УДК 004.056.53:004.9
DOI: 10.31673/2412-9070.2025.025620
М. Г. ТРЕНЬОВ, аспірант;
ORCID: 0009-0002-8459-0599
І. М. СРІБНА, доктор техн. наук, доцент,
ORCID: 0000-0001-9242-2021
Державний університет інформаційно-комунікаційних технологій, Київ

МЕТОДИ ПОБУДОВИ ЗВ'ЯЗНОГО ГРАФА ЗНАНЬ ДЛЯ АНАЛІЗУ НОРМАТИВНИХ ДОКУМЕНТІВ У СФЕРІ КІБЕРБЕЗПЕКИ

У статті розглядається метод автоматизованої побудови графа знань зв'язності формальних моделей норм і вимог у сфері інформаційної безпеки. Зростаюча складність нормативних документів та їх вимог вимагає нових підходів до їх формалізації та аналізу. Запропонований метод ґрунтується на застосуванні алгоритмів обробки природної мови та онтологічного моделювання для створення зв'язного графа знань, що представляє формальні моделі норм і вимог. Основна увага приділяється автоматизації процесу формування таких графів, що дозволяє мінімізувати людський фактор, знизити ймовірність помилок та підвищити ефективність обробки великих масивів нормативної інформації.

У роботі розглядається структура та принципи побудови графа знань, включаючи механізми ідентифікації зв'язків між нормами, вимогами та іншими сутностями. Запропоновано методи автоматичного вилучення та аналізу семантичних зв'язків, а також алгоритми їх подальшої обробки для забезпечення когерентності отриманого графа. Дослідження базується на сучасних методах штучного інтелекту та обробки текстової інформації, що дозволяє підвищити рівень автоматизації процесів формалізації та систематизації вимог інформаційної безпеки.

Результати проведених експериментів демонструють ефективність розробленого підходу та його можливості щодо інтеграції з існуючими системами управління інформаційною безпекою. Використання графа знань дозволяє суттєво покращити аналіз нормативної документації, забезпечуючи глибше розуміння зв'язків між її елементами та підвищуючи якість прийняття рішень у сфері інформаційної безпеки. Запропонований метод може бути застосований у різних сферах, що вимагають автоматизації обробки нормативних актів та їх взаємозв'язків, зокрема у кібербезпеці, управлінні ризиками та регуляторному аналізі.

Ключові слова: граф знань, автоматизація, інформаційна безпека, формальні моделі, штучний інтелект, обробка природної мови, онтологічне моделювання.

Вступ

У цій статті розглядаються всі етапи реалізації графа знань, а також проблеми, з якими, можливо, доведеться зіткнутися під час створення власного екземпляра цієї абстракції. Крім цього, розглянуто методи створення інтерактивного візуального подання інформації для її ефективного зберігання у графі, а також практичні кроки щодо його реалізації та використання.

Розробку методу автоматизованої побудови графа знань зв'язності формальних моделей норм та вимог у галузі інформаційної безпеки пропонується розділити на чотири основні етапи.

По-перше, аналіз розробок у галузі візуалізації нормативно-правової бази, а також використання графової моделі у сфері інформаційної безпеки.

По-друге, дослідження способів формального представлення норм права та вимог стандартів для використання комп'ютерних технологій у сфері права.

По-третє, аналіз норм права щодо можливості створення їх машинозчитуваного подання та автоматизованого способу поповнення бази знань.

По-четверте, розробка інструментів для роботи експертів у сфері інформаційної безпеки над формальними моделями норм та вимог.

Аналіз досліджень щодо способів візуалізації нормативно-правової бази, а також використання графової моделі у сфері інформаційної безпеки

Автори статті [1] пропонують аналіз існуючих підходів представлення знань у вигляді графів. Дослідження включає вилучення відношень з тексту, методи вбудовування відношень для створення посилань, огляд існуючих графів знань, порівняння сховищ графів знань, імпорт графів знань. Стаття корисна для первинного аналізу роботи з графами знань, визначення концепції побудови, зберігання та імпорту графів знань. Також зазначається, що підходи, запропоновані у статті, застосовні до будь-якої галузі знань, тобто універсальні.

