Automatic Diagrammatic Multiple Choice Question Generation from Knowledge Bases

Khalil Bouzekri, Liu Qiang, Husam N. Yasin, Benjamin Chu Min Xian, Dickson Lukose
Artificial Intelligence Department, MIMOS Berhad, Kuala Lumpur, Malaysia
khalil.ben@mimos.my, qiang.liu@mimos.my, husam.yasin@mimos.my, mx.chu@mimos.my, dickson.lukose@mimos.my

Abstract—**In this paper, we present a methodology to generate diagrammatical multiple choice questions from a knowledge base. When considering a knowledge base, the main strategies discussed in the literature are class-based, property-based and terminology-based, and the generated multiple choice questions are typical with all the choices (correct answer and distractors) being of atomic types such as plain text. In this paper, we introduce graph-based strategies, enabling the generation of choices (correct answer and distractors) in the form of complex structures such as diagrams, and discuss different approaches to take into account difficulty levels when generating the questions.**

Keywords-Question Generation; MCQ; Diagram; Graph; Question Difficulty; Knowledge Base; Ontology; Islamic Finance.

I. INTRODUCTION

A Multiple Choice Question (MCQ) comprises of a short text describing the question, and a number of alternative choices as answer, where usually one of the choices is the correct answer and the others are wrong alternative choices called distractors. MCQs are popular in e-learning systems due to several characteristics: simplicity of generation (header and choices), ease for scalability (systems can generate a large number of different MCQs), and automation and objectivity of the assessments.

The use of the semantic web to generate questions has been studied at length [1][2][3][8][9]. Indeed, domain ontologies can be considered a proper formalism as the basis for automatic generation of MCQs. Ontology contains domain knowledge in the form of concepts, instances and relationships between concepts and/or instances. Moreover, the concepts and relationships are structured in a hierarchical manner based on their semantics. Papasalouros et al. proposed an ontology-based approach to automatically generate MCQs [1]. They used Natural Language Processing (NLP) techniques to generate the question header, and a series of strategies based on the structure of the ontology (class-based, property-based and terminology-based) to generate the correct answer and the distractors. Their strategies were further enriched and implemented as a plugin in Protégé [2]. Cubric and Tosic extended the work by considering new elements such as annotations, and leveraging on question templates instead of NLP to generate the question header [3].

However, all the research has been focusing on the generation of "simple" MCQs, which means MCQs containing a question header and choices (correct answer and distractors) in an atomic format (e.g., plain text, numbers). To the best of our knowledge, more complex structures as choices (e.g., diagrams) have not been studied when automatically generating MCQs. Therefore, the main novelty of our research work resides in two elements: firstly, we propose a more complex type of MCQ that we call "Diagrammatic MCQ", in which the correct answer is a graph composed of labelled nodes and arcs (e.g., a graph representing a diagram) generated from a knowledge base, and the distractors are generated by combining existing strategies from the literature and new strategies relying on the structure of the graph itself; secondly, we discuss the difficulty levels of the different question generation strategies.

The paper is structured as follows: Section 2 recalls basic notions utilized throughout the paper. In Section 3, we introduce the diagrammatic MCQs and discuss the utilized strategies. Section 4 discusses the difficulty of generated questions. Finally, the prospects of this work are outlined in Section 5.

II. BASIC NOTIONS

In this paper, we use the Islamic Finance Knowledge Base (IFKB) from [4], which comprises a total of 2281 RDF triples [7]. An RDF triple consists of three components: subject, predicate and object. The predicate expresses a relationship between the subject and the object. For example, the relationship between "*Deposit*" and "*CashInvestment*" can be represented using the following triple: "*Deposit subClassOf CashInvestment*" where "*Deposit*" is the subject, *subClassOf* is the predicate and *CashInvestment* is the object (this triple can be read as "Deposit is a specialization/kind of CashInvestment"). Islamic Finance concepts are organized in a hierarchy, as shown in Figure 1: the concept "*Bank*" is a subclass of "*Party*" (also read as "*Bank*" is a child of "*Party*", "*Party*" is a parent of "*Bank*"), and the siblings of "*Bank*" are "*Bank Client*", "*Partner*" and "*Supplier*".

