

Can We Use LOD to Generate Meaningful Questions for History Learning?

Corentin Jouault^{*1}Kazuhisa Seta^{*1*2}Yuki Hayashi^{*2}^{*1} Graduate School of Science, Osaka Prefecture University^{*2} College of Sustainable System Sciences, Osaka Prefecture University

This research aims to clarify the potential of LOD (Linked Open Data) as a learning resource. In this paper, we describe a method that uses LOD to generate content-dependent questions in the history domain. Our method combines history domain ontology to LOD to create questions about any historical topic. To prove whether the generated questions have the potential to support learning, a human expert conducted an evaluation comparing our automatically generated questions with questions generated manually. The results of the evaluation showed that the generated questions could cover most (87%) of the questions supporting knowledge acquisition generated by humans. In addition, we confirmed that the system can generate questions that enhance history thinking of the same quality as human generated questions.

1. Introduction

The current state of linked open data (LOD) provides a large amount of content. It is possible to access semantic information about many domains. In this paper, we aim to verify whether it is possible to generate meaningful content-dependent questions that support history learning in an open learning space by using the current state of LOD sources.

Questions from the teacher are an important and integral part of the learning to deepen learners' understanding [Roth 96]. More specifically, in the history domain, asking questions to learners encourages them to form an opinion and reinforce their understanding [Husbands 96]. Learners also naturally ask questions to themselves during their learning. However, they cannot always generate good questions by themselves [Otero 09].

Because the quality of the learning is dependent on the quality of the questions [Bransford 99], asking good questions is important for performing satisfying learning. Learners are required to generate good questions to perform good quality of learning. This is one of the difficulties of learners performing their learning by themselves.

Our approach to solve this problem is to support learners with automatically generate meaningful questions depending on the contents of the documents studied by each learner. Our question generation method adopts a semantic approach that uses the LOD and ontologies to create content-dependent questions. Our function is part of a novel learning environment [Jouault 13] that aims to provide meaningful support in history learning about any historical topic.

The first issue to be clarified to build the question generation function, which is applicable to an open learning space, is (a) how we build scalable and reliable knowledge resource based on LOD and (b) how we build history dependent question ontology to generate content-dependent questions.

The second issue to be clarified is whether the quality of the questions generated by the system is sufficient to support history learning. We must evaluate the function before using it because,

as we mentioned, the quality of the learning is dependent on the quality of the question generated by the system.

2. Question Generation

2.1 Integrating LOD

The problem to be solved is to build a reliable knowledge resource for history learning that has both semantic information and natural language information. Natural language information is required for learners to use as learning materials. The requirements for the knowledge sources are:

- (a) **Unity:** The semantic information should closely represent the contents of the natural language documents.
- (b) **Scalability:** The natural language and semantic information should cover most historical important events that the learners may want to study.
- (c) **Reliability:** The semantic information represented should be reliable enough to provide meaningful concept instances and context information to learners.

Regarding (a), for the natural language learning materials, we selected Wikipedia because its fast evolution and growth give reliable information about a huge number of topics. An advantage of selecting Wikipedia is that two semantic knowledge resources, DBpedia [Bizer 09] and Freebase [Bolacker 08], are available. Both projects aim to create a semantic copy of the knowledge on Wikipedia; thus, the requirement of (a) is satisfied by using Wikipedia and its semantic resources.

On both sources, the information differs because DBpedia's information is automatically generated by analyzing Wikipedia and Freebase's information is provided by human users.

Figure 1 shows an example of the relation instances defined for two concept instances: "First Battle of the Marne" and "1st Army (German Empire)." These concept instances appear in red. The upper part of the figure shows the graphical representation of the relation instances defined for the two concept instances for Freebase (on the left) and DBpedia (on the right).

On one side, for major topics, the quality of the information from Freebase is better than the information from DBpedia. The first concept instance "First Battle of the Marne" is an interesting topic for many people. Thus, a lot of relation instances have been

