Wong, L.-H. et al. (Eds.) (2013). Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

Building a Semantic Open Learning Space with Adaptive Question Generation Support

Corentin JOUAULT* & Kazuhisa SETA
Graduate School of Science, Osaka Prefecture University, Japan
*jouault.corentin@gmail.com

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. To enable this support, we built 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 and documents. The learners request questions when they need and the system will generate the questions depending on the concept map of the learner. The generated questions aim to leads the learners to new knowledge deepening their understanding.

Keywords: Semantic Open Learning Space, Self-directed Learning, Question Generation, History Learning, Adaptive Learning Support

1. Introduction

When learners are confronted with a self-directed situation, their interests will influence their learning. The advantage of this is that the learners will be more motivated than they are by with classroom learning. Learners can proceed at their own rhythm and take more time to study the concepts in which they are interested. However, the disadvantage is that to study in a self-directed way and reach their learning objective, learners must use their self-regulation skills (Biswas, Roscoe, Jeong and Sulcer, 2009). If these skills are insufficiently developed, the resulting learning will be of a lesser quality than classroom learning, and learners will waste a considerable amount of time trying to extract the information they need, especially in an open learning space where the information is not limited to the studied subject.

Previous research already created systems to overcome this disadvantage such as the Navigation Planning Assistant (Kashihara and Taira 2009), which provides an environment used to describe learners' learning plans and state of 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 Web. Of course, we overcome the difficulty when we can prepare learning materials in advance. Teachers, however, cannot regulate the learning materials in principle in self-directed 'exploratory' learning: if teachers specify the learning materials, self-directed learning loses its meaningful advantages. Other notable related research is Kit Build Method (Hirashima, Yamasaki, Fukuda and Funaoi, 2011), which provides a knowledge externalization environment for building a concept map and providing support during the concept map construction. Also notable is the Betty's Brain (Leelawong and Biswas, 2008) system which also uses concept map in an environment for learning by teaching. However, in both cases, as the learning material (kits of the concept map) needs to be prepared beforehand, this requires a considerable amount of time even for constructing the closed space of learning. 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. This is build based on Wikipedia for the natural language information and enhanced by semantic information using open linked data (Heath and Bizer, 2011) to make it a semantic open learning space. This research has two advantages for learning support:

- The system can provide content dependent questions in accordance with the learners' interests to deepen their understanding by enhancing their internal self-conversation.
- The system can provide suitable documents in accordance with the questions that learners try to answer and highlight the information on which they should focus. Even learners less skilled at self-regulated learning can continue motivated learning, since they are released from extracting suitable information from huge amounts of information.

Both for A) and B), one key issue is the adaptability to the learners' interests and learning topics. To realize the above advantages, we adopt ontology and a linked open data technique to eliminate the difficulty of the natural language understanding problem in the history domain. Then, the system can automatically construct respective concept structures of the learning topic in accordance with learners' circumstances.

Regarding A), the problem is that learners cannot always generate good questions (Otero, 2009). The quality of the learning depends on the quality of the questions during this process (Bransford, Brown and Cocking, 1999), it is important to support the learners' question asking and answering activities in the learner's internal self-conversation in self-directed learning. By answering good questions that lead to a deeper understanding, the learner will be motivated to pursue his/her learning. Thus, learners need to be able to generate good questions by themselves. However, learners without support tend to focus on their interests and may not explore others subjects. However, if the support ignores these interests, the motivation of the learner will be reduced. To make this self-conversation explicit, the system suggests a list of possible questions depending on the learner's situation represented by the concept map to help the learner's question asking activity[3].

Regarding B), to support learners without strong self-regulation skills, the system introduces semantic information into the documents. When studying an important number of different concepts, learners may have difficulties in managing new information. It can become difficult to extract the suitable resources to answer the current questions and separate the information in a document into what is already known and what is (ir-)relevant to answer the questions. The system provides documents with a dynamic enhancement based on Wikipedia text with semantic information. These documents can highlight on demand concepts that appear in the text.

