CHAPTER 6

QUESTION GENERATION SYSTEM WITH SEMANTIC RELATIONSHIPS (Q-Genesis)

Question Generation system generates the reasonable questions from structured information and unstructured information. A wide range of questions can be generated based on the semantics of the information. The semantic relationships play a vital role in generating the questions. The knowledge level questions could not be generated without semantic relationships. In this work, semantic relationships based approach has been employed to generate the questions. Question generation system is used in intelligent tutoring system. Generally, the generated questions are widely used to improve the learner’s knowledge. Thus, the Question Generation system based on Semantic Relationships (Q-Genesis) has been proposed in this work. This chapter begins with the need of the proposed question generation system. The next section describes Q-Genesis system in detail.

6.1 NEED FOR Q-GENESIS

The task of generating questions is a tedious task for the trainer. Question generation system generates questions based on the learning materials to check the learner’s knowledge and also helps them to enhance their understanding about a particular field. It helps the trainer to generate various levels of questions which are intended to assess the knowledge level of the learners. Generally, the questions play a significant role in understanding the facts. The question generation system will be used to generate the various types of questions from the information provided. The questions generated by the subject experts and making entry by the data entry operator are time consuming and costly process. Hence, there is need of a system which generates the questions automatically. The variety of questions can be generated such as knowledge level, comprehensive level and application level, etc. Knowledge level questions are very popular for assessing the understandability level of the learner. The semantic relationship that exists between two entities plays a vital role in generating knowledge level questions. In this work, an effort has been made to generate knowledge level questions from structured information with the use of semantic relationships.
Q-Genesis helps to generate Bloom’s taxonomy level 1 and level 2 questions. It contains six levels of questions which incorporate knowledge, comprehension, application, analysis, synthesis and evaluation. The knowledge level questions usually required for the learners to understand the basic information in the same form as it was presented. It includes the questions such as know, how, list, what, when, where, list, name, etc. Comprehensive level questions are required to combine the knowledge together. The questions such as explain, discuss, compare and outline are included in comprehension. The learner has to apply their knowledge to a new situation in application level questions. It includes modify, apply, demonstrate, etc. Analysis requires the understanding of the content and structural information. It includes the questions such as how the entities are related, compare, contrast, why, how, etc. The synthesis level questions are a challenging one which expects the learners to develop new ideas and solve the problems. The questions such as revise, design, construct, plan, formulate, etc. are included in synthesis level. Evaluation questions require the judgment about something. The questions incorporated in this level are judge, assess, evaluate, etc. The next section explains Q-Genesis system elaborately.

6.2 QUESTION GENERATION SYSTEM BASED ON SEMANTIC RELATIONSHIPS (Q-Genesis)

Generally, the questions are not only used to assess whether the learners have learned something and also guides the learners to learn sufficient information. The questions are also used to impart the knowledge to the learners. The trainers spend a great deal of their time in testing the learners through the questions. Usually the trainers ask factual questions which depends on short-term memory. Although questions are widely used and assist in many tasks, trainers are likely to overuse the factual questions, literal and knowledge based questions. The knowledge level questions help the learners to think more creatively. The learners are constantly bombarded with questions; it makes them to think more innovatively and divergently. The generation of questions from various resources is a critical task for the trainer. Thus, Q-Genesis system helps to generate variety of superficial questions automatically with the use of the semantic relationships.

The architecture Q-Genesis is shown in Figure 6.1. In this approach, domain knowledge has been collected from web using crawler and stored in the database.
Generation of questions requires robust knowledge about the domain for which the questions to be generated. But given domain knowledge, it is quite impossible to generate questions automatically unless the knowledge can be represented in a proper way. In this approach, the domain knowledge is represented in the form of domain ontologies with the use of OWL parser. The ontologies represent concepts, semantic relationships and instances of the particular domain. The ontology is represented in OWL format. It is a standard Web Ontology Language based on the Description Logic representation formalism.

In Description Logic, the following notations have been used.

\[ C(a) : a \text{ is an instance of the concept } C \]

\[ R(a,b) : \text{instance } a \text{ is related to } b \text{ with the relationship } R \]

For example,

Leader (Mahatma Gandhi) - Mahatma Gandhi is a leader

Country (India) - India is a country

Situated_in (Gujarat, India) – Gujarat is situated in India
In this, Mahatma Gandhi, India and Gujarat are individuals. Leader and Country are concepts. Situated_in is a binary semantic relationship that exists between the individual Gujarat and India.