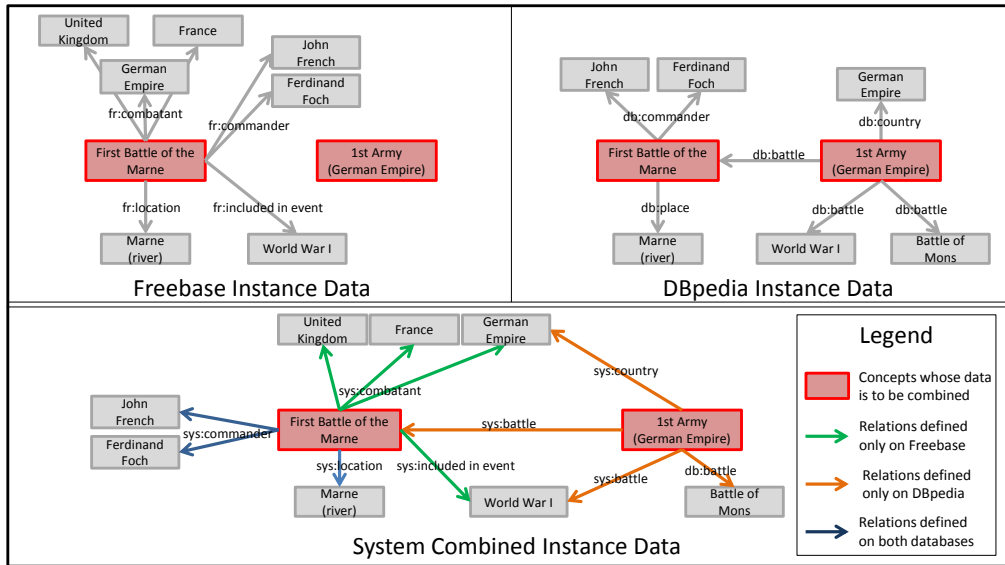


Fig. 1. Combination of Freebase and DBpedia information for two concepts (Concept map built by the system)

defined on Freebase. DBpedia also has relation instances, but not as many as Freebase.

On the other side, the information from DBpedia is also necessary because more minor topics have little information on Freebase. The second concept instance “1st Army (German Empire)” is minor for most learners, although advanced learners might have interest. For this concept instance, the advantage of DBpedia is particularly visible since Freebase did not define any relation instances about it.

By following the above consideration, we developed a method to satisfy the (b) Scalability and (c) Reliability requirements that combines the two semantic information resources. The bottom part of Fig. 1 shows the combined relation instances with color depending on their provenances. To combine the information, the system uses its own history domain ontology to recognize equivalent types from Freebase and DBpedia. The details are out of scope of this paper.

2.2 Question generation ontology

The system requires an understanding of a proper question’s structure to generate meaningful history domain questions. To understand the structure and function of a question, we refer to Graesser’s taxonomy [Graesser 10] to build an ontology for the history domain. This taxonomy describes domain-independent question types that are meaningful to support learning.

Figure 2 shows the History Dependent Question Ontology (HDQ Ontology) and examples of the natural language questions generated. The upper part of the Fig. 2 shows the HDQ Ontology which is divided in two parts. The left part shows the definition of the domain independent question concept classes based on Graesser’s taxonomy. The right part shows the history domain question concept classes we defined.

The HDQ Ontology specifies that there are four major categories of questions defined by Graesser, i.e. (a) ‘Description Question’, (b) ‘Method Question’, (c) ‘Explanation Question’ and (d) ‘Comparison Question’. Furthermore, more specialized question concept classes are defined and organized hierarchically, e.g. ‘Concept Comparison’ question, ‘Judgement’ question and

‘Improvement’ question are defined as a subclass of the class (d) ‘Comparison Question’.

Each definition of the question concept class in the HDQ Ontology specifies the relation among the concept (relation) classes specified in the history domain ontology and the natural language patterns (NLP). Each NLP specifies a template of a natural language question for each question concept class and is used to generate the natural language text of the question. For example, the history domain question concept class ‘Concept Completion location of MilitaryConflict’ subclass of the domain independent question type ‘Concept Completion’ associates the concept class ‘MilitaryConflict’, the relation class ‘location’ and the NLP “Where did \$x take place?” which means that, to create the natural language question, the \$x marker is replaced by the ‘label’ of a concept instance of ‘MilitaryConflict’ having a ‘location’ relation instance.

Currently, the HDQ ontology defines 28 history domain questions. Two kinds of questions can be generated:

- A) **R&C based Question:** Relation and Concept based Question. These questions require a concept instance with a relation instance. The answer to this type of question is identified based on a triple described explicitly.
- B) **C based Question:** Concept based question. These questions require only a concept instance. These questions ask even about information not explicitly described in the documents, thus, learners need to think about their knowledge to answer.

The table at the bottom of Fig. 2 shows examples of questions for R&C based question (first line) and C based question (second line). In both cases, the natural language questions are generated by filling the history domain concept instance and relation instance, which satisfy the constraints of a question concept class, into the NLP. For example, in the case of the C based question of Fig. 2, the NLP “How would \$MAIN_TOPIC have been different with \$x?” is filled by replacing the \$x and \$MAIN_TOPIC markers by respectively the name of the concept instance “Wilhelm II, German Emperor” and the name of the main topic of study, in this case “WWI (World War I).”