In this paper, we mainly discuss the technical issues to overcome to enable semantic open learning space with adaptive learning support. After overviewing our system in Sect. 2, we describe how the system builds a concept map on the basis of the ontology and semantic database in Sect. 3. Concept maps constructed by the system and the learner play an important role in helping the learner navigate his/ her learning regarding both A) and B). Different learners have different learning situations and often do not share the same learning experiences and interests. If we want to take these differences into account, we have to provide them dynamic advice that adapts to suit the learner. Then in Sect. 4, we discuss the question generation and highlight functions in accordance with both concept maps without forcing learners to follow a fixed path.

[3] We also expect that it may enhance the learner's internalization of his/her question asking and answering activities, although this is out of the scope of this paper.

2. Overview of the System

2.1 Learning Activity on the System

The system has three main windows: the question window in Fig. 1(a), the document window in Fig. 1(b), and the concept map window in Fig. 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 Fig. 1(c). To learn history in a satisfying way, learners need to understand the relations between the events (Stow and Haydn, 2000). They must study the events as a whole, not every event separately. The learner's concept map is designed on this principle. In the center of the concept map in Fig. 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 backbone of history learning. The surrounding concepts contribute to deepening their causal understanding. The timeline has to contain the important events of the studied period, which are chosen by a teacher from a timeline of all events in the studied period generated by the system to limit the preparation time, but the learner is free to add every event he/she considers relevant. The required events are present in the concept map of the learner at the beginning but they appear in grey until the learner adds them from a document.

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 subject is World War 1, and the starting document is the introduction of the Wikipedia page about WW1. The learners' task is to create a timeline of the events of WW1. The period of

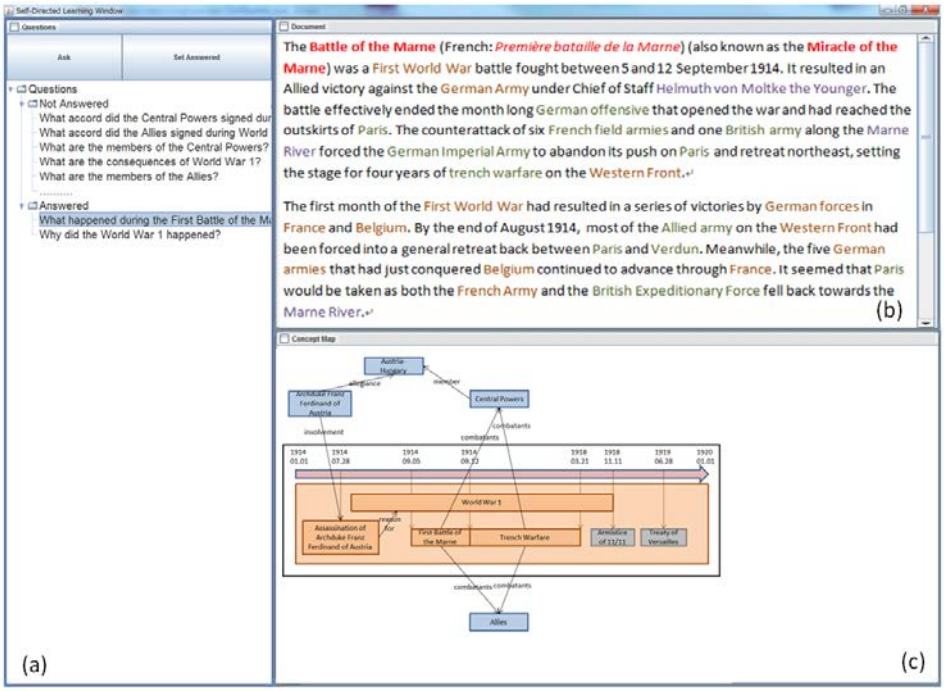

Figure 1. System Image

study we defined is from the beginning of 1914 to 1920 so the learners can also study the reasons for and consequences of the war.

When learning without support in self-directed exploratory learning, learners have to keep their focus on achieving their objective. If they do not, they will lose their way among irrelevant information. For example, when using Wikipedia, an unfocused learner will follow the links that most interest him/her and quickly move away from the domain of study. On Wikipedia, it is not difficult to start on an historical subject and end up on a page about sport or cooking in just a few clicks. It is also possible for learners to stay on topic but study documents of little importance. A learner could study about all the weapons used in WW1, which are still in the studied domain but do not help deepen the understanding of reasons of the War. While the interests of the learners are important for motivation, they often become a problem in self-directed open learning. Our system aims to give advice adapted to suit the learner. The interests of the learners influence their concept map, which is used to generate the questions to be answered (Fig. 1(a)).