To generate the questions from the ontology, all the concepts, relationships and instances that are defined in ontology will be useful. The number of questions that can be generated is directly proportional to the number of concepts and semantic relationships represented in the ontology. The semantic relationships are not defined properly in the ontology; there is a possibility of generating the definitional questions only. The collateral semantic relationships are inferred with SemPER. The context specification is incorporated with ontologies to generate the superficial questions. The context specification is shown in Table 6.1.

**Table 6.1. Context Specification**

<table>
<thead>
<tr>
<th>Context</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>people</td>
<td>who</td>
</tr>
<tr>
<td>time</td>
<td>when</td>
</tr>
<tr>
<td>location</td>
<td>where</td>
</tr>
<tr>
<td>country</td>
<td>where, what</td>
</tr>
<tr>
<td>things</td>
<td>what</td>
</tr>
<tr>
<td>organization</td>
<td>where, what</td>
</tr>
<tr>
<td>animal</td>
<td>what</td>
</tr>
<tr>
<td>abstract</td>
<td>what</td>
</tr>
<tr>
<td>transfer</td>
<td>what, who, where</td>
</tr>
<tr>
<td>information</td>
<td>what, who</td>
</tr>
<tr>
<td>emotional</td>
<td>what</td>
</tr>
<tr>
<td>thought</td>
<td>who, what</td>
</tr>
<tr>
<td>event</td>
<td>what, where</td>
</tr>
</tbody>
</table>

The questions to be generated are used to assess the various knowledge levels of the learner. It follows Bloom’s Taxonomy. In Q-Genesis, the semantic based approach is followed for generating Bloom’s level 1 and level 2 questions. These questions are
generated based on the concepts, semantic relationships and instances. Three semantic based methods have been formulated to generate the questions. The approaches are:

- Concept based question generation
- Association based question generation
- Case based question generation

### 6.2.1 Concept Based Question Generation

In ontology, the concepts are generally organized in subsumption (class/subclass) hierarchies. The semantic relationships are not explicitly specified, it takes ‘is-a’ relationship between the concepts as depicted in Figure 6.2. It shows E-Tourism ontology which consists of 21 concepts and 12 relationships. In this approach, the key concepts such as sub concepts, super concepts, equivalent concepts and disjoint concepts have been identified from the ontology by using format as specified in Figure 6.3. The questions such as define, what, compare and contrast type of questions are generated by using the concepts.

Fig 6.2. E-Tourism Ontology
In association based approaches, it identifies the object properties and related concepts. Once the association has been identified from ontology, it generates questions such as list, what, when, where, list, name and in addition, it finds relations between the two concepts.

### 6.2.3 Case Based Question Generation

Case based approaches incorporate concepts and semantic relationships with the instances. It can generate both knowledge level questions and comprehensive level questions. It generates the questions such as who, how, where, when, etc.
A small part of the sample leader’s ontology is shown in Figure 6.4. In this, the relationships like born_in, located_in and studied_in which specify how the instances are related to other instances.

Fig 6.4. Leader’s Ontology – A part

In this ontology, the following knowledge is represented as follows:

- Born in (Mahatma Gandhi, Gujarat)
- Hypernymy (Gujarat, State)
- Hypernymy (India, country)
- Born in (Gandhi, Porbandar)
- Born in (Gandhi, 1869)
- Studied (Gandhi, law)
- Studied_in (Gandhi, London)

Based on the represented knowledge and context specification, the following questions are generated.

- Where was Gandhi born?
- Who was born in Porbandar?
• In which year Gandhi was born?
• What did Gandhi study?
• Where did Gandhi study?

The following questions are also generated by inferring the semantic relationships.
• In which state Gandhi was born?
• In which country Gandhi was studied?
• In which century Gandhi was born?

In the similar way, huge numbers of questions are generated from the concepts, relationships and instances which are defined in ontology. The generated questions are verified for the correctness. The Q-Genesis system brings down the human intervention for preparing superficial questions. The detailed implementation and experiment results of a Q-Genesis have been dealt in chapter 7.

6.3 SUMMARY

This chapter discussed the necessity of semantic relationships in question generation system. The working model of question generation system based on semantic relationships (Q-Genesis) system is discussed elaborately in this chapter. The following chapter discusses the implementation of the Relation Extraction, SeRelaC, SemPER, RISer, RaSeR and Q-Genesis algorithm and it also deals with the different parameters which are used to evaluate the proposed algorithms. It also compares the proposed algorithms with existing algorithms against various parameters.