Generation of the Structure and Parameters of Training Tasks

ID Chulpan Bakieva
Minnegalieva¹
ID Anis Fuatovich Galimyanov²
ID Veronika Vladimirovna Bronskaya³

¹²Kazan Federal University

ABSTRACT

Electronic learning resource, online courses are widely applied in a learning process. As a rule, they obligatory include knowledge control block, which should contain sufficient quantity of tasks. Filling in tasks’ databases of such resources takes a considerable amount of time. Therefore it is a crucial task to explore ways of automatic generation of learning tasks. The article describes the techniques for generating tasks based on enumeration, based on information about the values of properties of objects and information about the relationship between objects, and based on information provided in the form of formulas and graphs. The properties of objects were stored in the databases. Algorithms for compiling questions to control knowledge in history, geography, and computer science have been developed. The obtained algorithms can be used for preparing the control measuring materials for such disciplines as chemistry, information theory. Value of given examples is in their possibility to replicate, extend to areas of other subjects. This allows inclusion into tasks databases of online courses sufficient number of tasks. Given algorithm can be used in composition of real time training tasks, without preservation of tasks wording in database.

Key words: training tasks generation, knowledge control, e-learning, online courses, databases.

Introduction

The creation of the necessary number of training tasks is sufficiently labour-consuming work. With the introduction of distance technologies into the educational process and the development of e-learning, the requirements for the volume of the base of control measuring materials have changed [1]. Currently, hundreds or more students can study the same e-learning course at the same time. For the objectivity of control, it is necessary that the tasks, presented to students, would not be repeated. Therefore, the actual trend is the automatic generation (that is, automatic creation) of the required number of learning tasks for electronic educational resources.
Results and Discussion

Knowledge Control Issues in Distance Learning

Use of distance learning technologies expands every year. Mass online courses take their own niche in the education system [1], [2]. Lecturers also consider possibility of using educational games, interactive courses, electronic cases [3], [4], [5]. When using all new technologies, an independent interactive and controlled intensive work of the student with educational materials becomes the main part of the educational process.

Educational materials include lectures, slides, guidelines for studying the discipline and performing control tasks, control and final tests. Successful use of distance technologies is possible when conditions for the functioning of an e-learning environment are created, which includes electronic information resources, electronic educational resources, a set of information technologies, telecommunications technologies, and appropriate technological tools [6]. Training of teachers is also important at such an organization of the educational process [7].

Electronic educational resources can be used by lecturers as an addition to traditional forms education, in mixed formats or totally distance learning formats [8], [9]. Electronic educational resources can have different structure. But block of knowledge control is obligatory in these resources. Different methods of evaluation and knowledge control can be applied [10]. In distance courses knowledge control often takes place in the form of test. Test question bank filling remains as an urgent task [11], [12]. Currently methods of using artificial intelligence, computer linguistics for questions composition and learning tasks are being studied. Modern support systems of distance learning provide lecturers with possibilities to compile tasks by some template. For instance, in Moodle system one can create calculable question, combined question and other types of tasks.

Methods of Tasks Generation

According to the method of creating a condition, training tasks that are part of distance courses can be divided into tasks with fixed conditions and tasks with conditions that include generated components. Bashmakov A. I. and Bashmakov I. A. presented the model of the problem in the form of a tuple, including the goal, initial data, restrictions, the model of the situation, results, evaluation functions of results and activities. They identified potentially generated tuple elements: $M_g = (A, C, \text{Str}(M), \text{Val}(M), \text{Int}(M), V, M_0)$. Here $A$ – purpose, $C$ – constraints that must be taken into account when performing the task, $\text{Str}(M)$ – the structure of the situation model, $\text{Val}(M)$ parameter values, $\text{Int}(M)$ – interpretation, $V$ – answers, $M_0$ – reference model of activity [13]. The authors provided a complete hierarchy of methods for generating training tasks. From this set, we will consider the following generation methods: generating tasks based on enumerations (roster method), based on information about relations between objects; based on information about the values of object properties; based on information provided in artificial languages (Fig. 1).

