

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ,
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.056

DOI: 10.17212/2782-2230-2022-1-85-98

**ИССЛЕДОВАНИЕ ПРИМЕНЕНИЯ ГРАФОВ ЗНАНИЙ
ДЛЯ ГЕНЕРАЦИИ ВОПРОСОВ АУДИТА
ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ***

А.В. ЛЕМЯКИН¹, А.Б. АРХИПОВА²

¹ 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, лаборант кафедры защиты информации. E-mail: a.v.lemiyakin@gmail.com

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доцент кафедры защиты информации. E-mail: arhipova@corp.nstu.ru

В статье представлены подходы к исследованию вопросов применения графов знаний для генерации вопросов аудита информационной безопасности. Показана значимость человеческого фактора в аудите информационной безопасности как элемента социально значимой деятельности. В качестве решения формирования объективной базы исследования предложен метод автоматической генерации вопросов на основе графов знаний, способных выявлять набор картежей, представляющих собой субъект-объектные отношения. Проанализированы существующие модели и набор динамически формируемых метрик. В частности, субъективный метод пятибалльной шкалы Макото Нагао и машинный метод автоматической системы оценки на основе метода *N*-грамм. Проанализированы достоинства и недостатки методов исследования, представлена постановка задачи на дальнейшее исследование.

Ключевые слова: граф знаний, повышение осведомленности, политики безопасности, автоматическая генерация вопросов, семантический анализ, контроль знаний, тесты

ВВЕДЕНИЕ

Самый опасный нарушитель в любой компании – это внутренний нарушитель. Так, например, по статистике компании Infowatch, доля утечек по вине сотрудников за 2020 год составила 72 % [1]. Для решения этой проблемы призван целый ряд мер. Так, в статье [2] приводится пять основных областей: повышение культуры информационной безопасности (ИБ) организации

* Статья получена 10 февраля 2022 г.

и обучение персонала, ИТ-компетенция сотрудников, внедрение и соблюдение сотрудниками политик безопасности организации, управление рисками, менеджмент ИБ. Самыми эффективными были признаны повышение культуры ИБ организации, обучение персонала и менеджмент ИБ. Обучение персонала также подразумевает тестирование. Это две взаимосвязанные вещи. Однако обычный контроль осведомленности о политиках безопасности имеет значительный недостаток – ручной способ создания вопросов.

Решением данной проблемы могла бы стать автоматическая генерация вопросов на основе графов знаний, способных выявить из анализируемого текста так называемые триплеты – картежи, состоящие из трех элементов: объекта, субъекта и предиката (отношение между объектом и субъектом). При этом в идеале необходимо соблюсти универсальность генератора для любых политик безопасности, чтобы любая организация могла без лишних усилий автоматически сгенерировать тесты на контроль знаний.

1. АВТОМАТИЧЕСКАЯ ГЕНЕРАЦИЯ ВОПРОСОВ

Генерация вопросов является задачей синтеза текста наряду с задачей анализа текста. Анализ может быть представлен следующими этапами [3].

1. Считывание текстовой информации из входного файла.
2. Выделение отдельных членов предложения. Реализуется любым из множества доступных способов (например, выделять слова от пробела до пробела и добавлять в массив).
3. Определение принадлежности части речи для каждого члена предложения (например, с помощью библиотеки на языке C# – DeepMorphy).
4. Подготовка семантических моделей, представленных в виде графов, где узлами являются слова, а ребрами – их отношения.

Синтез текста (вопросов) может быть представлен комбинацией двух способов из нижеприведенных классификаций (по методам генерации и по структуре исходных данных).

1.1. КЛАССИФИКАЦИЯ СПОСОБОВ ГЕНЕРАЦИИ ВОПРОСОВ

По методам генерации вопросов [4].

1. Логические методы (LOGIC) – методы, основанные на представлении функций и моделей в виде заранее заданных формул, на основании которых задача генерации вопросов сводится к применению этих формул для преобразования исходных данных.

2. Методы на основе машинного обучения (ML) – методы, способные обучиться по прецедентам, т. е. сделать общие выводы на основе частных наблюдений.

3. Гибридные методы (HM) – методы, сочетающие в себе как логические методы, так и методы на основе машинного обучения.

По структуре исходных данных [4].

1. Использующие текстовые данные (TEXT) – методы, которые используют входные данные в виде человекочитаемого текста.