The learning is divided into two mains step. First, the learners start their learning with a document given by the system in the document window, in Fig. 1(b). While working on a document, they update their concept map, shown in Fig. 1(c), with all concepts and relations they consider interesting. They can add concepts in Fig. 1(b) to the concept map Fig. 1(c) by selecting them in the document, shown in Fig. 1(b). To add relationships between concepts, learners need to select two related concepts in their concept map, shown in Fig. 1(c). Then the system will generate a list of possible types of relation which includes invalid answers: (i) if the system shows one valid answer, they can easily construct their concept maps without enough understanding, and (ii) the system can understand the meaning of each link chosen by learners.

The concept map created will be different depending on 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 shown in Fig. 1(c), a list of questions from the system appears in the question window, shown Fig. 1(a). The timing of showing the questions when they complete learning is important to become aware of the importance of questioning and answering activities. This list of questions will contain questions that will lead the learner to new relevant information that can deepen his/her understanding. The process used to generate this list is explained in section 4.1. The learner can also simply choose a question from the list without refreshing it. The list of questions is only refreshed when the learner requests it.

2.2 Inside of the System

The Fig. 2 shows the inside of the system in relation to a learner's activities described in the section 2.1. We adopt the server & client method to share ontologies among clients. Middle of the Fig. 2 shows the module structure in a client system. Top shows the ontologies and endpoints to retrieve semantic information from Freebase and DBpedia described below. The server stores semantic information once retrieved by a client.

When the learner needs a document, the system uses the resources of the server to generate a document with highlighted information.

The system's concept map is updated with every modification of the learner's one. The system's concept map will always contain more concepts than the learner's map. Every time the learner adds a concept to the map, the system will also integrate the concept to its map. Moreover, the system will also extract the semantic information about the concept, including other related concepts. For example, if the learner adds to