 ISBN: 978-1-61208-386-5

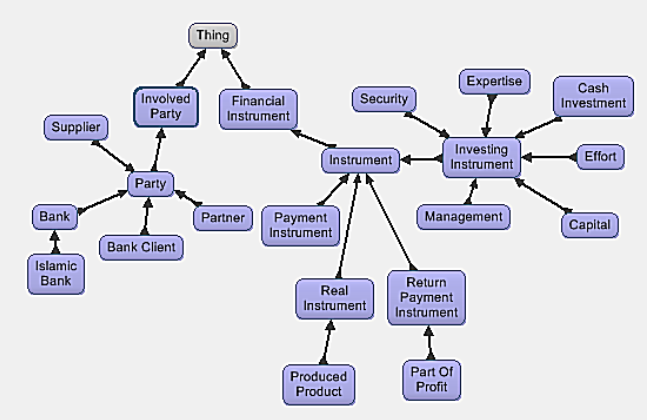

Figure 1. Excerpt of IFKB Concept Hierarchy

IFKB also contains 45 Islamic Finance contract models represented as sets of RDF triples. An Islamic Finance contract model shows the parties involved in the contract, the financial instruments utilized, as well as the process flow. Figure 2 illustrates the Salam contract model in IFKB (rendered using Gruff tool [6]):

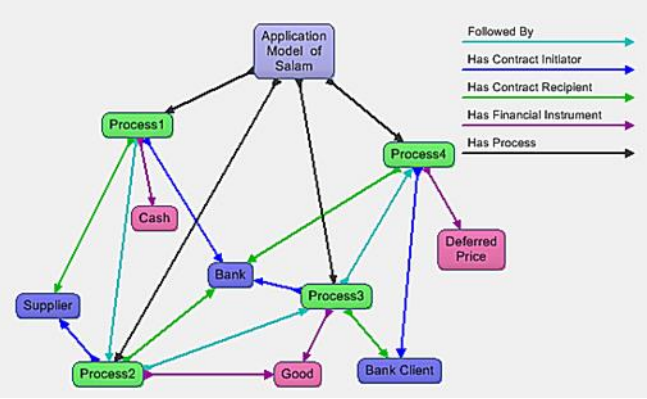

Figure 2. Salam Contract Model in IFKB

There are four processes, where Process1 is followed by Process2, then Process3 and subsequently Process 4. In this application model, the Bank plays the role of a buyer/purchaser as well as seller. The financial instruments used are "*Cash*", "*Deferred Price*" and "*Good*", and the involved parties are "*Bank Client*", "*Bank*" and "*Supplier*. They are all concepts from IFKB. As an example, a triple is "*Process1 HasContractInitiator Bank*". Figure 3 shows the Salam contract model in a diagrammatical form, which is the visualization used to display the DMCQs to the user.

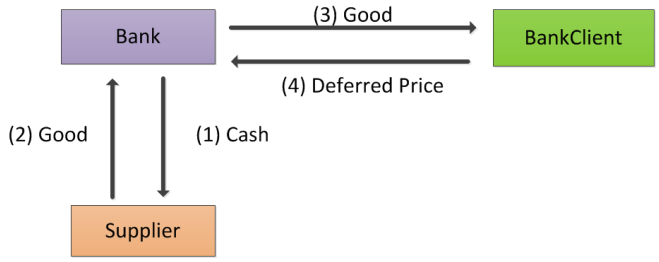

Figure 3. Salam Contract Model in Diagrammatical Form

Salam contract model (lit: forward payment) is the sale of a deferred item in exchange for an immediate (forward) price [5].