2. Использующие графовые данные (GRAPH) – методы, входные данные которых представляют собой знания, использующие графовое описание. Примером является формат представления знаний RDF (Resource Description Framework), в котором знания представлены в виде триплетов (субъект, предикат, объект), описывающих направленную связь субъекта с объектом.

3. Использующие гибридные данные (HD) – методы, для которых входные данные подаются в гибридном представлении описанных выше структур исходных данных. Примером такого представления может быть человек, читающий текст, дополненный онтологией в формате RDF, которая описывает предметную область текста.

В работе [5] для интеллектуальной обучающейся системы (ИОС) используется технология OSTIS (Open Semantic Technology for Intelligent Systems) и автоматическая генерация субъективных и объективных вопросов.

Генерация вопросов происходит логическими методами, которые реализуются аппаратом дискретной математики. Относительного этого подхода можно выделить следующие виды стратегии генерации вопросов [5].

1. На основе элементов:

- на основе ролевого отношения – то, которое задает роль элементов в рамках некоторого множества;

- на основе бинарного отношения – множество отношений на множестве M , являющихся подмножеством декартова произведения множества M на самого себя.

2. На основе различных отношений между классами:

- отношения «включение», когда некоторые классы в базе знаний содержат несколько подклассов. Пример: «Частным случаем бинарного дерева является: вариант 1, вариант 2, и т. д.»;

- отношения «разбиение», когда в результате разбиения множества получается множество попарно непересекающихся множеств, объединение которых есть исходное множество;

- отношения «строгое включение» – частный случай отношения включения.

3. На основе идентификаторов, когда некоторые сущности в базе знаний могут иметь несколько идентификаторов (синонимов). Пример: «Ориентированное множество также называется ___?».

4. На основе аксиом.

5. На основе свойств отношений – использование свойств рефлексивности, симметричности и транзитивности, которым могут соответствовать некоторые отношения в базе знаний. Пример: «Если $A = B$, и при этом $B = C$, то равны ли A и C ?».

6. На основе примеров изображений, когда в базе знаний хранятся некоторые понятия, отношения, теоремы и поясняющие их изображения.

7. Генерация субъективных вопросов – вопросы на определение понятий и доказательства теорем.

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 [5].

1.2. ОЦЕНКА КАЧЕСТВА СГЕНЕРИРОВАННЫХ ВОПРОСОВ

Существует два принципиальных способа оценки [6]: субъективный и машинный.

Первый может быть реализован с помощью пятибалльной шкалы Макото Нагао, в которой количество баллов означает следующее:

1 балл – смысл предложения понятен, и не возникает никаких вопросов; грамматика, словоупотребление и стиль соответствуют общей структуре текста и не требуют постредактирования;

2 балла – смысл предложения понятен, но возникают большие проблемы с грамматикой, словоупотреблением и стилем;

3 балла – общий смысл предложения понятен, но смысл некоторых его частей вызывает сомнение из-за неправильного грамматического строя;

4 балла – присутствуют ошибки словоупотребления и стилистики, требуется обращение к оригиналу;

5 баллов – в предложении имеется большое количество грамматических, словоупотребительных и стилистических ошибок, смысл предложения с трудом можно понять после внимательного изучения.

Второй способ может быть реализован с помощью автоматической системы оценки на основе метода N -грамм (вероятности появления цепочки букв N -го порядка (N -грамм) в анализируемых текстах). Он представлен таким метриками [7]: BLEU, NIST, METEOR.

BLUE (Bilingual Evaluation Understudy) – основная идея: «Чем ближе машинный текст к профессиональному человеческому тексту, тем лучше». Происходит вычисление точности выявленных n -грамм, так как, например, при переводе текста с одного языка на другой машинным способом оригинал часто оказывается более емким.

NIST (Национальный институт стандартов и технологий США) – в отличие от метода BLEU, помимо точности вычисляется еще и то, насколько информативным является конкретный N -грамм. Так, например, предлог будет иметь значительно меньший вес, чем последовательность из прилагательного и существительного.

METEOR (Metric for Evaluation of Translation with Explicit Ordering) использует функции сопоставления синонимов вместе с точным соответствием слов.

1.3. ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ ГЕНЕРАЦИИ ВОПРОСОВ

В работе [8] приводится самостоятельно разработанное решение на основе уже существующего приложения linkeddata trivia. Результаты приведены в сравнительной табл. 1.

