concept maps.

Due to the hierarchical nature of concept maps (i.e. taxonomic relations) and the inclusion of directed labelled edges between arbitrary pair of nodes (i.e. non-taxonomic relations and cross-links), a suitable data structure to store concept maps is a graph with a tree skeleton. **TreeGraph converter** reuses and customises the 'Tree-

Graph' data structure implemented by Stanford NLP group to store grammar trees [10].

Question-Specific Concept Map Extractor. is the core of the framework which extracts concept maps correspond to the input question. The process of QSCMap extraction is identified as a problem of 'sub graph matching' [13];

A data graph $G = (V, E)$, composed of a set of vertices V and a set of edges E. Each $e \in E$ is a pair (v_i, v_j) where $v_i, v_j \in V$.

A pattern ('question triple') $P = (Vp, Ep)$, which specifies the structural and semantic requirements that a sub graph of G must satisfy in order to match the pattern P.

The task is to find the set M of sub graphs of G that 'match' the pattern P. A graph $G' = (V', E')$ is a sub graph of G if and only if $V' \subseteq V$ and $E' \subseteq E$.

This work utilises the 'similarity-based exact sub graph' matching technique. This process is supported through a lemmatisation technique [10] where labels of concepts and relations in both question triple and domain model are mapped to their base form. Additionally, synonyms of the labels of relations are utilised [11]. This process maintains a threshold of 15 concepts based on the suggestions by Novak and Canas [3] and the feedback received from domain experts.

When determining the boundary of a sub graph, our algorithm considered the features listed in Table 2.

Table 2. Features for question-specific concept map extraction

Feature	**Description**
Question type	'Descriptive' or 'comparison'
In-degree and out-degree	Number of incoming and outgoing links of each node. This determines the importance of each node
Root	Most general node of the map
Leaf nodes	Nodes without children, generally these are the most specific nodes of the map
Number of overlapping nodes	This determines the boundary of the sub graph; Overlapping nodes should always be greater than 0. If this number is greater than 1, the boundary of the sub graph needs to accommodate all the overlapping nodes and its common parents, children and siblings
Number of overlapping relations	If relations in the question triple are not overlapping with the relations in the TreeGraph, synonyms are considered. The overlapping of relations are not mandatory, particularly in 'comparison' type questions
Distance between overlapping nodes	This considers the distance in every path including cross-links

The extracted sub graphs are converted to CXL format for **visualisation** using IHMC CMap Tools (see Figure 1 (b)) [12].

The benefits for learners who utilise this framework include the support for both interrogative (i.e. starts with *wh*-clause) and imperative (i.e. starts with words like *identify*) questions without restrictions in the grammatical structure. Additionally, this work derived two question types (descriptive and comparison) correspond to three objective levels of Bloom's taxonomy (i.e. knowledge, comprehension and analysis) [5].

Concept Map Mining Framework
Our system models the domain using concept maps generated from lecture slides (known as topic concept maps – TCMap) using the *concept map mining framework (CMMF)* [6, 7]. CMMF includes automated noise elimination (e.g. *course announcements, references*), resolution of syntactically and semantically missing and ambiguous sentences (e.g. pronouns, incomplete sentence fragments), useful knowledge acquisition in the form of concept-relation-concept triples using NLP-based algorithms, arrangement of extracted knowledge in a hierarchy, and ranking of concepts using structural and graph-based features (e.g. *term frequency, degree centrality, proximity*). The design and validation of CMMF is described further in [6, 7].

4 Evaluation

4.1 Method

A study was conducted with a sample of 59 second year Undergraduates in the University of Adelaide, who had enrolled in the Software Engineering (SE) course in Semester 1, 2014. A randomised experimental design was used with three treatment groups and a control group based on the *answer type* received by the participants. The Control group (LS) received text-based answers through a selected lecture slide segments. Treatment group 1 (TCMap) received a customised version of topic concept map with those concepts necessary for question answering, in addition to other related concepts. Treatment group 2 (HLCMap) received the same concept maps as group 1; however, in their maps, the context to answer the question is emphasised (manually using CMap tools [12]). Treatment group 3 (QSCMap) received the question-specific maps extracted from the proposed framework. The extracted QSCmaps were reviewed by two domain experts. Their feedback included manually removing some of the concepts from QSCmaps which did not illustrate a 'relation label'. This was an issue reported in CMMF, which occurred due to the point-based nature of lecture slides [7].

Ten multiple-choice questions (MCQs) were constructed by the researchers with the use of previous examinations and a text book of SE. These questions were provided to the participants in order to reduce the issues arise from forming questions by learners including incorrect terminology and grammatical issues. These questions covered 68% of the important concepts included in the 'software testing' topic.