III. DIAGRAMMATIC MULTIPLE CHOICE QUESTION GENERATION

Diagrams are useful mechanisms for information summarization and displaying relationships between objects. They are used to explain concepts and amplify understanding. In this section, we present four types of diagrammatic multiple choice question (DMCQ), and discuss the strategies involved to generate these DMCQs such as choosing a question header, choosing a correct answer by extracting a group of concepts and relationships representing a contract model from a knowledge base, and forming the distractors. In the following, we use the Islamic contracts of IFKB to build DMCQs.

A. Diagrammatic Multiple Choice Question

A DMCQ consists of a question header, a correct answer and distractors (wrong answers). In the following, each DMCQ will have one correct answer and three distractors. In DMCQ Type 1, the correct answer is a set of nodes or arcs from the diagram. DMCQ Type 2 is based on conventional fill-in-the-blank questions, where one or more labels, nodes or arcs will be removed from the correct diagram and the user needs to select the correct answer from the given choices to fill in the blank. The correct answer of the DMCQ Type 3 is the original diagram from the domain knowledge base. Finally, the correct answer of DMCQ Type 4 is a partial diagram which is extracted from the original diagram. For each DMCQ type, the distractors are derived from the correct answer using various strategies. The distractors generation strategies and difficulty levels are discussed in Section III. D.

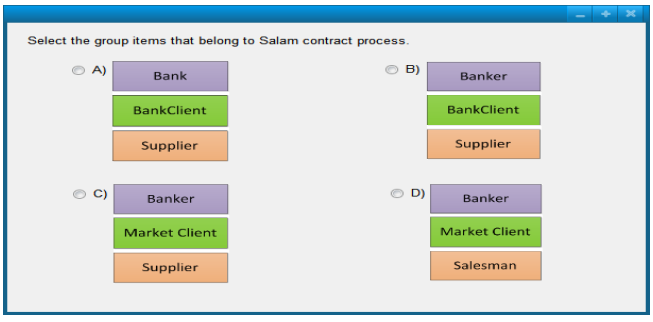

Figure 4. DMCQ Type 1 Sample

Figure 4 shows a sample of DMCQ Type 1, where the question header is "Select the group items that belong to Salam contract process", and the possible answers are sets of parties. Choice "A" is the correct answer and choices "B", "C" and "D" are distractors.

 ISBN: 978-1-61208-386-5

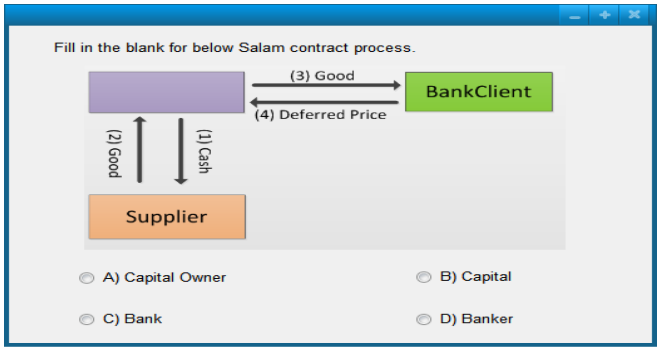

Figure 5. DMCQ Type 2 Sample

A sample of DMCQ Type 2 is shown in Figure 5. The party label "Bank" has been removed from the diagram representing the Salam contract model. In this sample, the modified diagram is part of the question header. Question choice "C" is the correct answer, and "B", "C" and "D" are distractors.

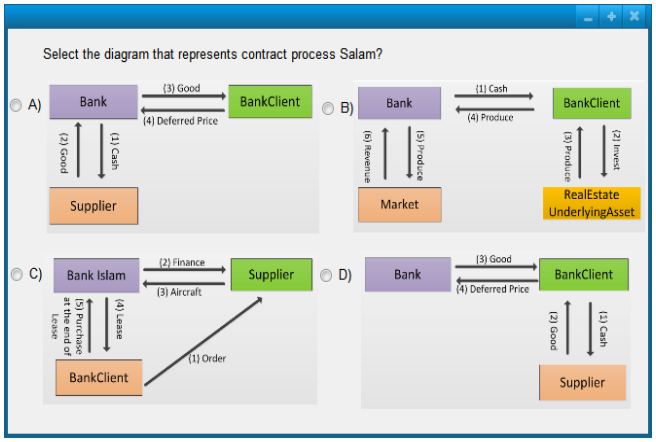

Figure 6. DMCQ Type 3 Sample

Figure 6 shows a sample of DMCQ Type 3, where the correct answer is the diagram of the Salam contract model (choice "A"), and "B", "C" and "D" are the distractors.