Таблица 1

Table 1

Сравнение решений
Comparison of solutions

	linkeddata trivia	Самостоятельная разработка
Способ построения вопросов	По случайному субъекту	По поиску субъекта; случайному субъекту, определенному по классу; случайному субъекту, определенному по тематике вопросов
Время на поиск субъектов	Около 23 с	< 1 с
Время на построение вопросов	–	Около 5 с

Окончание табл. 1

End of Tab. 1

	linkeddata trivia	Самостоятельная разработка
Полный перевод	Нет	Неполный (возможна ситуация, когда литерал на русском отсутствует в базе данных, и тогда он заменяется на английский, поэтому некоторые предикаты и субъекты могут быть на английском)
Примеры результатов	Что является частью военного конфликта операции «Радуга»?	Кто, или что, или какой автор Мона Лиза?
Особенности	–	Поиск реализован за счет Google Custom Search. Есть возможность поиска по географической карте
Источник исходных данных	Семантическая сеть DBPedia	

В статье [9] приводится модель Graph2Seq, которая улучшает современный показатель BLEU-4 с 11,57 до 29,40 и с 25,99 до 59,59 в тестах WQ (WebQuestions) и PQ (PathQuestions) соответственно.

1.4. ПРОБЛЕМЫ ГЕНЕРАЦИИ ВОПРОСОВ

Самые распространенные проблемы в области автоматической генерации вопросов [3].

1. Сложность моделирования семантики. Наличие в текстах синонимов, омонимов, неточностей, неопределенностей высказывания – всё это вызывает определенные проблемы при анализе текста.

2. Использование контекста. Зачастую человек воспринимает информацию лишь благодаря тому, что понимает контекст, в ходе которого эта информация была получена. Программа же контекста не понимает.

3. Трудности формирования сложных вопросов. Большинство решений позволяют генерировать только самые простые вопросы (вопросы на выбор, вопросы на заполнение пробелов и т. д.) [5].

Проблему контекста можно решить разными способами. Так, в статье [10] используется генерация последовательности вопросов на основе заранее известной последовательности ответов.

Для генерации вопросов выполняется взаимодействие с двумя графами:

1) об информации из отрывка текста; используется для лучшего захвата контекстных зависимостей;

2) об информации из ответов; используется, чтобы сделать сгенерированные вопросы более релевантными заданным ответам.

Пример исходного отрывка текста с отметками ответов:

(1) A small boy named [John]1 was at the park one day.

(2) He was [swinging]2 [on the swings]3 and [his friend]4 named [Tim]5 [played on the slide]6.

(3) John wanted to play on the slide now.

(4) He asked Tim [if he could play on the slide]7.

(5) Tim said [no]8, and he cried.

Т а б л и ц а 2

T a b l e 2

Последовательность вопросов на основе последовательности ответов

Sequence of questions based on sequence of answers

Сгенерированный вопрос	Ответ
Who was at the park?	John
What was he doing there?	swinging
On what?	on the wings
Who was he with?	his friend
Named?	Tim
What was he doing?	played on the side
What did John asked him?	if he could play on the slide
What did he say?	no

В статье [11] используется модель Refine Network (RefNet), которая пытается имитировать человеческий процесс создания вопросов, сначала создавая первоначальный черновик, а затем уточняя его. Модель обращает внимание как на исходный отрывок, так и на черновой вопрос и создает грамматически более правильный и полный вопрос. Реализуется это за счет двух декодеров: один формирует черновой вариант, второй формирует конечный вариант.

При этом второй декодер поддается улучшению по таким показателям, как беглость и возможность быть отвеченным (answerability) за счет использования дополнительной модели Reward-RefNet. Она проводит сравнение уточненного вопроса с результатами базового алгоритма (baseline) и вознаграждает систему за изменения, которые улучшают соответствующую метрику во время обучения.

Для оценки беглости используется балл BLEU. Для оценки возможности быть ответственным используется балл, фиксирующий, содержит ли вопрос требуемые именованные сущности, важные слова, служебные слова, типы вопросов.

Пример исходного отрывка текста (жирным шрифтом выделен ответ, относительно которого генерируется вопрос): Liberated by **Napoleon’s** army in 1806, Warsaw was made the capital of the newly created Duchy of Warsaw.

Таблица 3
Table 3

Результаты генерации вопросов для разных моделей
Results of generating questions for different models

Модель	Сгенерированный вопрос
Baseline	What was the capital of the newly duchy of Warsaw?
RefNet	Who liberated Warsaw in 1806?
Reward-RefNet	Whose army liberated Warsaw in 1806?