Основна робота у статті [2] поділена на частини. Обговорюється створення бази знань з кібербезпеки відповідно до триетапної процедури та пропонується структура для створення бази знань:

– по-перше, вилучення інформації шляхом збору та аналізу структурованих та неструктурованих даних;

- по-друге, побудова онтології відповідно до отриманої інформації комп'ютерних атак;

- по-третє, метод машинного навчання для отримання об'єктів, пов'язаних з комп'ютерною атакою.

Дослідження [3] визначає створення онтології з інформаційної безпеки в рамках прецедентного підходу до опису комп'ютерних загроз. Онтологія, що розроблюється досить повно описує предметну область комп'ютерних загроз та є концептуальною. Автори запропонували спосіб створення на основі онтології бази даних основних відомих прецедентів комп'ютерних атак та порушень інформаційної безпеки, з їх описом, методами ідентифікації та способами усунення наслідків.

У статті [4] автори пропонують нову загальну структуру захисту промислової мережі на основі даних. З метою підвищення якості аналізу відношення сутностей даних про кібербезпеку, авторами пропонується прототип нової моделі вилучення взаємоз’вязків кібербезпеки ResPCNN -ATT.

В дослідженні [5] розроблено модель інформаційної безпеки баз знань, що включає такі складові: функції захисту в базах знань; методи захисту баз знань, техніко-економічні показники методів захисту в базах знань.

Порівняно з представленими роботами, у статті пропонується вирішення кількох завдань:

– вивчення методів формального представлення норм права, вимог та стандартів для використання комп'ютерних технологій у галузі права;

– розробка формального подання предметної області;

– аналіз норм права щодо можливості створення їх машинозчитуваного представлення та автоматизованого способу поповнення бази знань;

– розробка інструментів для роботи експертів у галузі інформаційної безпеки над формальними моделями норм та вимог.

Опис способу формального подання норм права та вимог стандартів для використання комп'ютерних технологій у галузі права

Найчастіше інструментом, що застосовується під час візуалізації реляційних типів даних є графова модель, яка також застосовується і в інших напрямках візуалізації. Модель представляє не тільки дані, а й відносини між даними. Графова модель є структурою даних, що складається із зв'язаних сутностей. Подання інформації у вигляді графа знань не нове, але останнім часом воно набуло популярності завдяки його використанню у додатках штучного інтелекту. Нижче наведено основні терміни та визначення, що використовуються у статті.

Граф знань - це семантична мережа, в якій зберігаються сутності та взаємоз’язки між сутностями у вигляді графа. Перевага використання графа знань – це можливість обробки більшої кількості запитів, ніж традиційні методи зберігання. Граф знань – гнучка форма зберігання та візуалізації даних, які легко оновити [6].

Ключове слово – особливо важливе, загальнозрозуміле, ємне та показове слово в тексті, набір яких може дати високорівневий опис його змісту.

Нормативно-правовий акт – концепт, що відповідає офіційному документу, виданому в установленому порядку органом державної влади, органом місцевого самоврядування або посадовою особою, що містить правові норми.

Онтологія – класи, властивості класів, правила, що у сукупності відбивають формальну концепцію предметної області; спроба всеосяжної та докладної формалізації певної галузі знань за допомогою концептуальної схеми.

Ставлення – зв'язок між сутностями, що є ребром графа знань.

Посилання – зв'язок, який веде від фрагмента тексту одного нормативно-правового акта до іншого нормативно-правового акта.

Сутність - норми права, що є вузлами (вершинами) графа знань.

Тип відносини – спосіб вказати, що означає відношення між сутностями, семантику відповідного відношення.

Сьогодні терміни, які застосовуються під час побудови баз і графів знань використовують у різних контекстах. Моделювання та формальне подання схеми даних у вигляді баз та графів знань забезпечує набагато більші можливості, ніж традиційні бази даних або об'єктно-орієнтований підхід.

Розробка формального подання предметної галузі

У цьому розділі представлено формальне подання онтології графа знань нормативно-правової бази у сфері інформаційної безпеки.

Відповідно до визначення, формалізація онтології – процес вираження концептуалізації предметної галузі відповідно до парадигми представлення знань, запропонованої мовою моделювання.

Формалізацію онтології проведено відповідно до умов:

– онтологія один із інструментів, необхідний для моделювання предметної області;

– онтологія містить перелік ключових слів даної предметної галузі;

– знання про зміст ключових слів, подане онтологією, має бути очевидним для будь-якого експерта в даній предметній галузі.