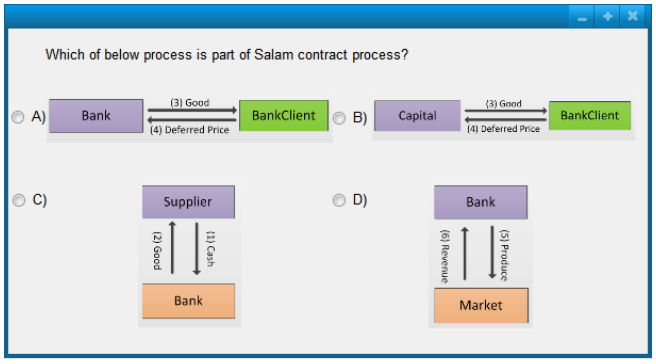

Figure 7. DMCQ Type 4 Sample

DMCQ Type 4 question sample is shown in Figure 7. The correct answer is a subset of the full diagram of Salam contract model (choice "A"), and choices "B", "C" and "D" are distractors.

B. *Question Header Generation*

As discussed in Section 1, using predefined question header templates allow the multiplication of generated questions. We use the same method, as shown in Table 1. An example of question header template is "*Select the group items that belong to <Concept X>*", where <Concept X> is called the *missing concept*, and is instantiated during the question header generation by selecting a contract model concept from IFKB. The selected concept is called the *target concept*, and it will be used to retrieve/create the correct answer and the distractors in the next steps.

For example in Figure 4, the chosen question header template is "*Select the group items that belong to <Concept X>*", and the target concept is "*Salam Contract Model*".

TABLE I. EXCERPT OF QUESTION TEMPLATES

Types	Question Templates
DMCQ Type 1	• Which of the following group items belong to <Concept X>? • For the below choices, which one can be considered as part of < Concept X>? • Select the group items that belong to <Concept X>
DMCQ Type 2	• Fill in the blank for below diagram representing <Concept X>. • For the diagram below representing <Concept X>, please fill in the blank. • Fill in the blank with the correct answer for the diagram below representing <Concept X>.
DMCQ Type 3	• Select the diagram that represents <Concept X>. • Which of the following diagrams represent <Concept X>? • Please select the diagram that represents <Concept X>.
DMCQ Type 4	• Which of the below process is part of <Concept X> process? • Select the diagram which is part of<Concept X> process • Please select the diagram which is part of<Concept X> process.

For DMCQ Type 2, the question header also contains a modified diagram (see Figure 5), where parts of the original diagram have been randomly removed.

C. *Correct Answer Generation*

The correct answer of DMCQ Type 1 is generated by randomly choosing a subset of nodes/arcs from the original diagram. For DMCQ Types 2 and 3, the correct answer is straightforward, respectively the removed labels/nodes/arcs and the complete diagram. Finally, the correct answer of DMCQ Type 4 is a randomly chosen subset of the original diagram.

