



<http://dx.doi.org/10.35596/1729-7648-2020-18-5-44-52>

Оригинальная статья
Original paper

УДК 004.89

ОНТОЛОГИЧЕСКИЙ ПОДХОД К АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ВОПРОСОВ В ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМАХ

ЛИ ВЭНЬЦЗУ

*Белорусский государственный университет информатики и радиоэлектроники
(г. Минск, Республика Беларусь)*

Поступила в редакцию 22 мая 2020

© Белорусский государственный университет информатики и радиоэлектроники, 2020

Аннотация. В работе предложен подход к проектированию общей подсистемы автоматической генерации вопросов в интеллектуальных обучающих системах. Проектируемая подсистема позволяет автоматически генерировать различные типы вопросов на основе информации из базы знаний и сохранять генерируемые вопросы в базе знаний подсистемы для последующего использования. Основной частью подсистемы является модуль автоматической генерации вопросов, позволяющий генерировать вопросы разных типов на основе существующих стратегий генерации вопросов в сочетании со структурными характеристиками баз знаний, построенных по технологии OSTIS. Предложен ряд стратегий автоматической генерации вопросов, использование которых позволяет автоматически генерировать вопросы различных видов, такие как вопросы на выбор, вопросы на заполнение пробелов, вопросы на толкование определений и т. д. Важнейшей частью подсистемы является база знаний, в рамках которой хранится онтология вопросов, включая сами экземпляры вопросов. В данной работе база знаний построена на основе стандартов технологии OSTIS. Разработана классификация типов автоматически генерируемых вопросов, а также предметная область для хранения генерируемых вопросов и соответствующая им онтология, описанная в базе знаний подсистемы. Генерируемые вопросы хранятся в базе знаний подсистемы в виде SC-кода, который является стандартом технологии OSTIS. При тестировании пользователей данные генерируемые вопросы преобразуются в соответствующие вопросы на естественном языке с помощью естественно-языкового интерфейса. В сравнении с существующими подходами, предложенный в данной работе подход имеет определенные преимущества, а подсистема, разработанная с использованием предложенного подхода, может быть применена в различных обучающих системах, построенных по технологии OSTIS.

Ключевые слова: генерация вопросов, базы знаний, обучающие системы, технология OSTIS.

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Для цитирования. Ли Вэньцзу. Онтологический подход к автоматической генерации вопросов в интеллектуальных обучающих системах. Доклады БГУИР. 2020; 18(5): 44-52.

ONTOLOGICAL APPROACH TO AUTOMATIC GENERATION OF QUESTIONS IN INTELLIGENT LEARNING SYSTEMS

LI WENZU

Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)

Submitted 22 May 2020

© Belarusian State University of Informatics and Radioelectronics, 2020

Abstract. This article proposes an approach for designing a general subsystem of automatic generation of questions in intelligent learning systems. The designed subsystem allows various types of questions to be automatically generated based on information from the knowledge bases and save the generated questions in the subsystem knowledge base for future use. The main part of the subsystem is the automatic generation module of questions, which allows one to generate questions of various types based on existing question generation strategies in combination with the structural characteristics of knowledge bases built using OSTIS technology. In this article, a variety of strategies for automatically generated questions are proposed, the use of which allows various types of questions to be automatically generated, such as multiple-choice questions, fill-in-the-blank questions, questions of definition interpretation and etc. The most important part of the subsystem is the knowledge base, which stores the ontology of questions, including the question instances themselves. In this article, the knowledge base is constructed based on OSTIS technical standards. The type classification of automatically generated questions was developed, as well as the subject area for storing generated questions and the corresponding ontology described in the knowledge base of the subsystem. The generated questions are stored in the subsystem knowledge base in the form of SC-code, which is the OSTIS technology standard. When testing users, these automatically generated questions are converted to the corresponding natural language form through the natural language interface. Compared with the existing approaches, the approach proposed in this article has certain advantages, and the subsystem designed using this approach can be used in various OSTIS-based systems driven by OSTIS technology.

Keywords: question generation, knowledge base, learning systems, OSTIS technology.

Conflict of interests. The author declares no conflict of interests.