Figure 1. Methods for generating tasks (methods not considered in this paper are in italics)

Generating tasks based on text fragments (methods that provide building local fragments of the subject area representation)

Generation methods based on training material can be divided into two groups: first that using methods of computational linguistics and methods providing for the building local fragments of the subject area representation
The second group can include the generation of learning tasks on the basis of enumerations, on the basis of information on the relationship between objects, on the basis of information on the values of objects properties. Information about objects and their properties can be conveniently stored in structured databases.

Then it will be possible to implement the aforementioned method types of generation.

The tabular form of information presentation is the most convenient. The table can store a set of terms, historical dates, and geographic data. Let’s consider this technique using the example of creating tasks on geography.

The table “Objects” stores information about the geographic objects themselves. Geographical objects are countries, republics, cities, oceans, seas, rivers and lakes. Geographical coordinates and type of object are written besides the names in these tables. Types of objects: country, republic, city, ocean, sea, river and lake. The table “Properties” contains all basic parameters of geographical objects. These include numerical characteristics: population, population density, area, and percentage of the water surface. In the system, comprising of tasks, there can be other auxiliary tables for the realization of the connection between the data. (fig. 2)

![Diagram](image)

**Figure 2. Example of databases scheme.**

For example, we will select the necessary data about certain countries from the table with object properties. You can choose them a lot, table 1 shows part of the data. The continent can be defined by coordinates or written separately.

![Table](image)

**Table 1.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Population (mln people)</th>
<th>Population density (people per km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>South America</td>
<td>43,42</td>
<td>14,4</td>
</tr>
<tr>
<td>Mexico</td>
<td>North America</td>
<td>133,14</td>
<td>62</td>
</tr>
<tr>
<td>Venezuela</td>
<td>South America</td>
<td>30,76</td>
<td>32</td>
</tr>
<tr>
<td>Botswana</td>
<td>Africa</td>
<td>2,11</td>
<td>3,4</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Eurasia (Asia)</td>
<td>266,36</td>
<td>138,77</td>
</tr>
<tr>
<td>Slovakia</td>
<td>Eurasia (Europe)</td>
<td>5,46</td>
<td>111</td>
</tr>
<tr>
<td>Canada</td>
<td>North America</td>
<td>37,6</td>
<td>3,76</td>
</tr>
</tbody>
</table>

We will show how one can perform generation based on enumerations of this table.

In order to perform this generation a set of S classes and a set of E elements are introduced. Here are examples of classes. S1 – countries located in North America, S2 – countries located in Africa, S3 – countries with a population density of more than 100 people per square kilometre, S4 – countries with a population of 100 million people to 200 million people. In this case, the elements are countries. Then we can implement different types of tasks. Bashmakov A. I. and Bashmakov I. A. wrote about 31 types. Examples cited:

Specify elements that belong to the S1 class. (Specify countries located in North America).
Specify elements that do not belong to the S1 class. (Specify countries that are not located in North America).
Specify elements that belong to S3 and S4 classes simultaneously. (Specify countries with a population density of more than 100 people per square kilometre and a population of 100 million to 200 million people).
Specify classes among the data that this element belongs to. (On which continent is Botswana located-in Africa or in South America?).

In the resulting algorithms, we can take cities, rivers, lakes, and any other objects instead of countries.
Generation of training tasks on the basis of information about the values of object properties.

In this case, we consider E – a set of elements (objects) and P – a set of properties that characterize objects from E. We can specify the following types of training tasks:

Arrange objects in descending / ascending order for the values of this property.
Specify properties whose values are the same for objects in the set E.
Specify the object that has the lowest value of this property, and etc.

Here are the approximate templates for generation of questions on the numerical characteristics of the object:
– Wh[at] [city] has [the lowest] [area]?

Words in brackets change. City – country – republic – sea, etc. The smallest – the greatest, area – population density – the percentage of the water surface. Depending on the estimated complexity of the question, 4-6 objects are selected from the database. The program takes into account one range of options. For example, if you ask a question about the area of cities, the presence among the answers at the same time the cities of Beijing and Bruges will serve as prompt for the trainee. Therefore, objects are preliminarily sorted according to the property indicated and are divided into the ranges.
– Rank [countries] in the [descending] order of their [population density]. In this case, the names of countries (cities, lakes) are also chosen so that their population density (area, percentage of water surface) differs from each other by small values.

Training tasks based on information between objects will be generated similarly.