Мета онтології – підвищити кількість і якість загальних властивостей предметної області, що описуються, не залежать від її конкретних реалізацій. Формальною онтологією предметної області *EA* називається пара $< e, a >$, де *e* – безліч ключових слів предметної області, *a* – безліч аналітичних пропозицій, відносин, посилань, що описує зміст цих ключових понять.

Онтологія предметної області включає:

- словник ключових понять, що використовуються в даній предметній галузі;

- сукупність взаємозв’язків, що забезпечують коректну інтерпретацію понять та їх правильне використання.

Подання знань як онтологій застосовується заради семантичної інтеграції інформаційних ресурсів, коректної інтерпретації змісту чи назви текстових документів, представлених з допомогою природної мови. На основі онтологій розробляються бази знань та графи знань.

Тому, всю множину *S* пропозицій, які є вірними в предметній області *EA*, називатимемо теорією предметної області *EA*.

Тоді, якщо *E, A* – формальна онтологія предметної галузі *EA*, а *S* – теорія предметної галузі *EA*, то кожен елемент *E* є елементом *S*.

Візьмемо до уваги, що якщо деяке твердження бази даних не є істинним, то твердження є хибним. Тоді як для бази знань у цьому випадку твердження є ні істинним, ні хибним. Ця властивість суттєво впливає на те, які факти вважаються логічно наступними із заданої бази знань і поняття логічного слідування в ці факти бази знань.

Аналіз норм права щодо можливості створення їх машинозчитуваного подання та автоматизованого способу поповнення бази знань

Законодавча база України містить множину нормативно-правових актів (НПА) у сфері інформаційної безпеки. Аналіз нормативно-правових актів вільного доступу дозволить сформувати основу знань. Алгоритм аналізу нормативно-правового акта представлений у вигляді алгоритму на рис. 1.