For citation. Li Wenzu. Ontological approach to automatic generation of questions in intelligent learning systems. Doklady BGUIR. 2020; 18(5): 44-52.

Введение

В последние годы с развитием технологий искусственного интеллекта образовательные средства перешли от традиционных мультимедийных образовательных технологий к эпохе интеллектуального образования [1]. Использование в процессе обучения передовых способов, предоставляемых технологиями искусственного интеллекта, может стимулировать интерес пользователей к обучению и повысить его эффективность.

По сравнению с традиционными мультимедийными обучающими системами (ТМОС), наиболее важной особенностью интеллектуальных обучающих систем (ИОС) является то, что ИОС может использовать базу знаний для автоматического генерирования различных типов вопросов и автоматической проверки ответов пользователей. Но подходы к генерации вопросов, предложенные большинством исследователей, позволяют генерировать только самые простые вопросы (вопросы на выбор, вопросы на заполнение пробелов и т. д.), и корреляция между правильными и неправильными (мешающими) вариантами генерируемых вопросов на выбор невысока. Поэтому большинство подходов к автоматической генерации вопросов не отвечают требованиям практического применения, а предложенные подходы реализуются на конкретных системах, так что они не являются универсальными.

В данной работе предложен подход к автоматической генерации вопросов для решения вышеуказанных проблем. Предлагаемый подход направлен на разработку общей подсистемы автоматической генерации вопросов для интеллектуальных обучающих систем, построенных по технологии OSTIS [2]. Следует отметить, что в качестве демонстрационной системы выбрана система по дискретной математике, разрабатываемая при участии студентов и преподавателей кафедры интеллектуальных информационных технологий Белорусского государственного университета информатики и радиоэлектроники.

Существующие подходы и проблемы

Методы автоматической генерации вопросов (МАГВ) позволяют автоматически генерировать тестовые вопросы из электронных документов, корпусов текстов и баз знаний с помощью компьютерных технологий. По сравнению с традиционным подходом, предполагающим использование базы данных для извлечения вопросов, МАГВ основаны на знаниях, поэтому генерируемые вопросы являются более гибкими [1, 3, 4].

МАГВ можно разделить на следующие категории по источнику знаний:

- на основе электронных документов;
- на основе корпуса текстов;
- на основе базы знаний.

С развитием технологий семантических сетей использование баз знаний для автоматической генерации вопросов стало основным направлением исследований. База знаний создается после того, как специалисты предметной области проанализируют полученные знания. Данные знания будут профильтрованы и организованы в определенную структуру, поэтому вопросы, автоматически генерируемые с помощью базы знаний, являются более гибкими и разнообразными [1, 4]. Ниже будут представлены некоторые результаты исследований по автоматической генерации вопросов:

– система автоматической генерации вопросов на выбор, разработанная Andreas Papasalouros, в основном использует отношения между родительскими классами и подклассами, классами и элементами, элементами и атрибутами в онтологии OWL для автоматической генерации вопросов на выбор [5];

– на основе вышеупомянутого подхода Li Hui была разработана онтология в сфере компьютерной теории с помощью Protégé и использованы различные онтологические отношения, чтобы автоматически генерировать вопросы на выбор, вопросы на заполнение пробелов и вопросы суждения [1, 4, 5].

Хотя подходы, рассмотренные выше, имеют много преимуществ, существует также много проблем:

– подход к автоматической генерации вопросов с использованием электронных документов и шаблонов предложений требует большого количества шаблонов предложений, а качество генерируемых вопросов невысоко;

– объем и качество автоматически генерируемых вопросов с использованием корпуса текстов зависят от размера и качества корпуса текстов, и корреляция между неправильными вариантами и правильными вариантами генерируемых вопросов на выбор невысока;

– в настоящее время не существует единого стандарта разработки базы знаний, поэтому различные базы знаний имеют различную структуру и несовместимы друг с другом. Поскольку базы знаний несовместимы друг с другом, подход, использующий базу знаний для автоматической генерации вопросов, может быть использован только в соответствующей конкретной базе знаний, а для баз знаний, разработанных в рамках других подходов, должны разрабатываться новые соответствующие подходы к генерации вопросов;

– существующие подходы к автоматической генерации вопросов позволяют генерировать только простые объективные вопросы (те, на которые существует только один стандартный правильный ответ).