A web-based prototype was developed for the experiments. Participants attempted each MCQ once before getting the answers through the QA system. This score was recorded as students' prior knowledge on the subject matter (pre-test). If the answers to the first attempt is incorrect, participants are expected to learn the required knowledge through the resources allocated for them and re-attempt any number of times within the allocated time for the study. Their scores, time spent on each resources (i.e. time spent between '*getting help*' and '*go back to question*') and attempts were collected for a quantitative data analysis.

After two week gap, a post-test was conducted to measure the learning outcome. Post-test was paper-based, consisting 10 questions with a combination of MCQs, fill-in-the-blanks and open-ended questions in order to minimise the memorisation of answers from previous study. These questions were constructed from similar topic, covering similar concepts; however, with a different presentation. Among the 59 participants of the main study, only 30 students were able to participate to the post-tests.

After the experiments, participants were requested to complete a questionnaire consisted of open-ended and close-ended questions.

4.2 Quantitative Analysis

The results of Table 3 illustrates that the QSCMap group has the numerically highest mean for *learning gain* (*M = 5.0, SD = 1.5, n = 9*) while the control group (LS) has the smallest mean (*M = 2.5, SD = 2.0, n = 7*)

Table 3. Descriptive statistics of pre-test and learning gain

Group	Pre-test M (SD)	Learning gain M (SD)	n
LS	2.9 (1.9)	2.5 (2.0)	7
TCMap	1.8 (1.5)	3.2 (1.4)	7
HLCMap	2.6 (1.5)	3.6 (2.0)	7
QSCMap	2.5 (1.5)	5.0 (1.5)	9
Total	2.4 (1.6)	4.0 (1.9)	30

Analysis of Variance (one-way ANOVA) was conducted to compare means of learning gain. Prior to that three assumptions of ANOVA were evaluated. The homogeneity of variances, $F\ (3,26) = .698,\ p = .562\ (>.05)$ was not violated. Shapiro-Wilk test indicated that the dependent variable was normally distributed around means. Due to the randomised assign of participants into groups, there was no influence of the performance of an individual in one group to the others in the same group. Results of the one-way ANOVA (Table 4) indicated that the means between groups were significant; $F(3,26) = 3.103,\ p = .044\ (< .05),\ \eta^2 = .263$.

Table 4. Summary results of one-way ANOVA

	Sum of Squares	df	Mean square	F	Sig.
Between Groups	28.292	3	9.431	3.103	.044
Within Groups	79.008	26	3.039		
Total	107.300	29			

Table 5. Summary of Tukey HSD post-hoc test

(I) Group	(J) Group	Mean Difference (I-J)	Std. error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
LS	TCMap	-.71429	.93178	.450	-2.6296	1.2020
	HLCMap	-1.14286	.93178	.231	-3.0582	.7724
	QSCMap	-2.55556*	.87849	.007	-4.3613	-.7498
TCMap	HLCMap	-.42857	.93178	.649	-2.3439	1.4867
	QSCMap	-1.84127*	.87849	.046	-3.6470	-.0355
HLCMap	QSCMap	-1.41270	.87849	.120	-3.2185	.3931

*. The mean difference is significant at the 0.05 level

Tukey HSD post-hoc test was conducted to compare the mean differences (Table 5). According to the results, means between QSCMap and LS groups are statistically significant; $p = .007$ ($< .05$). This suggests that the use of question-specific concept map-based answers is beneficial for learners in contrast to text-based answers. However, due to the smaller sample size in each group ($n = 9$ in QSCMap and $n = 7$ in LS), these findings cannot be generalised to a wider population [14]. Even though the effect of QSCMap has not been studied previously within EQA, utilising knowledge organisation techniques such as concept maps or knowledge maps over text-representations proved to be beneficial in many studies [15].

The means between QSCMap and TCMap groups are also statistically significant; $p = .046$ ($< .05$). This suggests that the use of *question-specific concept map*-based answers have higher learning gain than the topic concept map-based answers. The primary reason for this could be the amount of information included in each answers and their relevancy to the context. According to Novak & Canas [3], concept maps constructed to answer a question is more effective than a concept map which represent a domain/topic. The former involves more dynamic thinking and a deeper understanding.

However, the means between QSCMap and HLCMap are not statistically significant. This could be due to the fact that students in the HLCMap group might only have looked at the highlighted area of the map without being overloaded by the number of concepts and relations provided in the *topic concept map*. The idea of highlighting the relevant context is further supported in the feedback of students.

Pearson correlation analysis was conducted to measure the correlation between the time spent on each form of resource and the learning gain. There was a positive correlation between time spent on concept maps and learning gain; QSCMap ($\gamma = .801, p < .05$) and TCMap ($\gamma = .594, p < .05$). However, control group had a negative correlation. This could have occurred if the students in the control group spent more time scrolling the lecture slides to formulate an answer when the relevant information is scattered throughout the slides.