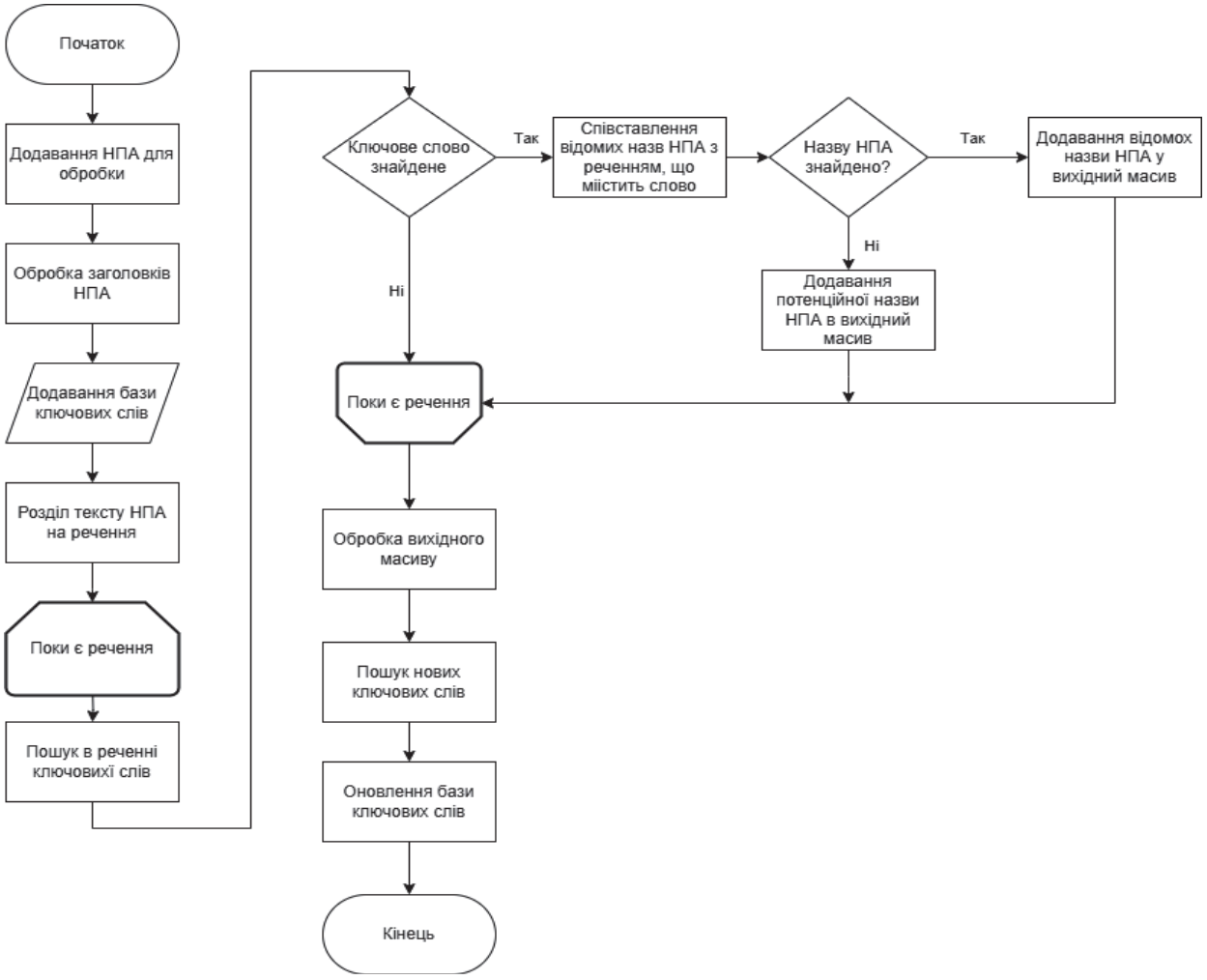

Рис. 1. Алгоритм аналізу нормативно-правового акту

Опис основних етапів алгоритму наведено нижче.

1. Додавання нормативно-правових актів. На даному етапі відбувається додавання користувачем нормативно-правових актів, які надалі вважатимуться нормативно-правовою базою і на підставі яких пропонується будувати граф знань. Додавання НПА відбувається через вибір директорії з необхідним набором документів текстового формату.

2. Обробка заголовків НПА. У завантажених документах визначаються їх заголовки та вносяться до окремого масиву. При цьому мається на увазі, що файли, знайдені в каталозі, можуть мати, а можуть не відповідати шаблону найменування нормам законодавства назви (%тип документа% %видавець% від %дата % %номер% %назва%). Набір ключових слів визначається на основі шаблонів найменування норм законодавства в напівавтоматичному режимі.

3. Поділ тексту НПА на речення. У цьому циклі здійснюється основна робота з визначення зв'язків між НПА таким чином: здійснюється пошук за всіма пропозиціями щодо явного змісту в реченні або пропозицій, що потенційно містять ключові слова заголовків НПА.

4. Обробка вихідного масиву. Масив отриманих із пропозицій назв НПА (посилань на документи) співвідноситься із заголовками з поточної множини завантажених НПА. Наступним кроком виявляються збіги між отриманим посиланням та заголовком за допомогою ключових

слів. Знайдені ключові слова або потенційні ключові слова додаються до масиву, який згодом редагується.

5. Пошук нових ключових слів. Користувачеві пропонується аналізувати потенційні ключові слова щодо додавання їх до бази ключових слів для подальшого використання при обробці НПА. Результатом аналізу користувача на цьому етапі є оновлена база ключових слів.

Насправді вилучення сутностей полягає у витягу з тексту нормативно-правового акта посилань на інші нормативно-правові акти. Процес побудови графа знань складається з вилучення сутностей та їхнього взаємозв'язку відносинами з базою нормативно-правових актів через онтологію. Це сформує граф знань і надасть можливість експертам у галузі інформаційної безпеки проводити міркування щодо зв'язності документів. Розроблену онтологію також можна пов'язати із зовнішніми знаннями та онтологіями.

Основний підхід до вилучення взаємозв’язків ґрунтується на техніках обробки природної мови, включаючи тегування частини мови, синтаксичний аналіз, розпізнавання іменованих сутностей.

Онтології представляються у вигляді графів знань, щоб візуалізувати взаємозв’язки між сутностями галузі знань. Графи знань – найкращі засоби представлення сутностей та відносин між ними. Онтологія складається з набору класів з атрибутами і відносинами.

База знань є семантичним графом знань, що описує семантику джерел інформації. Таким чином, онтологія, що вийшла, пропонує для аналітиків і професіоналів в даній області інструмент для аналізу. Граф знань є результатом зв'язування галузі знання та моделі подання даних, а саме онтології. У нашому випадку граф знань є зв'язністю нормативно-правових актів, витягнутих у спеціальну онтологію.