Examples of relationships between objects can be relationships that reflect physical connections between material objects (connected to, attached to, etc.), spatial relationships (above, next to, to the left, etc.), relationships that reflect comparison of properties of objects (larger, smaller, stronger, more efficient, etc.), relationships that reflect structure (includes, consists of, is included to, etc.).

To find out whether objects belong to each other it is possible to assign the following types of questions:
– In which [country] is [city]?

Similar to the cases discussed above, it is possible to change the country to the city, city to the lake or the sea. Objects are selected so that the answer is not obvious. For example, the question “What country is Rome in?” for most trainees will prove easy.
– Select [lakes] located in [country].

The generation of questions occurs analogously with the examples examined. In the program or when designing a database, it is envisaged to check the belonging of data to the same range. If you need to choose a lake located in Norway, the presence of “Lake Baikal” among the possible answers will serve as an unintended prompt for the student.

Using data in the tables it is possible to compose the tasks, where it is necessary to fill the gaps – to write down the missing numerical values or dates. In this case, the text from the training material is supplied at the input. The program finds the numbers, dates in the text and “cut” them. After this the dates (number) are generated and are proposed in the form of response options together with the correct answers.

Having the large database for geographical objects it is easy to form other templates and to create a fair number of questions, training tasks.

The created structures and templates can be used to fill the base of learning activities of electronic educational questions in other disciplines, such as history, chemistry, physics, and computer science. Using the methods listed above, it is possible to compose questions concerning the history: “Match the events with the dates”, “Choose the correct date of the event”, “Specify the name of the historical person you are talking about”, “Arrange historical events in a chronological sequence”, etc [14].

In chemistry it is possible to make up questions, for example, with this formulation: “Choose the three elements that are in the same period in the periodic system, and arrange these elements in order of decreasing the radius of the atom”, “Match the formula of the substance to the class/group to which the substance belongs”. When compiling test cases in chemistry, you can use the Periodic system of chemical elements of D.I. Mendeleev, presented in a format convenient for developers.

Here is an example of a computer science question: “There are requests to the search server. Arrange the numbers in ascending order of the number of pages that will find the search server for each request. For the designation of logical operation “OR” in the request is used the symbol |, while for the logical operation “AND” – &. 1) printers & scanners & sale; 2) printers & sale; 3) printers | sale; printers | scanners | sale”. To generate tasks of this type, the methods discussed above are combined with the parameter generation.

Generating tasks based on information provided in artificial languages

The training material uses various formalized representations: diagrams, program fragments, formulas, etc. That is, such representations have formal rules for their construction and processing. There are three main types of
tasks whose models are based on formalized views. The first type is the interpretation of a formalized representation. The second type is when a formalized representation is obtained from the description of a situation, object, or process. The third type is converting one formalized representation to another. If we consider the first type, interpretation is performed according to three schemes: meaningful interpretation of formalized representation components, calculation of the value, and solving typical problems based on the formalized representation.

If we take the disciplines of the physics and mathematics cycle, some topics can generate parameters and task structure using computer algebra systems, libraries of mathematical functions of programming languages [15]. In this case, the scheme for calculating the value of the formalized representation is often used. Many tasks are posted on the Wolfram Problem Generator (https://www.wolframalpha.com/). The capabilities of this resource can be used in integration with other systems [16].

Together with students we completed the generation of various tasks. Considered example.

Example 1. Analysis of computer science programs. Students must determine the outcome of a program that contains a cycle. The example is shown in the Basic and Python programming languages (Fig. 3):

![Figure 3. Example of a computer science task](image)

The algorithm of task generation:

1. Values n and s are generated from a specific set of values, and the resulting values are checked for correctness and compatibility.
2. Function is called to generate a task on the screen, that is, the task lines are compiled, including n and s.
3. The job is displayed on the screen.
4. The student’s response is red.
5. Calculates the value of the variable that is obtained during the execution of the cycle (the reference response).
6. The resulting value is compared to user response and displays the appropriate message.

Only parameters are generated in this example, by changing the function, you can create tasks to consolidate knowledge about work with operator FOR, SWITCH—CASE, and so on.

