Wikipedia-Based Concept-Map Building and Question Generation

Corentin Jouault* and Kazuhisa Seta*

(Received 7 June 2013 and accepted in revised form 7 January 2014)

Abstract This research aims to give learners more content-dependent scaffolding in the self-directed learning of history. Learners use a system to build a concept map containing a chronology. The system is able to generate content-dependent support adapted to the learners. For this support, we intend to build a semantic open learning space using a natural language online encyclopedia and semantic information using the open linked data. The support is provided by the automatically generated questions according to a learner’s request by referring to the concept maps of the learner and the system. The generated questions aim to lead learners to new knowledge, deepening their understanding.

Keywords: question generation, adaptive learning support, open linked data, concept map building

1. Introduction

Our research objective is to build a learning environment to support learners during self-directed exploratory learning in an open learning space. By open learning space, we mean the WWW as a learning space in contrast to a closed learning space where learning materials are limited and arranged by authors and teachers. Thus, learners in the closed learning space can follow learning processes designed by authors of learning materials to achieve learning goals, whereas learners in the open learning space need to set their own learning goals and make plans by themselves during learning, which requires their self-regulated skills. We call the latter learning style in the open learning space self-directed exploratory learning. If the learners’ skills are insufficiently developed, however, the learning will be of a lesser quality than classroom learning, and learners will waste a considerable amount of time doing undesirable self-exploration in their learning processes. On the other hand, the advantage of self-directed exploratory learning is that the learners will be more motivated than they are with classroom learning. Learners can proceed at their own rhythm and take more time to study the concepts in which they are interested.

Previous research has already created systems to overcome the disadvantage such as the Navigation Planning Assistant (1), which provides an environment used to describe learners’ learning plans and the state of their understanding to prompt their self-regulation in an open learning space. The limitation of this system, however, is that its support is content independent due to the difficulty of working with natural language information on the WWW. Of course, we overcome the difficulty when we can prepare learning materials and texts in advance. Teachers, in principle, cannot regulate the learning materials in self-directed exploratory learning: if teachers specify well-designed learning materials, self-directed learning loses its meaningful advantages. Another notable related research study was on the Kit Build Method (2), which provides a knowledge externalization environment for building a concept map and providing support during the concept map construction. However, as the learning material (concept map kits) needs to be prepared beforehand, this requires a considerable amount of time even for constructing the closed learning space. The underlying difficulty of this is also that the system cannot use the semantic information to prepare the domain concept structure of a target field.

Therefore, our approach for building a system able to generate content-dependent support in an open learning space is to use semantic information. More concretely, we have developed a question generation function that can generate content-dependent questions to prompt learners’ setting of meaningful learning goals and the exploration to achieve the goals in an open learning space.

Our question generation function is built based on Wikipedia for the natural language information and en-
enhanced by semantic information using open linked data (3). We call this semantically enhanced WWW a semantic open learning space: the difference between an ordinary open learning space and the semantic one is that the information in the latter one is represented in RDF in order for it to be machine understandable. Thus, the system can understand the contents stored in it. In our research, we use Freebase and DBpedia, whereby the clone information of Wikipedia is stored, for constructing the semantic open learning space.

Consequently, our research is positioned as a challenge to get a technical breakthrough that provides the content-dependent question generation function by using semantically enhanced information to avoid working with natural language information.

2. Prototype of the System

The system has three main windows: the question window in Figure 1(a), the document window in Figure 1(b), and the concept map window in Figure 1(c). The question window contains the list of questions generated by the system and the questions already answered by the learner. The document window displays the document answering the selected questions.

In our research, the learner is given the task to build a timeline of the events of the studied period with causal relationships between these events. Our system provides a specific knowledge externalization environment to improve the understanding of chronology depicted as Figure 1(c). To learn history in a satisfying way, learners need to understand the relations between the events (4). They must study the events as a whole, not every event separately. The learner’s concept map is designed on this principle.

On the other side, we can recognize a concept map built by learners as an externalized representation of their internal understanding state. We use it as a medium for supporting their self-direction by giving learners a chance to promote checking of their own understanding state, because appropriate self-direction for learning requires appropriate monitoring of the individual’s own internal understanding state.