Розробка інструментів для роботи експертів у галузі інформаційної безпеки над формальними моделями норм та вимог

Реалізацією запропонованих алгоритмів та формалізованих уявлень є розроблена програма для ЕОМ. Програма є інструментом для роботи експертів у галузі інформаційної безпеки над формальними моделями норм та вимог у галузі законодавства. На рис. 2 представлений інтерфейс програми для ЕОМ.

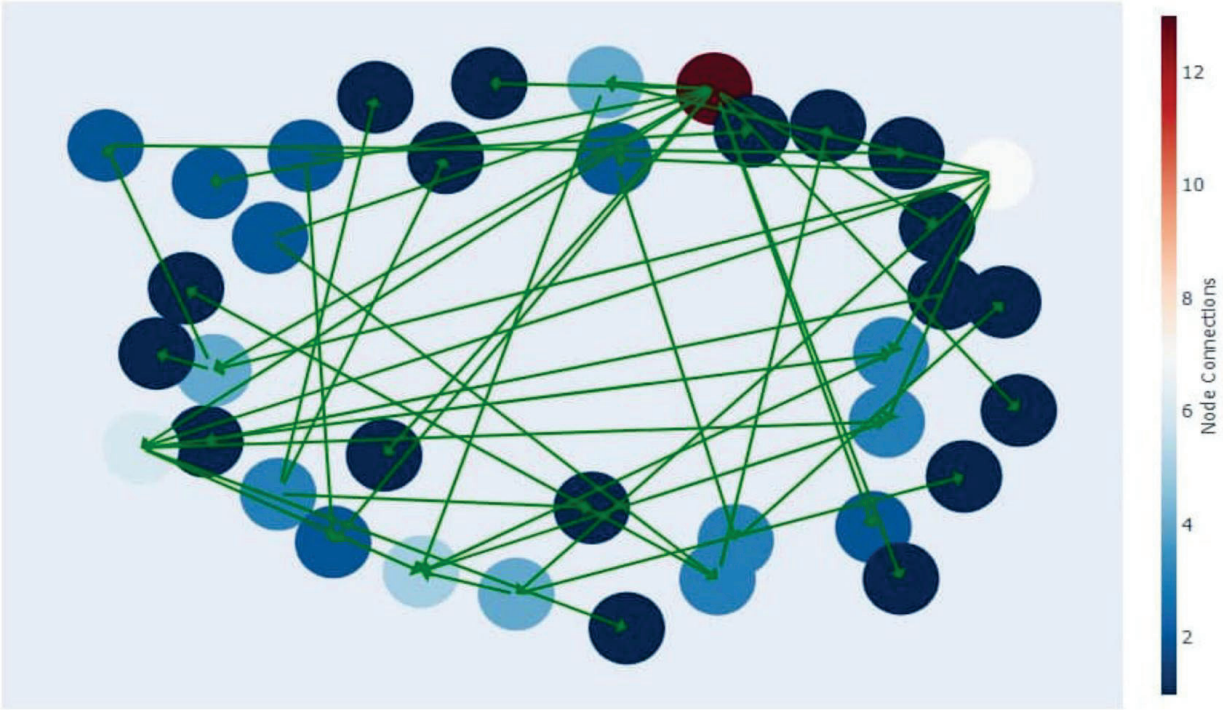

Рис. 2. Відносини сутностей нормативно-правової бази

Інтерфейс складається з:

- елементів управління;

- зони управління завантаженими нормативно-правовими актами;

- галузі візуалізації відносин сутностей нормативно-правової бази.

Елементи управління дозволяють користувачеві визначити вхідні дані для роботи програми, коригувати склад НПА, ознайомитись з посібниками користувача та адміністратора, зберегти чи відновити граф знань.

Зона керування завантаженими НПА відображає склад документів, вибраних користувачем як вхідні дані. Перелік документів, що відображаються у цій зоні, має функціонал інтерактивної взаємодії з сферою візуалізації відносин сутностей нормативно-правової бази.