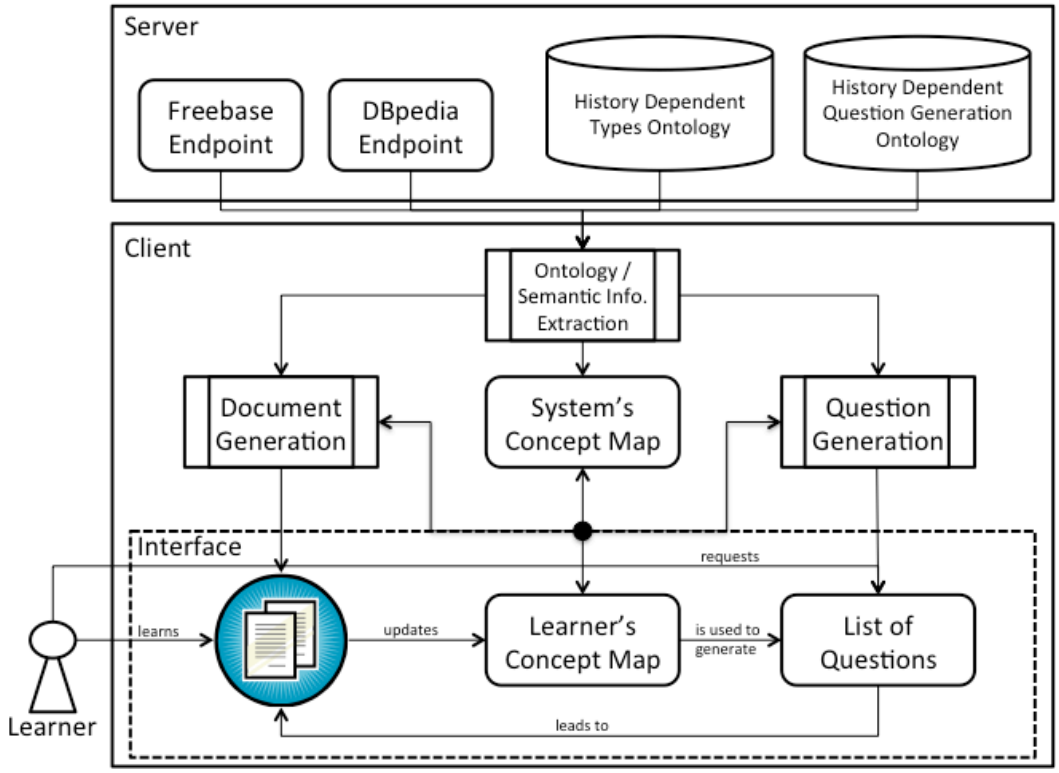

Figure 2. Structure of the system

his/her concept map the concept "Allies", the system will extract all the information on it. The system will then add its related concepts to the map. The map of the learner will only contain "Allies", although the system's concept map corresponding the learner's one also contains all the alliance members (e.g. France, United Kingdom, Russia…) and the battles (ex: First Battle of the Marne, Trench Warfare, Battle of Verdun…) in which the Allies fought, that the learner doesn't know yet. Thus, the role of the questions listed by the system is to lead the learner into discovering these concepts.

3. Concept Map Building based on the Ontology and the Semantic Database

Wikipedia is one of the largest information databases available online. It is also making efforts to add semantic information to the pages, though these pages are still not developed enough to be used in our system. The information present on Wikipedia is provided by users, so it can contain invalid information, but most of the information can be considered accurate. This is the case for most self-directed learning on the Internet. Wikipedia is always evolving: invalid information is regularly corrected, and missing information is regularly added. This makes it a powerful and reliable source of information that can be used as an open space for learning. A significant number of research projects are aiming to transform information from Wikipedia into semantic information. In our case, the system uses two of these projects as sources of semantic information: DBpedia (Bizer, Lehmann, Kobilarov, Auer, Becker, Cyganiak and Hellmann, 2009) and Freebase (Bollacker, Evans, Paritosh, Sturge and Taylor, 2008). The system uses DBpedia because the information is always up to date with the latest Wikipedia information. However, DBpedia contains many relations that are not meaningful for history learning and are difficult to manage. Freebase offers more meaningful relationships between concepts, but as the concept information is not always the same as on Wikipedia, there could be contradictions between the text and the

semantic information. The combination of DBpedia and Freebase is a satisfying information source for our system. Wikipedia is also translated into many languages, and the pages in different languages are related to each other, making multi-language use possible.

The system accesses the information using SPARQL requests on a server with an endpoint for both Freebase and DBpedia. The semantic information is then used to build the concept map as detailed in Fig. 3. Fig. 3(a) presents simple examples of the ontology. The history dependent types ontology in Fig. 3(a) unifies the DBpedia and Freebase types. 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. In the Fig. 3 are examples of the types association for a relation type on the right and for a concept type on the left. For every type used in the system, there is one Freebase type and one DBpedia type. Since the DBpedia information contains the Freebase URI, the system can identify the DBpedia and Freebase versions of an instance as the same concept.

Fig. 3(b) shows two examples of the semantic information of one instance of relationship and one instance of concept, and Fig. 3(c) shows an example of a concept map built by the system and integrating these examples. Once the information has been extracted, the system generates concepts and relation instances using the semantic information. The system prioritizes the concept information from DBpedia since it is updated at every modification of a Wikipedia document, so the text information and the semantic information will always be the same. For the relationships information, Freebase is prioritized since the relation information of Freebase contains less irrelevant 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 Fig. 3(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 Fig. 3(c). For the