На основе существующих результатов исследований и в соответствии со стандартами технологии OSTIS [2, 6] в данной работе предложен подход к автоматической генерации субъективных и объективных вопросов с использованием базы знаний для ИОС. Объективные

вопросы относятся к типу вопросов с уникальным стандартным ответом и включают в себя вопросы на выбор, вопросы суждения и т. д. Субъективные вопросы не имеют уникальных ответов и включают вопросы на доказательство, вопросы на толкование определений и т. д. Предлагаемый подход будет реализован в рамках интеллектуальной обучающей системы по дискретной математике.

Следует подчеркнуть, что при тестировании пользователей для субъективных типов вопросов близость между ними может быть определена путем вычисления подобия между стандартными ответами и ответами пользователей. Стандартные ответы хранятся в базе знаний подсистемы автоматической генерации вопросов в виде SC-кода (в качестве основы представления знаний в рамках технологии OSTIS используется унифицированный вариант кодирования информации любого рода на основе семантических сетей, названный SC-кодом), а ответы пользователей на естественных языках (английском, китайском, русском и др.) преобразуются в SC-код с помощью естественно-языкового интерфейса [2, 6]. Основная идея вычисления подобия состоит в том, чтобы сначала разложить семантический граф стандартного ответа и семантический граф ответа пользователя (семантический граф – это неориентированный граф или направленный граф, который визуальным образом представляет семантические отношения между понятиями) на подструктуры в соответствии со структурой представления знаний, а затем вычислить подобие между стандартным ответом и ответом пользователя в соответствии с тем же числом подструктур. Поскольку проверка ответов в рамках данной статьи не является основным направлением, конкретный процесс проверки ответов не будет подробно описан.

Предлагаемый подход

В данной работе предложен подход к автоматической генерации различных типов вопросов для ИОС. Используя предлагаемый подход и технологию OSTIS [2, 6], субъективные и объективные вопросы могут быть автоматически сформированы из базы знаний. Сгенерированные вопросы хранятся в базе знаний подсистемы автоматической генерации вопросов в виде представления SC-кода [6]. Данные сгенерированные вопросы преобразованы в соответствующие вопросы на естественном языке с помощью естественно-языкового интерфейса, что не является задачей данной работы.

Рассмотрим более подробно стратегии генерации вопросов.

1. Стратегия генерации на основе элементов.

На основе ролевого отношения.

На основе бинарного отношения.

Ролевое отношение (то, которое задает роль элементов в рамках некоторого множества) и бинарное отношение (это множество отношений на множестве M , являющихся подмножеством декартова произведения множества M самого на себя) – это типы отношений, часто используемых при построении базы знаний по технологии OSTIS [6]. Ниже приведен фрагмент базы знаний в SCn-коде (одном из языков внешнего отображения SC-кода), использующий бинарное отношение “автор*”:

Понятие условной сходимости

⇒ автор*:

Иоганн Петер Густав Лежён Дирихле

Взяв в качестве примера генерируемый вопрос на заполнение пробелов, можно записать его следующим образом: «Автором понятия условной сходимости является ____». Поэтому данные отношения между элементами могут использоваться для автоматической генерации объективных вопросов.

2. Стратегия генерации на основе классов.

Данная стратегия использует различные отношения между классами, чтобы автоматически генерировать объективные вопросы.

На основе отношения “включение*”.

В базе знаний интеллектуальной системы, построенной по технологии OSTIS, многие классы содержат несколько подклассов, поэтому отношение включения между классами и подклассами используется для автоматической генерации объективных вопросов. На языке теории множеств отношение включения между классами выражается следующим образом: $S_i \subseteq C$ ($i \geq 1$), (S – подкласс, i – номер подкласса, C – родительский класс). Взяв в качестве примера генерируемый вопрос суждения, можно записать его следующим образом: $S_i \subseteq C$ является “истинным” или $S_i \subseteq C$ является “ложным”. На рис. 1 приведен пример семантической структуры вопроса на выбор, автоматически генерируемого с помощью данной стратегии на SCg-коде (SCg-код – графический вариант внешнего визуального представления SC-кода) [2].