4.3 Qualitative Analysis

Participants were questioned about the issues and suggestions about the system. Students in the TCMap group mentioned that "*too much of information in concept maps*", "*they were kind of bland, it is difficult to navigate when more and more information added to the map*". Similarly, they suggested "*use colors to help identifying important sections*", "*improve appearance by providing partial concept maps that applies to the topic*", "*need colors and switches*", "*less concepts*", "*a way to toggle between maps of higher and lower densities of information*", "*color codes or smaller maps*". Since the participants of this group had no idea about maps in HLCMap or QSCMap groups, they repeatedly mentioned the requirement of colors to differentiate information or smaller maps to focus more relevant information.

The participants in the HLCMap group had minor issues or suggestions such as "*concept maps were useful for hints*", "*useful if ability to search within the map*", "*a feature in which you can click on a concept to retrieve more information*", "*more details in relation labels*".

Some students in the QSCMap group reported that "*not enough information provided in the concept map*", "*it was quite good*". However, similar to HLCMap group, they suggested to have more explanations in the concept map by allowing them to click the concepts to retrieve further information.

5 Conclusion

The use of question answering systems to automatically answer questions is a widespread area of research in NLP and information retrieval. The adoption of the QA to facilitate learning requires wide focus in the AIED research. This paper presented an approach to utilise concept maps in contrast to text-based answers for the questions presented to EQA. Additionally, this paper investigated whether adapting the returned concept map to the specific question context provides further learning benefit. We have obtained significant results on learning gain when presenting question-specific concept maps in contrast to lecture slides or topic concept maps. Further, time spent on concept maps were positively correlated with the learning gain.

The future works include expanding the system to support more question types with higher-level of learning objectives [5, 8]. In addition, questions and their corresponding QSCmaps generated from our framework are expected to evaluate using different dimensions (e.g. *coverage, correctness, pedagogy*) in multiple CS subjects with the use of two to three domain experts in order to calculate *inter-rater reliability*. More sophisticated natural language understanding techniques are required to accept questions from learners or integration of question generation is necessary in contrast to expert-constructed questions utilised in the study. Based on the results and the feedback obtained from the students, it is uncertain whether the students' preferred type of concept maps are QSCMaps or HLCMaps. Therefore, the current study can be expanded using a larger student cohort to find the answer to this.

Although the research discussed here does not support adaptive question answering through student modeling, the concept maps adapted to the context of questions improved the learning experience. Therefore, future research within the AIED community can be focused on adaptive question answering with the use of suitable form of concept maps.

References

1. Katz, B. Annotating the world wide web using natural language. In: Fifth RIAO Conference on Computer Assisted Information Searching on the Internet (1997)
2. Gurevych, I., et al., Educational question answering based on social media content. In: Proceedings of the 2009 conference on Artificial Intelligence in Education, p. 133-140 (2009)
3. Novak, J.D., Canas, A.J.: The Theory underlying Concept maps and How to construct and use them. Florida Institute for Human and Machine Cognition (2006)
4. Dali, L., et al.: Question answering based on semantic graphs. In: Language and Technology Conference, Poznan (2009)
5. Bloom, B., et al., Taxonomy of educational objectives: the classification of educational goals. In: Handbook I: Cognitive Domain, New York, Toronto (1956)
6. Atapattu, T., Falkner, K. Falkner, N.: Acquistion of triples of knowledge from lecture notes: a natural langauge processing approach. In: 7th International Conference on Educational Data Mining, London (2014)
7. Atapattu, T., Falkner, K., Falkner, N.: Evaluation of Concept Importance in Concept Maps Mined from Lecture Notes: Computer vs Human. In 6th International Conference on Computer Supported Education, Barcelona (2014)
8. Graesser, A., Gordon, S., Brainerd, L.: QUEST: A model of question answering. Computers & Mathematics with Applications **23**(6–9), 733–745 (1992)
9. Kolomiyets, O., Moens, M.: A survey on question answering technology from an information retrieval perspective. Information Sciences **181**(24) (2011)
10. Manning, C., et al.: The stanford CoreNLP natural language processing toolkit. In: 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore (2014)
11. Miller, G.A., et al.: Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography **3**(4), 235–244 (1990)
12. Canas, A.J., et al.: CmapTools: a knowledge modeling and sharing environment. In: First International Conference on Concept Mapping, Pamplona (2004)
13. Gallagher, B.: Matching structure and semantics: A survey on graph-based pattern matching. AAAI FS **6**, 45–53 (2006)
14. Kenny, D.A.: Statistics for the Social and Behavioural Sciences (1987)
15. O'Donnell, A.M., et al.: Knowledge Maps as Scaffolds for Cognitive Processing. Educational Psychology Review **14**(1), 71–86 (2002)