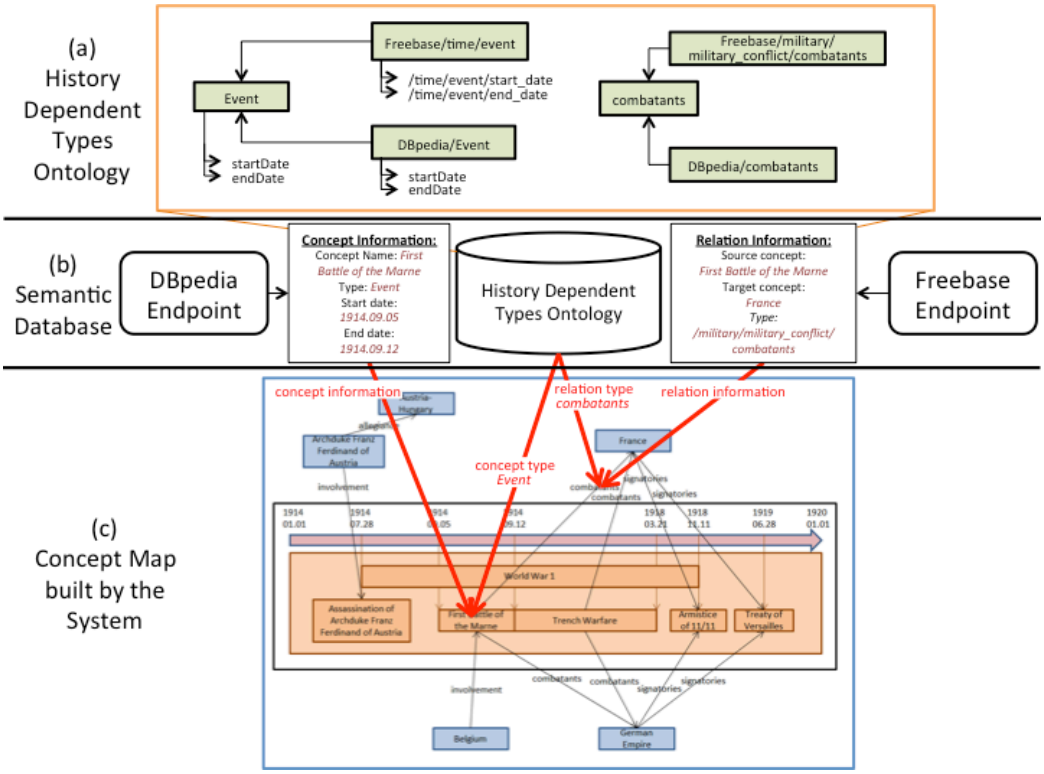

Figure 3. Information Extraction and Use

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 Fig. 3(c) shows an example of a concept map created using the starting document. The history dependent types ontology in Fig. 3(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.

As a result, the system manages two concept maps: its own map and the learner's map. The learner's concept map is 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 section 4.1.

4. Support Functions Using Semantic Information for Self-Directed Learning

4.1 Question Generation

To generate a question, the system uses history dependent question generation ontology and patterns to generate a natural language question. As opposed to Goto et al. (2010), our method generates the questions starting from the semantic information in the concept map and uses the natural language only to display to the learner. 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. (2010) as shown in Fig. 4.

These question types are domain independent. Thus, variables appears 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. These natural language questions are then displayed to the learner.

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 WW1 category on Wikipedia as well as all their related concepts. The ontology can easily be expended 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. The detailed process of the question generation can be seen in Fig. 5.

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 map. 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

- Description Questions
  - **Concept Completion**: *Who, what, when, where?*
  - **Definition**: *What does X mean?*
  - **Feature Specification**: *What features does X have?*
  - **Composition**: *What is the composition of X?*
  - **Example**: *What is an example of X?*
- Method Questions
  - **Calculation**: *Compute or calculate X.*
  - **Procedural**: *How do you perform X?*
- Explanation Questions
  - **Causal Antecedent**: *What caused X?*
  - **Causal Consequence**: *What will X cause?*
  - **Enablement**: *What enables the achievement of X?*
  - **Goal Orientation**: *What is the goal of X?*
  - **Justification**: *Why is X the case?*
- Comparison Questions
  - **Concept Comparison**: *Compare X to Y?*
  - **Judgment**: *What do you think of X?*
  - **Improvement**: *How could you improve upon X?*

Figure 4. Types of Questions referred to Build Ontology (Quoted from Grasser 2010)