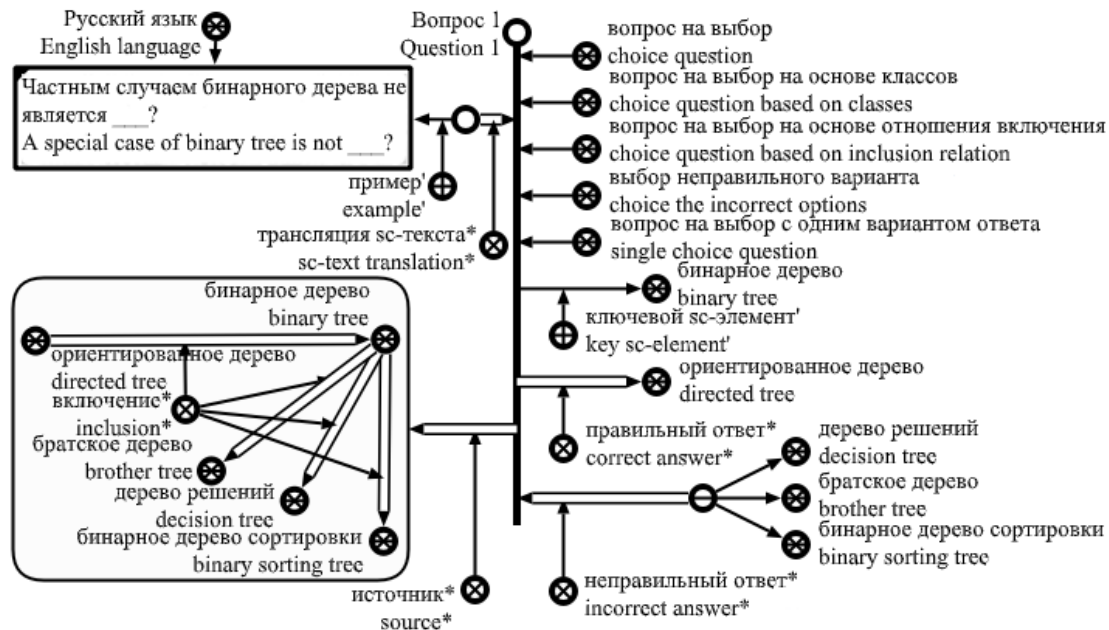


Рис. 1. Пример вопроса на выбор
Fig. 1. Example of a multiple-choice question

Пользователю предлагается выбрать один из четырех вариантов, связанных с бинарным деревом, который не является частным случаем бинарного дерева. Описание данного примера на естественном языке выглядит следующим образом:

«Частным случаем бинарного дерева не является ()?»

- A. дерево решений C. ориентированное дерево
B. братское дерево D. бинарное дерево сортировки

На основе отношения “разбиение*”.

Областью определения отношения разбиения является семейство всевозможных множеств. В результате разбиения множества получается множество попарно непересекающихся множеств, объединение которых есть исходное множество [6]. В базе знаний, построенной по технологии OSTIS, как правило, существует много классов, удовлетворяющих отношению разбиения, поэтому данное отношение может быть использовано для автоматической генерации различных типов объективных вопросов по аналогии с отношением “включение*”.

На основе отношения “строгое включение*”.

Строгое включение является частным случаем отношения включения. Подход к автоматической генерации объективных вопросов с использованием отношения строгого включения аналогичен подходу к генерации объективных вопросов с использованием отношения включения. На языке теории множеств отношение строгого включения между классами можно выразить следующим образом: $S_i \subset C$ ($i \geq 1$), (S – подкласс, i – номер подкласса, C – родительский класс).

3. Стратегия генерации на основе идентификаторов.

Обычно некоторые множества и отношения в базе знаний имеют несколько идентификаторов (под идентификатором в технологии OSTIS понимается файл, который может быть использован для обозначения (именования) той или иной сущности в рамках какого-либо внешнего языка) [6]. Например, «в дискретной математике ориентированное множество также называется кортежем; вектором и упорядоченной последовательностью;» Взяв в качестве примера вопрос на выбор, генерируемый с использованием идентификаторов ориентированного множества, можно записать его следующим образом: «Ориентированное множество также называется ()?» Правильными вариантами этого вопроса на выбор являются идентификаторы ориентированного множества, а неправильными (мешающими) вариантами могут быть идентификаторы множеств, связанных с ориентированным множеством. Таким образом, понятия, имеющие несколько идентификаторов, могут быть использованы для автоматической генерации объективных вопросов.