Проблема генерации сложных вопросов также поддается решению. Например, в статье [12] предлагается комплексное решение DQG (Deer Question Generation), состоящее из трех частей:

- 1) новый кодировщик графа, который включает механизм внимания в нейронную сеть Gated Graph (GGNN); используется для динамического моделирования взаимодействия между различными семантическими отношениями;
 - 2) усовершенствование вложений отрывков текста на уровне слов и представлений семантического графа на уровне узлов; используется для получения унифицированных представлений отрывков с учетом семантики для декодирования вопросов;
 - 3) введение вспомогательной задачи выбора контента, которая тренируется совместно с декодированием вопросов; используется для выбора подходящих контекстов в семантическом графе для формирования правильной цепочки рассуждений.
- Поверхностный вопрос генерируется из одного предложения и апеллирует лишь к одному факту, который можно извлечь из такого предложения.

Таблица 4

Table 4

Пример генерации простого вопроса**Example of generating a simple question**

Исходный текст	Кислород используется в клеточном дыхании и высвобождается в результате фотосинтеза, который использует энергию солнечного света для производства кислорода из воды
Вопрос	Какой жизненный процесс производит кислород в присутствии света?
Ответ	Фотосинтез

Сложный (глубокий) вопрос, требующий рассуждений и понимания текста в целом, формируется из нескольких непересекающихся релевантных предложений.

Таблица 5

Table 5

Пример генерации глубокого вопроса**Example of generating a deep question**

Отрывок А	Международный аэропорт Паго-Паго, также известный как аэропорт Тафуна, представляет собой общественный аэропорт, расположенный в 7 милях (11,3 км) к юго-западу от центрального делового района Паго-Паго, в деревне и на равнинах Тафуна на острове Тутила в Американском Самоа
Отрывок В	Аэропорт Хуна – это государственный общественный аэропорт, расположенный в одной морской миле (2 км) к юго-востоку от центрального делового района Хуна, Аляска
Вопрос	Находятся ли международный аэропорт Паго-Паго и аэропорт Хуны на американской территории?
Ответ	Да

ЗАКЛЮЧЕНИЕ

Среди существующих моделей генерации вопросов есть достаточно впечатляющие представители [10–12], которые способны устранить известные проблемы (поверхностные вопросы, невозможность учесть контекст). При этом есть возможность использовать открытый исходный код, например

модель DQG [12]. Также есть работы, которые могут послужить примером для проведения семантического анализа исходного документа [3, 8]. Для контроля же качества полученных результатов подойдут такие метрики, как BLUE, NIST и METEOR. Таким образом, задача автоматического создания тестов с приемлемым уровнем качества выглядит вполне реализуемой.

СПИСОК ЛИТЕРАТУРЫ

1. InfoWatch. Утечки информации ограниченного доступа: отчет за 9 месяцев 2020 г. – Экспертно-аналитический центр InfoWatch, 2020. – URL: https://d-russia.ru/wp-content/uploads/2020/12/infowatch_2020_9_monts_data_leak.pdf (дата обращения: 09.03.2022).

2. Information security awareness, a tool to mitigate information security risk: a literature review / O.S. Dada, E.A. Irunokhai, C.J. Shawulu, A.M. Nuhu, E.E. Daniel // Innovative Journal of Science. – 2021. – Vol. 3, N 3. – P. 29–54. – URL: https://www.researchgate.net/publication/353849520_Information_Security_Awareness_a_Tool_to_Mitigate_Information_Security_Risk_a_Literature_Review (accessed: 09.03.2022).

3. Клыч А.А., Кориунов А.Г., Стативко Р.У. Разработка подхода для генерирования вопросов // Актуальные проблемы теории и практики обучения физико-математическим и техническим дисциплинам в современном образовательном пространстве: IV Всероссийская (с международным участием) научно-практическая конференция. – Курск, 2020. – С. 367–373. – URL: <https://elibrary.ru/item.asp?id=44845075> (дата обращения: 09.03.2022).

4. Автоматическая генерация вопросов на основе текстов и графов знаний / М.А. Белянова, Г.И. Ревунков, Г.И. Афанасьев, Ю.Е. Гапанюк // Динамика сложных систем – XXI век. – 2020. – Т. 14, № 4. – С. 55–64. – DOI: 10.18127/j19997493-202004-06.

5. Вэньцзу Л. Онтологический подход к автоматической генерации вопросов в интеллектуальных обучающих системах // Доклады БГУИР. – 2020. – Т. 18, № 5. – С. 44–52. – DOI: 10.35596/1729-7648-2020-18-5-44-52.