Example 2. A task in chemistry. From the chemical elements listed in the row, select three elements that belong to the same period in the Periodic table of chemical elements of D. I. Mendeleev. Arrange the selected elements in ascending order of their metallic properties (Fig. 4)

![Figure 4. Example of a chemistry task](image)

Generation algorithm:

1. A random number from 2 to 7 is generated. This number will indicate the period of the periodic table.
2. Select three random chemical elements from the period, the number of which is taken from the first step.
3. 2 random numbers are generated that point to an element in the periodic table. They must be from other periods.
4. The resulting 5 elements are combined into an array and mixed.
When analysing the student’s response, it is first checked that the selected 3 elements are in the same period. Then the order is checked. If elements are arranged in ascending order of their ordinal numbers, then their metal properties are also ordered.

Example 3. Computer science task. The numbers are given $A = A7_{16}$, $B = 251_{16}$. Find the sum of $A + B$. The answer to be given in the binary system.

Generation algorithm:
1. The quantity of terms from 2 to 4 are select randomly.
2. Randomly chosen sign of the arithmetic operation, the bases for the summands and the result of expression from an array of bases {2, 8, 10, 16}, terms range from 10 to 256.
3. Calculated the model result and checked the response of the student. Using modern programming languages, you can represent a number in a certain number system.

Example 4. A similar task in computer science. Calculate: $10101010_{2} - 250_{8} + 7_{16}$. Write the answer in decimal notation.

But in this case, the generation is designed in such a way that the algorithm can be used for different examples.

Generation algorithm:
1. Number of numbers in the specified task line is checked
2. Number of mathematical operations ($+/-$) in the specified task line is checked
3. Randomly selected characters in the expression so that one character occurs only once.
4. Numbers are randomly selected from the range (0, 256) so that the result of calculating the expression is a number greater than zero.
5. The bases of the selected numbers are randomly selected so that one base occurs only once.
6. Meta-objects in the given line are replaced in accordance with the selected numbers, their bases and signs.
7. The resulting task is displayed to the user.

Present algorithm can be used for row with arbitrary text content, for example, with a different problem statement. However, the number of meta objects in a row must remain the same. The rows with the valid location and number of meta objects are shown below:

1. “Convert into (gr:To1) the system (gr:From) number (dig:1).”
2. “Calculate the sum of the numbers $x$ and $y$ for $x = (dig:1)$, $y = (dig:2)$. Present the result in (gr:To2) notation.”
3. “Calculate the value of the expression $(dig:1) (sign:1) (dig:2)$. In answer, write the calculated value in (gr:To2) notation.”
4. “Calculate the value of the expression $(dig:1) (sign:1) (dig:2)$. Write the answer in (gr:To2) notation.”
5. “Calculate: $(dig:1) (sign:1) (dig:2) (sign:2) (dig:3)$. Write the answer in (gr:To2) notation.”
6. “Calculate: $(dig:1) (sign:1) (dig:2) (sign:2) (dig:3)$. Write the answer in the (gr:To2) numeral system. In the answer, write only the number, number system base does not need to be written.”

Example 5. Computer science. Indicate comma-separated, in ascending order all decimal numbers not exceeding 25 (number c), which entry in the numeral system with base four (number e) ends with 11 (number d).

The algorithm is actualized by generating a random basis e, generating the upper limit of the search for numbers c (a random number from the range from e to e + max, where max is a given constant, for example, equal to 50). A random number is generated from the interval from the basis to the upper limit and its representation in the decimal system is calculated. Hence, the answer will always be there. Several characters are randomly taken from the resulting number and are outputted as a d in the question.

So that to calculate the answer, the program checks the representation of all numbers from the basis to the upper limit in decimal form. As per condition, if the number ends with a certain number of given digits, the number is written as the answer (Fig.5).
Conclusion

Thus, having information about objects and their properties, you can generate training tasks for different disciplines. Formulas, diagrams, and graphs also provide many opportunities for automatic task creation. Some tasks can be generated using different algorithms, for example, based on information about relationships between objects and using formalized representations. An example would be a chemistry task where you need to select elements from a single period. In this case, you need to choose which method is the most optimal for implementation. If you already have a ready-made database, it is preferable to use well-developed algorithms based on text fragments. The created tasks can be used when filling in the database of educational tasks of an electronic educational resource. When using distance learning technologies, tasks can be created in real time.

Acknowledgements

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References