4. Стратегия генерации на основе аксиом.

Многие аксиомы и их математические выражения хранятся в базе знаний по дискретной математике. Например, в базе знаний формально представлена аксиома объединения и соответствующее ей математическое выражение: $\forall a \exists d \forall c (c \in d \sim \exists b (b \in a \wedge c \in b))$. Таким образом, эта аксиома и ее математическое выражение могут быть использованы для генерации аналогичных типов вопросов, описанных выше.

5. Стратегия генерации на основе свойств отношений.

Многие отношения в базе знаний удовлетворяют свойствам рефлексивности, симметричности и транзитивности. Ниже приведен фрагмент базы знаний, описывающий свойства отношения на SCn-коде:

строгое включение*

∈ антирефлексивное отношение
∈ транзитивное отношение

Взяв в качестве примера вопрос суждения, генерируемый с помощью данной стратегии, можно записать его следующим образом: «Верно ли, что если множество А строго включается во множество В, а множество В строго включается во множество С, то множество А строго включается во множество С?». Поэтому данные отношения и их свойства используются для генерации объективных вопросов. Важно отметить, что как показано в данном примере, благодаря базе знаний система способна генерировать вопросы, проверяющие не только знания учащегося, но степень его понимания и способность применять эти знания при решении задач.

6. Стратегия генерации на основе примеров изображений.

В базе знаний хранятся некоторые понятия, отношения, теоремы и поясняющие их изображения, что позволяет автоматически генерировать некоторые объективные вопросы.

7. Стратегия генерации субъективных вопросов.

Определения понятий и доказательства теорем хранятся в базе знаний в соответствии с правилами представления знаний, поэтому процесс генерации субъективных вопросов можно разделить на следующие этапы:

1. Использование шаблонов логических формул для поиска определений некоторых понятий и доказательств некоторых теорем в базе знаний.

2. Найденные фрагменты баз знаний хранятся в соответствующей предметной области базы знаний подсистемы в соответствии с типом фрагмента (определение или доказательство).

3. Наконец, с использованием ручных или автоматических подходов для описания определения каждого конкретного понятия и процесса доказательства каждой конкретной теоремы на SCg-коде или SCL-коде (специальный подязык SC-кода, предназначенный для формализации логических формул) в соответствии с правилами представления знаний и с использованием отношения “стандартный ответ*” они связываются с соответствующими семантическими фрагментами [2, 6].

Вопросы, автоматически генерируемые с использованием перечисленных выше стратегий, могут содержать повторные и неправильные вопросы, поэтому для обеспечения качества генерируемых вопросов необходимо сначала сохранить эти автоматически генерируемые вопросы в базе знаний подсистемы автоматической генерации вопросов, а затем

использовать ручные или автоматические подходы (сравнение подобия между вопросами) для фильтрации повторных и неправильных вопросов. Основой базы знаний любой системы OSTIS является иерархическая система предметных областей и соответствующих им онтологий (онтология – это вид знаний, каждое из которых является спецификацией соответствующей предметной области, ориентированной на описание свойств и взаимосвязей понятий, входящих в состав указанной предметной области) [6]. Рассмотрим иерархию предметной области, используемой для хранения генерируемых вопросов в базе знаний подсистемы на SCn-коде.

Раздел. Предметная область вопросов

← декомпозиция раздела*:

- ```

{
 • Раздел. Предметная область субъективных вопросов
 ← декомпозиция раздела*:
 {
 • Раздел. Предметная область вопросов на доказательство
 • Раздел. Предметная область вопросов на толкование теорем
 • Раздел. Предметная область вопросов на толкование определений
 }
 • Раздел. Предметная область объективных вопросов
 ← декомпозиция раздела*:
 {
 • Раздел. Предметная область вопросов на заполнение пробелов
 • Раздел. Предметная область вопросов суждения
 • Раздел. Предметная область вопросов на выбор
 }
}

```

Среди них объективные типы вопросов могут быть разложены на более конкретные типы в соответствии со стратегиями генерации вопросов и характеристиками каждого типа вопроса (например, вопросы на выбор включают в себя: вопросы на выбор с одним вариантом ответа и вопросы на выбор с несколькими вариантами ответа). На рис. 2 приведен семантический фрагмент классификации вопросов на выбор в предметной области вопроса на выбор на SCg-коде [2].