In the center of the concept map in Figure 1(c) is a timeline of the events ordered by time as usual, but learners also need to add relations between these events as well as other related concepts. All non-event concepts are displayed around the timeline. The motivation for this is to enhance causal understanding of the historical epoch events according to the time series which are the backbone of history learning.

When using the system, all learners have the same starting point. They are given a document about the main subject of study. For this scenario, the studied sub-

![Figure 1. System Image.](image-url)
ject is World War I, and the starting document is the introduction of the Wikipedia page about WW I. The learners’ task is to create a timeline of the events of WW I. Thus, the interests of the learners influence their concept map, which is used to generate the questions to be answered (Figure 1(a)).

The concept map created will differ for each learner since all learners will perceive the importance of the concepts in accordance with their interests, and we do not regulate the parts used for constructing the concept map to enhance their self-exploratory learning.

Then, when the learner thinks all concepts considered important have been added to the concept map (Figure 1(c)), a list of questions from the system appears in the question window (Figure 1(a)). The timing of showing the questions is important for learners to become aware of the importance of questioning and answering activities.

However, it is notable that a question does not appear automatically but appears according to the learner’s request: if the system shows questions automatically without any request, the chance might be lost for learners to set their own learning goals by themselves which would result in losing the advantages of self-directed exploratory learning.

Figure 2 shows how the system extracts the information and uses it to build the concept map. The system accesses the information using SPARQL requests on a server with an endpoint for both Freebase and DBpedia. Figure 2(a) presents simple examples of the ontology. The history dependent types’ ontology in Figure 2(a) unifies the DBpedia and Freebase types. Figure 2(a) shows that the Event type is related to its equivalent on Freebase and DBpedia, as well as their parameters, which are often different. The ontology only contains the parameters that are relevant for the history learning. The ontology is created by us by merging the DBpedia and Freebase ontologies, all the similar types are associated to a new equivalent type used only by our system.

The ontology is domain-dependent for history learning only. The ontology is content-independent and only has to be created once by making a correspondence among concepts and relations specified in the DBpedia and Freebase ontologies. The creator does not need to be an expert to make the correspondence in the ontologies and it just costs about 10 hours.

Figure 2(c) shows an example of a concept map built by the system, integrating these examples. Once the information has been extracted, the system generates concepts and relation instances using the semantic information. A simplified example can be seen for one concept and one relationship: the First Battle of the Marne and the relationship between the First Battle of the
Marne and France. The semantic information used can be found in Figure 2(b). The First Battle of the Marne is identified as an ‘Event’, so the system will extract its start and end dates to place it at the right position on the time axis of the timeline in Figure 2(c). For the relationship between the First Battle of the Marne and France, the type of relationship ‘Combatants’ and the orientation ‘France’ are extracted and inserted into the concept map. The concept map in Figure 2(c) shows an example of a concept map created using the starting document.

As a result, the system manages two concept maps: its own map and each learner’s map. The learners’ concept maps are modified by the learners during learning, and they can only see their own concept map, which is then compared with the system’s map to generate the questions as described in the next section.

4. Question Generation Support Using Semantic Information

To generate a question, the system uses history dependent question generation ontology and patterns to generate a natural language question. For the system, a question consists of two main parts: the type of question and the target concept. The ontology contains the following types of questions defined by Graesser et al. as shown in Figure 3.

These question types are domain-independent. Thus, variables that appear in questions are manually associated to a history-dependent type in the history-dependent types ontology to be adapted to the history learning. The system also can access patterns depending on the question type and the target concept type to generate natural language questions.

The system compares its concept map with the learner’s one and finds a target concept or relation for the question. The types contained in the ontology are the types of all the concepts under the WW I category on Wikipedia as well as all their related concepts. The ontology can easily be expanded for other types. The number of types stays manageable for our purpose as DBpedia’s ontology currently contains 359 classes with some categories not relevant for our system, such as fiction or sports. From the DBpedia and Freebase types, we create the ontology and associate the question pattern to them.