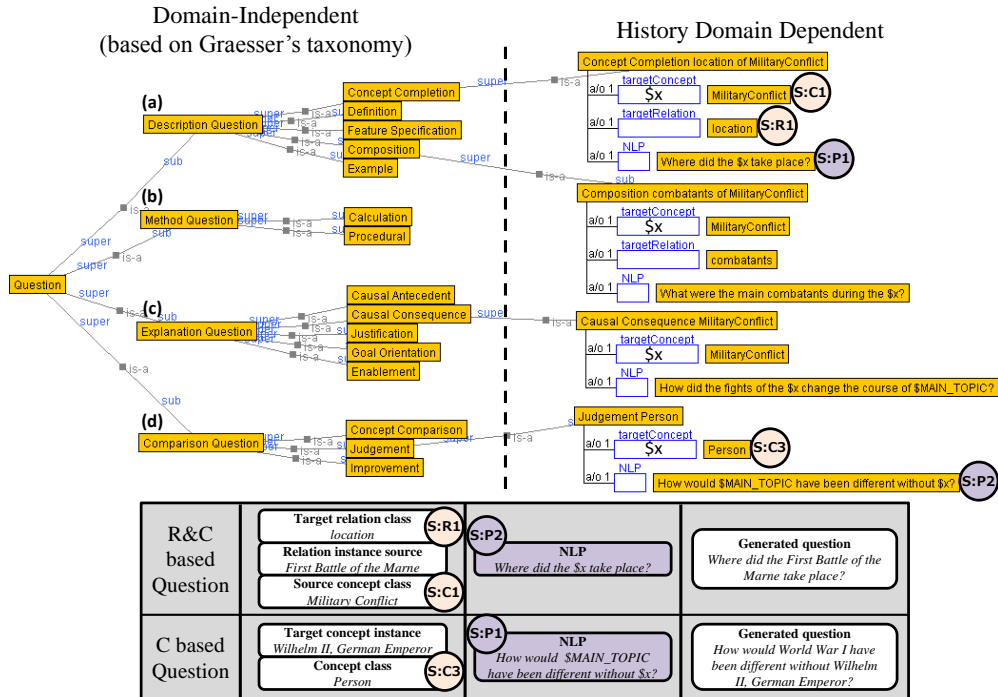


Fig. 2. History dependent question ontology and examples of natural language question generated

3. Evaluation Method

For this evaluation, we asked a history professor to compare the quality of questions according to his own criteria. More detailed information about the evaluation is as follows.

- **Topic: WWI.** The current version of the system can generate history domain questions for any topic. For this evaluation, the topic is set to WWI.
- **Source of human generated questions: SparkNotes** [SparkNotes Editors 14], a popular website, whose main target users are junior high-school and high-school students. The questions for our experimental study were taken from the multiple choice quiz and essay questions. We used all of the 58 questions about WWI in SparkNotes. 38 questions had an answer written explicitly in the learning materials can be compared to the R&C based questions. The remaining 20 questions did not have a definite answer, like essay questions, but contributing to deepening history understanding can be compared to the C based questions.
- **Evaluator: A history professor in a university** with over 20 years of history teaching experience.

3.1 R&C based questions

In the case of the R&C based questions, the system generates them by using explicitly described semantic relation instances representing fact (predicate) knowledge. The size of a set of questions generated by the system depends on the amount of semantic relation instances represented. Therefore, it is important to evaluate the coverage of human questions by the set of system generated questions to confirm whether the question generation function can build/assess learners' fundamental knowledge.

The evaluation of the R&C based questions considered the 38 manually generated questions with explicit answer taken from the website SparkNotes. Each manually generated question was

paired with an automatically generated question. In total, 38 couples of questions containing each one manually and one automatically generated question were created for the evaluation.

The evaluator was asked to grade each couple from 1 to 4, where 1 means the knowledge required to answer is different and 4 means the knowledge needed is the same. The concrete criteria from the viewpoint of history learning are set by the evaluator.

3.2 C based questions

The purpose of evaluating the C based questions is to clarify whether the system can generate questions that trigger historically deep thinking by using explicitly represented domain knowledge in LOD and specific topic independent ontologies.

The evaluation considered the totality of the 20 manually generated questions with an answer not explicitly in the documents and 30 C based questions automatically generated by the system. The automatically generated questions were generated by referring to the concept instances mentioned on the SparkNotes page "Key People and Terms" about WWI. Among 55 concept instances identified, 30 concept instances with the most semantic information were used to generate questions.

The evaluator was provided with the 50 questions in random order and instructed to categorize into 5 categories (C1-C5) the questions depending on their ability to deepen learners' understanding. C5 questions contribute to deepening the understanding of the learners, whereas C1 questions do not. The specific criteria were left up to the evaluator.