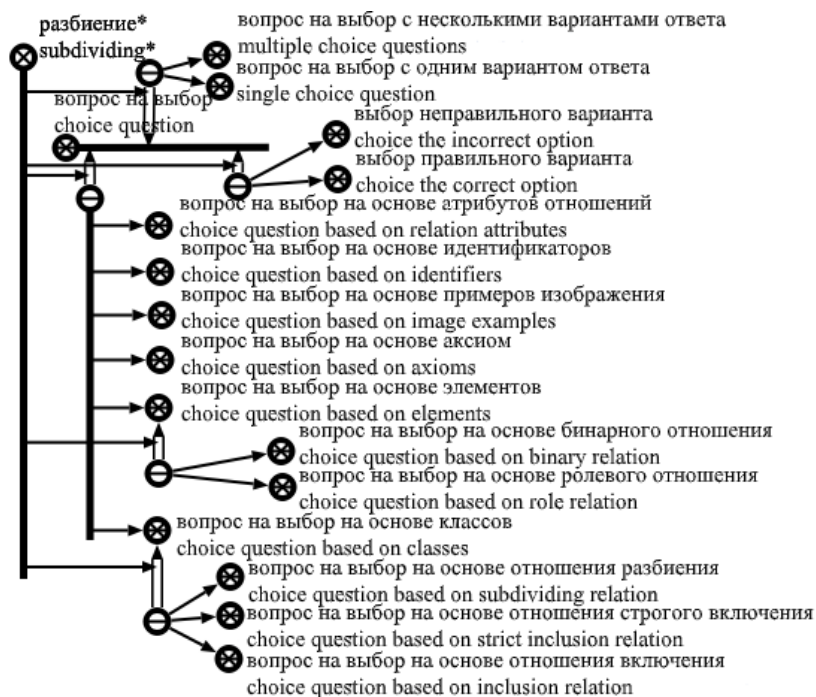


Рис. 2. Семантический фрагмент классификации вопросов на выбор  
Fig. 2. Semantic fragment of multiple-choice question classification

Вопросы на выбор, автоматически генерируемые с помощью базы знаний, будут храниться в соответствующих разделах базы знаний подсистемы в соответствии с классификацией, приведенной на рисунке. Поскольку генерируемые вопросы хранятся в базе знаний подсистемы автоматической генерации вопросов в соответствии с типом вопросов, при тестировании пользователей системой из базы знаний подсистемы извлекаются вопросы определенных типов в соответствии с требованиями пользователей, что значительно повышает эффективность извлечения вопросов. В текущей версии методы поддержки извлечения вопросов включают в себя: извлечение требуемых типов вопросов в соответствии с количеством, извлечение требуемых типов вопросов в соответствии со стратегией генерации вопросов и случайное извлечение вопросов в соответствии с количеством.

Предложенный в данной работе подход к автоматической генерации вопросов и подход использования онтологии для хранения генерируемых вопросов имеют следующие преимущества:

– поскольку базы знаний, разработанные с использованием технологии OSTIS, имеют одинаковую структуру хранения знаний, предлагаемый подход для автоматической генерации вопросов может использоваться в других интеллектуальных системах, построенных по технологии OSTIS;

– сгенерированные вопросы выражаются с помощью унифицированного языка кодирования знаний SC-кода, поэтому они не зависят от естественного языка;

– используя подход, предложенный в данной работе, можно не только генерировать субъективные и объективные вопросы, но и повысить качество генерируемых вопросов.

### Заключение

В данной работе анализируется состояние развития и преимущества ИОС, обсуждаются проблемы существующих подходов автоматической генерации вопросов. На основе объединения существующих подходов и технологии OSTIS предложен подход к автоматической генерации вопросов, который решает некоторые актуальные проблемы. Подсистема, разработанная с использованием предложенного подхода, обладает независимостью от естественного языка, и может широко использоваться в обучающих системах, построенных по технологии OSTIS.