First, the two concept maps on top are compared, and the target concepts and relations are identified. The target concepts are determined by comparing the system’s and learner’s concept maps. All concepts have a single ID and all entries in the concept map are controlled so all the information present in the concept map
can be analyzed and understood by the system. The system finds all missing information and identifies the targets related to the unknown concepts. There are two categories of targets: the target concepts and target relations. The target concepts are those that have the most unknown related concepts. For example, the learner’s concept map contains only the Allies without its members. The learner needs to learn about the members of the Allies to improve his/her understanding. Then, the system will select the Allies as a target for a question, so the learner will be able to complete his/her concept map. The system uses the natural language pattern in the ontology and generates a question: e “What countries were the members of the Allies?”

The target relations are those that are important for the understanding. For now, they contain the relations between events and between an important event and another concept. For example, the learner does not have the reason for the war. The system will create a question to explore the relation between WW I and the assassination of Archduke Franz Ferdinand of Austria. The generated question will use WW I as a target, and answering the question will lead the learner to learn about the assassination of Archduke Franz Ferdinand of Austria. The system uses the natural language pattern in the ontology for a “reason for” with the target concept being an Event and the generated question being “What was the reason for World War I?” If the learner chooses this...
question from the list, the system will present a document describing the reasons for the World War I which will lead the learner to learn about the events leading to the war, one of the most important being the assassination of Archduke Franz Ferdinand of Austria.

When all the questions have been generated, the system displays them in a list and the learner can choose them. By using the targets in Figure 4, the generated questions will be:

- What countries were the members of the Allies?
- What countries were the members of the Central Powers?
- Who was involved in World War I?
- Where did World War I take place?
- What was the reason for World War I?
- What were the consequences of World War I?
- Where did the First Battle of the Marne take place?

5. Conclusion and Future Work

To provide content dependent support in the open learning space to support self-directed learning, techniques using semantic information are necessary. This paper describes a way to build such a space by using a natural language encyclopedia (Wikipedia) and semantic information using open linked data. We discussed how to use this space to provide meaningful and adapted support to learners with different profiles.

We will evaluate and improve the quality of questions through an experimental study in the future.

References

(1) Kashihara, A. and Taira, K.: “Developing Navigation Planning Skill with Learner-Adaptable Scaffolding”, in Proc. of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, pp. 433–440, IOS Press, Amsterdam (2009).
(2) Hirashima, T., Yamasaki, K., Fukuda, H. et al.: “Kit-build Concept Map for Automatic Diagnosis”, in Artificial Intelligence in Education, pp. 466–468, Springer, Berlin (2011).
(3) Heath, T. and Bizer, C.: “Linked data: Evolving the web into a global data space,” Synthesis Lectures on the Semantic Web: Theory and Technology, Vol. 1, No. 1, pp. 1–136 (2011).
(4) Stow, W. and Haydn, T.: “7 Issues in the Teaching of Chronology” in Issues in History Teaching, pp. 83–97, Psychology Press, Hove (2000).
(5) Bizer, C., Lehmann, J., Kobilarov, G. et al.: “DBpedia-A Crystallization Point for the Web of Data”, Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 7, No. 3, pp. 154–165 (2009).
(6) Bollacker, K., Evans, C., Paritosh, P. et al.: “Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge”, in Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, pp. 1247–1250, ACM, New York City (2008).
(7) Graesser, A., Ozuru, Y. and Sullins, J.: “What is a good question?” in Bringing Reading Research to Life, pp. 112–142, Guilford Press, New York City (2010).

Corentin Jouault received the M.E. in Computer Science from EISTI, France in 2010. He is currently a Ph.D. candidate at the Graduate School of Science, Osaka Prefecture University. His research interests include ontological engineering, artificial intelligence, intelligent tutoring systems and software engineering.

Kazuhisa Seta received the B.E. and M.E. from Ryukoku University in 1993 and 1995, respectively. He received the Ph.D. from Osaka University in 1998. He is currently a professor in the College of Sustainable System Sciences and the Graduate School of Science, Osaka Prefecture University. His research interests include software engineering, intelligent tutoring systems, human resource management, and ontological engineering. He received a Best Paper Award of the Japanese Society for Information and Systems in Education in 2012. He is a member of JSAI, IEICE, IPSJ, JSiSE, and IAIED.