D. *Distractors Generation*

When generating the distractors, we have to consider two modifier strategies: labelling and structural strategies. They respectively concern the labels of the nodes and arcs,

 ISBN: 978-1-61208-386-5

and the structure of the contract model chosen as target concept. The labelling strategy is similar to existing class-based, property-based and terminology-based strategies [1], and consists in modifying the labels of nodes and/or arcs. The structural strategy modifies the correct contract model to create a distractor by adding new nodes and/or arcs, modifying direction of arcs, modifying the positioning of nodes, as well as removing nodes and/or arcs. By considering labelling as well as structural strategies, the number of possible generated distractors is growing immensely. The utilized strategies are summarized as follows:

- Strategy 1: Replacing a node label (resp. an arc label).
- Strategy 2: Inserting new nodes and/or arcs that do not belong to the correct contract model.
- Strategy 3: Modifying the direction and/or numbering of arcs.
- Strategy 4: Modifying the positioning of nodes by swapping nodes
- Strategy 5: deleting nodes and/or arcs belonging to the correct contract model.
- Strategy 6: Selecting a wrong contract model.

E. Difficulty of Distractors for each Strategy

In this research, we consider the generated distractors to be the core criterion for determining the question difficulty. The results of our preliminary investigation are as follows:

- In Strategy 1, the replacement label makes the question more difficult if its meaning is close to the replaced label, which is if it comes from a sibling or a direct parent in the hierarchy.
- In Strategy 2 (resp. 5), inserting (resp. deleting) *n* nodes/arcs makes the question easier as *n* grows bigger, since it totally alters the original diagram.
- Strategy 3 (resp. 4) introduces difficulty, as the diagram must be well understood in order to remember the direction as well as the numbering of the arcs (resp. the positioning of the nodes).
- Finally, Strategy 6 makes the question more difficult if the selected contract model is close in structure to the correct contract model.

IV. CONCLUSION

In this paper, we presented a methodology to automatically generate DMCQ from a knowledge base, and discussed our preliminary investigations on the difficulty of the generated questions. We introduced four types of DMCQs, and new generation strategies based on the structure of Islamic finance contract models in the form of diagrams.

In our future work, we plan to evaluate and extend our current prototype in several prospective directions: (1) evaluate the quality of the generated questions with regard to several criteria such as pedagogy; (2) pursue the investigation on question difficulty; (3) enrich the question templates; (3) extend the types of generated questions to free-hand drawing, and automatically assess the drawings.

REFERENCES

[1] A. Papasalouros, K. Kanaris, and K. Kotis, "Automatic Generation Of Multiple Choice Questions From Domain Ontologies," presented at the IADIS e-Learning conference, Amsterdam, The Netherlands, 2008, pp.427-434.

[2] M. Tosic and M. Cubric, "SeMCQ – Protégé Plugin for Automatic Ontology-Driven Multiple Choice Question Tests Generation", presented at the 11th International Protégé Conference, Amsterdam, The Netherlands, 2009.

[3] M. Cubric and M. Tosic, "Towards automatic generation of e-assessment using semantic web technologies". In Proceedings of the 2010 International Computer Assisted Assessment Conference, University of Southampton, July, 2010.

[4] A. Mamadolimova, N. Ambiah, and D. Lukose, "Modeling Islamic finance knowledge for contract compliance in Islamic banking", Proc. 15th international conference on Knowledge-based and intelligent information and engineering systems, Springer-Verlag, Volume Part III, 2011, pp. 346-355.

[5] W. Al-Zuhayli, "Financial Transactions in Islamic Jurisprudence" (originally alFiqhu al-Islāmī wa adillatūhu). Translated into English by Mahmoud Amin El-Gamal. Syria: Dār al-Fikr, 2003.

[6] Gruff: A Grapher-Based Triple-Store Browser for AllegroGraph. retrieved January, 2015, from: http://franz.com/agraph/gruff/.

[7] Concepts and Abstract Syntax, retrieved January, 2015, from: http://www.w3.org/TR/rdf11-concepts/.

[8] M. Al-Yahya, "Ontology-Based Multiple Choice Question Generation", The Scientific World Journal, 2014.

[9] M. Teitsma, J. Sandberg, M. Maris, and B. Wielinga, "Using an Ontology to Automatically Generate Questions for the Determination of Situations", 22nd International Conference, DEXA, 2011.

 ISBN: 978-1-61208-386-5