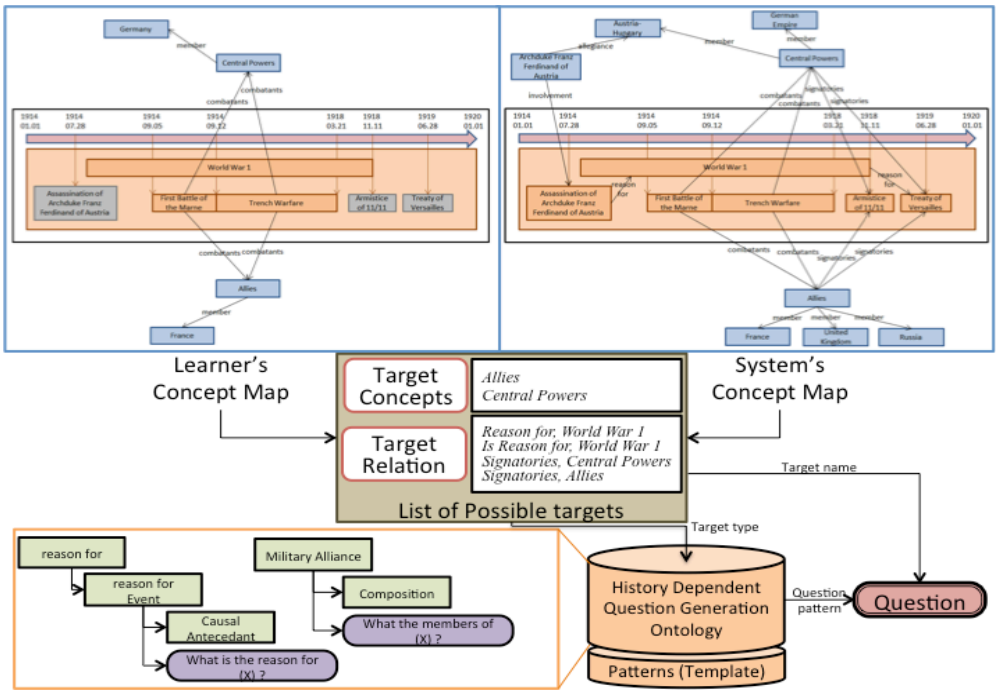

Figure 5. Question Generation Process

target concepts and target relations. The target concepts are those that have the most unknown related concepts. For example, in Fig. 5 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 will be "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, in Fig. 5, the learner does not have the reason for the war. The system will create a question to explore the relation between WW1 and the Assassination of Archduke Franz Ferdinand of Austria. The generated question will use WW1 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 1?" If the learner chooses this question from the list, the system will give him a document describing the reasons for the World War 1 which will lead him 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 Fig. 5, the generated questions will be:

What countries were the members of the Allies?
What countries were the members of the Central Powers?
What was the reason for World War 1?
What were the consequences of World War 1?
What accord did the Central Powers sign during World War 1?
What accord did the Allies sign during World War 1?

4.2 Dynamic Document Generation with Highlighted Information according to the Question

As the learning advances, learners will encounter more and more concepts, and it can become difficult for them to keep track of their knowledge if they do not have the necessary self-regulation skills. They have the concept map to help them, but judging the importance of the concepts in the document may become more difficult. The system provides a function to highlight the concepts in the document on demand, so learners who have difficulties in extracting the information will not be stopped in their learning. If a learner can already extract the information by him/herself, s/he will not need to use the function. Learners need to identify which concepts in the document they already know and which are related to the concept they are now studying. The dynamic document generation creates a document that uses a color code to represent the different kinds of concepts mentioned in the document. With the semantic information at our disposal, the information in the document can be made easier for the learner to extract.

The documents are enhanced with the gathered semantic information. The semantic information is used to make the important concepts appear in the document clearly. The process is described in Fig. 6. The learner's and system's concept map are used to categorize the concepts into four types: the main concept of the document, the concepts known by the learner, the concepts related to the main concept unknown to the learner, and the other mentioned concepts. To handle the natural language mentions of the concepts in the documents, the system uses alternates names gathered on DBpedia. The system uses the Wikipedia redirects as alternate names to identify the mention of the concepts in the document in natural language.