6. Переходько И.В., Мячин Д.А. Оценка качества компьютерного перевода // Вестник Оренбургского государственного университета. – 2017. – № 2 (202). – С. 92–96. – URL: <https://elibrary.ru/item.asp?id=28904673> (дата обращения: 09.03.2022).

7. Легостина М.С. Метрики оценки качества машинного перевода // Инноватика-2019: сборник материалов XV Международной школы-конференции студентов, аспирантов и молодых ученых. – Томск, 2019. – URL: <https://elibrary.ru/item.asp?id=39222633> (дата обращения: 09.03.2022).

8. *Андреичев М.Д., Ференец А.А.* Разработка программного комплекса генерации вопросов по заданным субъектам при помощи семантической сети // Электронные библиотеки. – 2019. – Т. 22, № 2. – С. 68–71. – DOI: 10.26907/1562-5419-2019-22-2-68-94.

9. *Chen Y., Wu L., Zaki M.J.* Toward subgraph guided knowledge graph question generation with graph neural networks // Journal of Latex Class Files. – 2015. – Vol. 14, N 8. – URL: <https://arxiv.org/abs/2004.06015> (accessed: 09.03.2022).

10. *Chai Z., Wan X.* Learning to ask more: semi-autoregressive sequential question generation under dual-graph interaction // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – Association for Computational Linguistics, 2020. – P. 225–237. – DOI: 10.18653/v1/2020.acl-main.21.

11. Let's ask again: refine network for automatic question generation / P. Nema, A.K. Mohankumar, M.M. Khapra, B. Ravindran, B.V. Srinivasan // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). – Hong Kong, China, 2019. – DOI: 10.18653/V1/D19-1326.

12. Semantic graphs for generating deep questions / L. Pan, Y. Xie, Y. Feng, T.-S. Chua, M.-Y. Kan // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – Association for Computational Linguistics, 2020. – P. 1463–1475. – DOI: 10.18653/v1/2020.acl-main.135.

Лемякин Артём Викторович, лаборант кафедры защиты информации Новосибирского государственного технического университета. Область научных интересов – информационная безопасность. E-mail: a.v.lemyakin@gmail.com

Архипова Анастасия Борисовна, доцент кафедры защиты информации Новосибирского государственного технического университета. Основное направление научных исследований – математическое моделирование в информационной безопасности, оценка качества социально значимой деятельности. E-mail: arhipova@corp.nstu.ru

DOI: 10.17212/2782-2230-2022-1-85-98

Research of application of knowledge graphs to generate information security audit questions*

A.V. Lemyakin¹, A.B. Arkhipova²

¹ Novosibirsk State Technical University, Karl Marx Avenue, 20, Novosibirsk, 630073, Russian Federation, laboratory assistant of the Department of Information Security. E-mail: a.v.lemiyakin@gmail.com

² Novosibirsk State Technical University, Karl Marx Avenue, 20, Novosibirsk, 630073, Russian Federation, Associate Professor of the Department of Information Security. E-mail: arhipova@corp.nstu.ru

The article presents approaches to research on the use of knowledge graphs to generate information security audit issues. The significance of the human factor in the audit of information security as an element of socially significant activity is shown. As a solution to the formation of an objective research base, a method of automatically generating questions based on knowledge graphs capable of identifying a set of cartoons that are subject-object relations is proposed. Existing models and a set of dynamically generated metrics were analyzed. In particular, the subjective method of the five-point Makoto Nagao scale and the machine method of the automatic evaluation system based on the N-gram method. Advantages and disadvantages of research methods are analyzed, task setting for further research is presented.

Keywords: knowledge graph, awareness raising, security policies, automatic question generation, semantic analysis, knowledge control, tests

REFERENCES

1. InfoWatch. *Utechki informatsii ogranichennogo dostupa: otchet za 9 mesyatsev 2020 g.* [Restricted Information Leaks: 9 Month 2020 Report]. Available at: https://d-russia.ru/wp-content/uploads/2020/12/infowatch_2020_9_monts_data_leak.pdf (accessed 09.03.2022).
2. Dada O.S., Irunokhai E.A., Shawulu C.J., Nuhu A.M., Daniel E.E. Information security awareness, a tool to mitigate information security risk: a literature review. *Innovative Journal of Science*, 2021, vol. 3, no. 3, pp. 29–54. Available at: https://www.researchgate.net/publication/353849520_Information_Security_Awareness_a_Tool_to_Mitigate_Information_Security_Risk_a_Literature_Review (accessed 09.03.2022).
3. Klych A.A., Korshunov A.G., Stativko R.U. [Development of an approach for generating questions]. *Aktual'nye problemy teorii i praktiki obucheniya fiziko-matematicheskim i tekhnicheskim distsiplinam v sovremennom obrazovatel'nom*