4. Results

4.1 R&C based questions

Table 1 shows the number of couples for each mark. The evaluator described the criteria for attributing grades as follows.

- 1: Both questions require different knowledge to be answered.

Table 1. R&C based Question Evaluation Results

Mark	1	2	3	4	Total
Number of couples	5 (13%)	15 (39%)	14 (37%)	4 (11%)	38

- 2: Both questions focus on different parts of the target relation instance, but they require the same knowledge to be answered, e.g. “Who won the Battle of the Falkland Islands?” (Human) and “What was the result of the Battle of the Falkland Islands?” (System).
- 3: Both questions assess the same knowledge from different viewpoints (they require the same knowledge to be answered), e.g. “Who assumed power in Germany and led negotiations with the Allies after Wilhelm II lost power?” (Human) and “Who succeeded Wilhelm II, German Emperor?” (System).
- 4: Both questions have the same meaning.

The evaluator judged that couples marked 2, 3, and 4 require the same knowledge for junior and high school students to be answered. In total, under the conditions of the system, 87% of the manually generated questions could be covered by the system.

For 39% (15/38) of the couples, the questions did not have a difference in quality. The system generated question was better in 29% (11/38) of the couples, and the human generated one was better in 32% (12/38) of the couples. These results show no significant difference between the quality of the system and human generated questions.

As a result, the question generation function has potential to generate useful questions to construct learners’ basic knowledge. Furthermore, it suggests that even R&C based question might be useful even for learners studying history in a university.

4.2 C based questions

Table 2 shows the number of questions for each category. The result shows that, even if the system cannot generate questions of the highest quality (C5), it can generate questions of the same quality as the manually generated questions in weighted average.

The criteria defined by the evaluator were:

- C1) Questions asking facts or yes/no questions.
- C2) Questions asking causal relations.
- C3) Others (more complex than C2 but does not require integrated knowledge).
- C4) Questions requiring integrated knowledge of the topic.
- C5) Questions requiring a deep historical or political thinking.

The system only uses information specific to one topic, i.e., using one of the key concepts defined in SparkNotes, to generate a question in this experiment. Although it depends on the templates that were used by the system, the majority of the questions are of reasonable quality (C3 or C4).

As a result, the question generation function has potential to generate useful questions to reinforce learners’ deep understanding of historical topics.

5. Concluding Remarks

In this paper, we described a method for generating questions automatically by using LOD. The history dependent question ontology makes possible to generate content dependent questions using domain dependent but content independent question concept

Table 2. C based Questions Evaluation Results

	C1	C2	C3	C4	C5	W. Avg.
Human (n=20)	1	5	8	3	3	3.1
System (n=30)	1	1	19	9	0	3.2

classes. The generated questions can be expected to support learners in acquiring basic knowledge and deepening their understanding of history. The system has been implemented and can generate questions about any historical topic.

The evaluation showed that the system could generate good quality questions by using the current LOD. The experimental results showed that the questions generated by the system can cover the majority of the questions generated by humans. In addition, the questions enhancing history thinking generated by the system and by human were of the same average quality.

In addition, by considering the system can generate much more questions not appearing on SparkNotes and its adaptability, the results described in this paper seems quite meaningful in the situation of individual learners’ support.

References

- [Bizer 09] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia-A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3), 154-165.
- [Bollacker 08] Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 1247-1250.
- [Bransford 99] Bransford, J. D., Brown, A., & Cocking, R. (1999). How people learn: Mind, brain, experience, and school. *Washington, DC: National Research Council*.
- [Graesser 10] Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question?. In *Bringing reading research to life*. Guilford Press.
- [Husbands 96] Husbands, C. (1996). *What is history teaching?: Language, ideas and meaning in learning about the past*. Berkshire: Open University Press.
- [Jouault 13] Jouault, C., & Seta, K. (2013). Wikipedia-Based Concept-Map Building and Question Generation. *The Journal of Information and Systems in Education*, 12(1), 50-55.
- [Otero 09] Otero, J. (2009). Question generation and anomaly detection in texts. *Handbook of metacognition in education*, 47-59.
- [Riley 00] Riley, M. (2000). "Into the Key Stage 3 history garden: choosing and planting your enquiry questions." In: *Teaching History*, London: Historical Association
- [Roth 96] Roth, W. M. (1996). Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student responses. *Journal of Research in Science Teaching*, 33(7), 709-736.
- [SparkNotes Editors 2014] SparkNotes Editors. “SparkNote on World War I (1914–1919).” SparkNotes LLC. 2005. <http://www.sparknotes.com/history/european/ww1/> (accessed December 12, 2014).