Fig. 6 shows the enhancement of the Wikipedia document about the First Battle of the Marne. The resulting document is on top, and the categorized concepts are shown under it. The process starts from a Wikipedia page about the main concept in the case of Fig.6 the main concept is the 'First Battle of the Marne.' The system identifies the concepts mentioned and organizes them in four categories: main concept, concept known by the learner, concept unknown by the learner but related to the main concept and the other mentioned concepts. Then, the system highlights all of the concept mentions in the

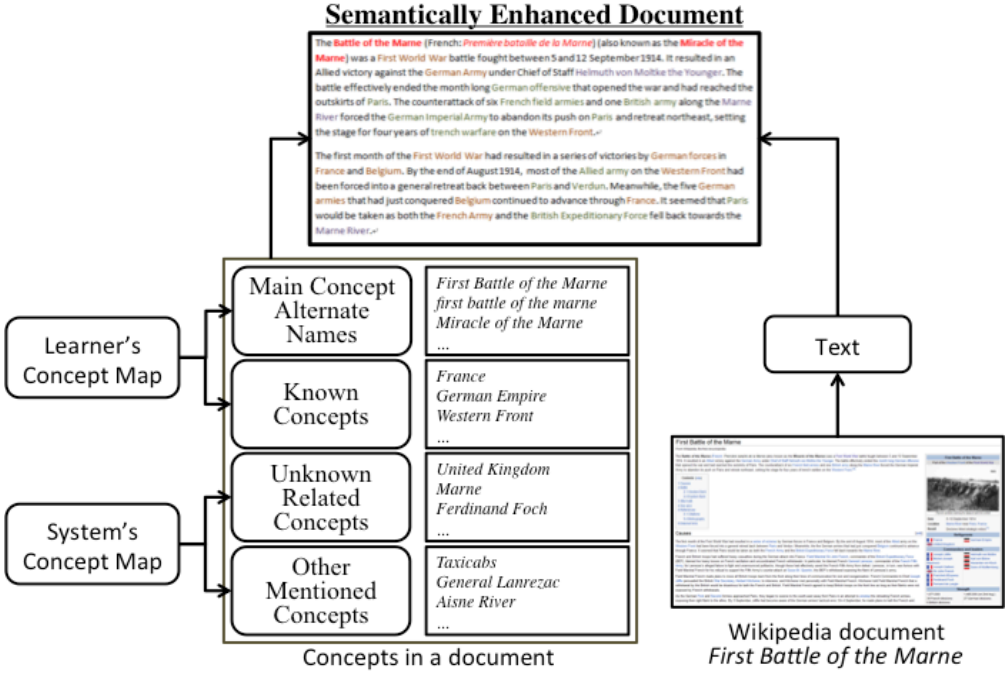

Figure 6. Dynamic Document Generation

document using a different color for each category.

5. Conclusion and Future Works

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 the open linked data. We discussed how to use this space to provide meaningful and adapted support to learners with different profiles.

The next step of this research is the evaluation. It will consist of two groups of learners. The first group will use the system, and the second will be a control group. First, the two groups will learn about the same subject, one group will use the system with advice and the other without advice. Both groups will have to create a timeline. Then in a second session, both groups will use the system without the advice. They will have to learn about a second subject by generating questions by themselves. They will have to use as many questions as possible to generate the best timeline they can. We will evaluate the quality of the timeline by taking into account its complexity and the density of the relations. We also want to evaluate the quality of the questions through the experiments. We will carefully address the specification of the ontology in another paper.

References

Biswas, G., Roscoe, R., Jeong, H., & Sulcer, B. (2009). Promoting self-regulated learning skills in agent-based learning environments. In *Proceedings of the 17th international conference on computers in education* (pp. 67-74).

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, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008, June). 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.

Bransford, J. D., Brown, A., & Cocking, R. (1999). How people learn: Mind, brain, experience, and school. *Washington, DC: National Research Council*.

Goto, T., Kojiri, T., Watanabe, T., Iwata, T., & Yamada, T. (2010). Automatic generation system of multiple-choice cloze questions and its evaluation. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, *2*(3), 210-224.

Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question?. In *Bringing reading research to life*. Guilford Press.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), 1-136.

Hirashima, T., Yamasaki, K., Fukuda, H., & Funaoi, H. (2011, January). Kit-build concept map for automatic diagnosis. In *Artificial Intelligence in Education* (pp. 466-468). Springer Berlin Heidelberg.

Kashihara, A., & Taira, K. (2009, July). Developing Navigation Planning Skill with Learner-Adaptable Scaffolding. In *Proceedings 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.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education*, *18*(3), 181-208.

Otero, J. (2009). Question generation and anomaly detection in texts. *Handbook of metacognition in education*, 47-59.

Stow, W., & Haydn, T. (2000). 7 Issues in the teaching of chronology. *Issues in history teaching*, 83.