Предложенные подходы к автоматической генерации вопросов и использованию онтологий для хранения сгенерированных вопросов позволяют устранить недостатки автоматической генерации вопросов, перечисленные в начале данной работы. Несмотря на то, что предложенный подход имеет ряд преимуществ, существует также ряд проблем, которые необходимо решить:

– контроль качества и частоты повторения автоматически генерируемых вопросов в онтологии вопросов;

– несоответствие набора терминов (имен понятий) при одинаковой системе понятий в разных источниках информации.

Следующим шагом является внедрение данной подсистемы в обучающую систему по дискретной математике.

### Список литературы

1. Li H. Research on item automatic generation based on DL and domain ontology. *Journal of Changchun University of Technology (Natural Science Edition)*. 2012;33(04):461-464. DOI: 1674-1374(2012)04-0460-05.
2. Голенков В.В., Гулякина Н.А. Проект открытой семантической технологии компонентного проектирования интеллектуальных систем. *Онтология проектирования*. 2014;1:42-64.
3. Li W., Grakova N.V., Qian L. Ontological approach to automating the processes of question generation and knowledge control in intelligent learning systems. *Open Semantic Technologies for Intelligent Systems (OSTIS-2020)*. 2020;4:215-224.
4. Ruslan M., Le N.H., Nikiforos K. A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*. 2006;12(02):177-194. DOI: 10.1017/S1351324906004177.



5. Andreas P., Konstantinos K., Konstantinos K. Automatic generation of multiple-choice questions from domain ontologies. *IADIS International Conference e-Learning*. 2008;1:427-434.
6. Golenkov V.V., Ivashenko V.P., Taberko V.V., Kroshchanka I.I. Integration of artificial neural networks and knowledge bases. *Ontology of designing*. 2018;29:366-386. DOI: 10.18287/2223-9537-2018-8-3-366-386.

### References

1. Li H. Research on item automatic generation based on DL and domain ontology. *Journal of Changchun University of Technology (Natural Science Edition)*. 2012;33(04):461-464. DOI: 1674-1374(2012)04-0460-05.
2. Golenkov V.V., Guljakina N.A. [Project of open semantic technology for component design of intelligent systems]. *Ontologija projektovanija = Ontology of designing*. 2014;1:42-64. (In Russ.)
3. Li W., Grakova N.V., Qian L. Ontological approach to automating the processes of question generation and knowledge control in intelligent learning systems. *Open Semantic Technologies for Intelligent Systems (OSTIS-2020)*. 2020;4:215-224.
4. Ruslan M., Le N.H., Nikiforos K. A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*. 2006;12(02):177-194. DOI: 10.1017/S1351324906004177.
5. Andreas P., Konstantinos K., Konstantinos K. Automatic generation of multiple-choice questions from domain ontologies. *IADIS International Conference e-Learning*. 2008;1:427-434.
6. Golenkov V.V., Ivashenko V.P., Taberko V.V., Kroshchanka I.I. Integration of artificial neural networks and knowledge bases. *Ontology of designing*. 2018;29:366-386. DOI: 10.18287/2223-9537-2018-8-3-366-386.

### Сведения об авторе

Ли Вэньцзу, аспирант кафедры интеллектуальных информационных технологий Белорусского государственного университета информатики и радиоэлектроники.

### Information about the author

Li Wenzu, PG Student of the Department of Intelligent Information Technologies of Belarusian State University of Informatics and Radioelectronics.

### Адрес для корреспонденции

220013, Республика Беларусь,  
г. Минск, ул. П. Бровки, 6,  
Белорусский государственный университет  
информатики и радиоэлектроники  
тел. +375-29-851-60-84;  
e-mail: lwzzggml@gmail.com  
Ли Вэньцзу

### Address for correspondence

220013, Republic of Belarus,  
Minsk, P. Brovka str., 6,  
Belarusian State University  
of Informatics and Radioelectronics  
tel. +375-29-851-60-84;  
e-mail: lwzzggml@gmail.com  
Li Wenzu