* Received 10 February 2022.

prostranstve [Actual problems of the theory and practice of teaching physical, mathematical and technical disciplines in the modern educational space]. IV All-Russian (with international participation) scientific and practical conference, Kursk, 2020, pp. 367–373. (In Russian). Available at: <https://elibrary.ru/item.asp?id=44845075> (accessed 09.03.2022).

4. Belyanova M.A., Revunkov G.I., Afanasyev G.I., Gapanyuk Yu.E. Avtomaticheskaya generatsiya voprosov na osnove tekstov i grafov znaniy [Automatic question generation based on texts and knowledge graphs]. *Dinamika slozhnykh sistem – XXI vek = Dynamics of Complex Systems – XXI Century*, 2020, vol. 14, no. 4, pp. 55–64. DOI: 10.18127/j19997493-202004-06.

5. Wenzu L. Ontologicheskii podkhod k avtomaticheskoi generatsii voprosov v intellektual'nykh obuchayushchikh sistemakh [Ontological approach to automatic generation of questions in intelligent learning systems]. *Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki = Doklady BGUIR*, 2020, vol. 18, no. 5, pp. 44–52. DOI: 10.35596/1729-7648-2020-18-5-44-52.

6. Perekhodko I.V., Myachin D.A. Otsenka kachestva komp'yuternogo perevoda [Quality evaluation of computer translation]. *Vestnik Orenburgskogo gosudarstvennogo universiteta = Vestnik of the Orenburg State University*, 2017, no. 2 (202), pp. 92–96. Available at: <https://elibrary.ru/item.asp?id=28904673> (accessed 09.03.2022).

7. Legostina M.S. [Metrics of quality assessment of machine translate]. *Innovatika-2019* [Innovation-2019]. Collection of materials of the XV International School-Conference of Students, Graduate Students and Young Scientists, Tomsk, 2019. (In Russian). Available at: <https://elibrary.ru/item.asp?id=39222633> (accessed 09.03.2022).

8. Andreichev M.D., Ferenets A.A. Razrabotka programmnogo kompleksa generatsii voprosov po zadannym sub'ektam pri pomoshchi semanticheskoi seti [Development of a program complex for generating questions on given subjects using a semantic network]. *Elektronnye biblioteki = Russian Digital Libraries Journal*, 2019, vol. 22, no. 2, pp. 68–94. DOI: 10.26907/1562-5419-2019-22-2-68-94.

9. Chen Y., Wu L., Zaki M.J. Toward subgraph guided knowledge graph question generation with graph neural networks. *Journal of Latex Class Files*, 2015, vol. 14, no. 8. – Available at: <https://arxiv.org/abs/2004.06015> (accessed 09.03.2022).

10. Chai Z., Wan X. Learning to ask more: semi-autoregressive sequential question generation under dual-graph interaction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020, pp. 225–237. DOI: 10.18653/v1/2020.acl-main.21.

11. Nema P., Mohankumar A.K., Khapra M.M., Ravindran B., Srinivasan B.V. Let's ask again: refine network for automatic question generation. *Proceedings of*

the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019. DOI: 10.18653/V1/D19-1326.

12. Pan L., Xie Y., Feng Y., Chua T.-S., Kan M.-Y. Semantic graphs for generating deep questions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020, pp. 1463–1475. DOI: 10.18653/v1/2020.acl-main.135.

Для цитирования:

Лемякин А.В., Архипова А.Б. Исследование применения графов знаний для генерации вопросов аудита информационной безопасности // Безопасность цифровых технологий. – 2022. – № 1 (104). – С. 85–98. – DOI: 10.17212/2782-2230-2022-1-85-98.

For citation:

Lemyakin A.V., Arkhipova A.B. Issledovanie primeneniya grafov znanii dlya generatsii voprosov audita informatsionnoi bezopasnosti [Research of application of knowledge graphs to generate information security audit questions]. *Bezopasnost' tsifrovyykh tekhnologii = Digital Technology Security*, 2022, no. 1 (104), pp. 85–98. DOI: 10.17212/2782-2230-2022-1-85-98.