Область візуалізації відносин сутностей нормативно-правової бази є основним полем роботи з програмою для ЕОМ. Область візуалізації взаємозв'язків відображає зв'язок всіх оброблюваних користувачем НПА. Візуалізація взаємозв'язків реалізована у вигляді графа знань та є інтерактивною. Користувачеві доступний функціонал переміщення вершин графа, взаємодії з вершинами та ребрами графа для отримання додаткової інформації. Візуально кількість зга-док про документ в інших НПА відрізняється кольором. В галузі візуалізації взаємозв'язків представлена шкала, що демонструє співвідношення кількості згадок документа в інших НПА (кількість зв'язкових ребер графа знань з конкретною вершиною) та колірного відтінку.

Метою застосування засобу автоматизованої побудови нормативно-правової бази є зниження трудоємності задачі аналізу вимог нормативно-правової бази. Аналіз вимог нормативно-правової бази проводиться аналітиками інформаційної безпеки періодично.

Висновки

Представлені розробки пропонують метод автоматизованої візуалізації взаємозв'язків між нормативно-правовими актами у вигляді побудови графа знань, що дозволяє прискорити аналіз інформації фахівцем у сфері інформаційної безпеки. Основною перевагою запропонованого методу є автоматизований спосіб поповнення бази ключових слів, а також швидкість аналізу даних, оскільки автоматизована реалізація здатна в десятки разів скоротити час, що витрачається на пошук, аналіз і систематизацію інформації, викладеної в нормативно-правовій базі.

Список літератури

1. Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge Graph Embedding: A Survey of Approaches and Applications. IEEE Transactions on Knowledge and Data Engineering, 29, 2724-2743. https://doi.org/10.1109/TKDE.2017.2754499.

2. A Practical Approach to Constructing a Knowledge Graph for Cybersecurity Yan Jia, Yulu Qi, Huaijun Shang, Rong Jiang, Aiping Li https://doi.org/10.1016/j.eng.2018.01.004.

3. Arogundade, O., Abayomi-Alli, A., & Misra, S. (2020). An Ontology-Based Security Risk Management Model for Information Systems. Arabian Journal for Science and Engineering, 45, 6183 - 6198. https://doi.org/10.1007/s13369-020-04524-4.

4. Guowei Shen, Wanling Wang, Qilin Mu, Yanhong Pu, Ya Qin, and Miao Yu Data-Driven Cybersecurity Knowledge Graph Construction for Industrial Control System Security doi.org/10.1155/2020/8883696.

5. Kireeva, N., Pozdnyak, I., & Gazizulina, A. (2019). Filling a Knowledge Base for Expert System in Information Security. IOP Conference Series: Materials Science and Engineering, 618. https://doi.org/10.1088/1757-899X/618/1/012085.

6. Shen, T., Zhang, F., & Cheng, J. (2022). A comprehensive overview of knowledge graph completion. Knowl. Based Syst., 255, 109597. https://doi.org/10.1016/j.knosys.2022.109597.

M. Trenov, I. Sribna

METHODS FOR CONSTRUCTING A CONNECTED KNOWLEDGE GRAPH FOR ANALYZING REGULATORY DOCUMENTS IN THE FIELD OF CYBERSECURITY

The article discusses the method of automated construction of a knowledge graph of the conectivity of formal models of norms and requirements in the field of information security. The growing complexity of regulations and their requirements requires new approaches to their formalisation and analysis. The proposed method is based on the use of natural language processing and ontological modelling algorithms to create a connected knowledge graph representing formal models of norms and requirements. The main focus is on automating the process of forming such graphs, which minimises the human factor, reduces the likelihood of errors and increases the efficiency of processing large amounts of regulatory information.

The paper discusses the structure and principles of building a knowledge graph, including mechanisms for identifying links between norms, requirements and other entities. Methods of automatic extraction and analysis of semantic relations, as well as algorithms for their further processing to ensure the coherence of the resulting graph are proposed. The study is based on modern methods of artificial intelligence and textual information processing, which allows to increase the level of automation of the processes of formalisation and systematisation of information security requirements.

The results of the experiments demonstrate the effectiveness of the developed approach and its ability to integrate with existing information security management systems. The use of a knowledge graph can significantly improve the analysis of regulatory documentation, providing a deeper understanding of the relationships between its elements and improving the quality of decision-making in the field of information security. The proposed method can be applied in various areas requiring automation of processing of regulations and their interrelationships, including cybersecurity, risk management and regulatory analysis.

Keywords: knowledge graph, automation, information security, formal models, artificial intelligence, natural language processing, ontological modelling.