Хмельницький національний університет
Факультет інформаційних технологій
Кафедра кібербезпеки

ДИПЛОМНА РОБОТА МАГІСТРА

Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів
Назва теми

Галузь знань 12 – Інформаційні технології

Спеціальність 125 –Кібербезпека

ДРКб. 170151.21.01.02 ПЗ

Виконав: студент 2 курсу, група КБм-21-1 _____ Пахар О.В.
Підпис

Керівник доц., д. т. н, професор кафедри КБ _____ Касянчук М.М.
Підпис

Нормоконтролер ст. викладач кафедри КБ _____ Мостовий С.В.
Підпис

До захисту допускаю:
Зав. кафедри КБ к.т.н., доц _____ Кльоц Ю.П.
Підпис

7 12 2022р.

Хмельницький, 2022

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет Інформаційних технологій
Кафедра Кібербезпеки
Освітній рівень магістр
Галузь знань 12 Інформаційні технології
Спеціальність 125 Кібербезпека
Освітня програма Освітньо-професійна програма підготовки магістра

ЗАТВЕРДЖУЮ
Завідувач кафедри кібербезпеки
к.т.н. доцент Кльоц Ю.П.
"4" 04 2022 року

ЗАВДАННЯ
НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ)

Пахару Олександру Валерійовичу
(прізвище, ім'я, по батькові)

1. Тема проекту (роботи) Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів
Науковий керівник Касянчук Михайло Миколайович, д.т.н., доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджена наказом №87 ректора університету додаток №26 від 01.07.2022
2. Строк подання студентом проекту (роботи) на кафедру 5.12.2022.
3. Вихідні дані до проекту (роботи) Провести дослідження сучасного стану вразливостей та загроз інформаційної безпеки та засобів захисту інформації. Розробити моделі бази даних інтернет-ресурсів, потоку даних тематичних ресурсів з метою виявлення вразливостей, загроз безпеки інформації. Розробити алгоритм прогнозування вразливостей, загроз інформаційної безпеки, на основі отриманих результатів обробки інформації тематичних ресурсів та алгоритм фільтрації потоку текстових повідомлень інтернет-ресурсів. Розробити інформаційно-аналітичну систему автоматизації проведення аналізу потоку даних тематичних інтернет-ресурсів та прогнозування появи нових вразливостей, загроз безпеки інформації з використанням логічного нечіткого виводу.
4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)
Дослідження сучасних вразливостей та загроз інформаційної безпеки та засобів захисту інформації. Моделі та алгоритми прогнозування вразливостей та загроз інформаційної безпеки на основі інтернет-ресурсів. Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів. Інформаційно-аналітична системи прогнозування вразливостей та загроз інформаційної безпеки.
5. Перелік графічного матеріалу (із зазначенням обов'язкових креслень) 1,2.Тема, мета магістерської роботи, об'єкт, предмет, задачі дослідження, наукова новизна, практична цінність, апробація роботи. 3. Класифікація вразливостей та загроз інформаційної безпеки, характерних для інтернет-ресурсів. 4. Інформаційне наповнення тематичних інтернет-ресурсів. 5.Алгоритм прогнозування вразливостей та загроз інформаційної безпеки. 6. Модель потоку текстових повідомлень та бази даних інтернет-форуму. 7. Модель потоку текстових повідомлень тематичних форумів. 8. Алгоритм фільтрації потоку текстових повідомлень та статистичного аналізу інформаційної безпеки. 9. Метод прогнозування вразливостей та загроз інформаційної безпеки. 10. Структура та функції інформаційно-аналітичної системи. 11. Діаграма діяльності інформаційно-аналітичної системи. 12. Показники якості прогнозування інформаційно - аналітичної системи. 13. Висновки.

6. Консультанти розділів дипломного проекту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання
Відповідальний за оформлення ДП	Мостовий С.В., ст. викладач		

7. Дата видачі завдання: «01» лютого 2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської роботи	Строк виконання етапів роботи	Прим
1	Грунтовне дослідження предметної області	23.02.2022	Вик
2	Визначення структури, змісту магістерської роботи	10.03. 2022	Вик
3	Опрацювання магістерської роботи - перший розділ	5.04. 2022	Вик
4	Робота статтею за результатами дослідження	2.05. 2022	Вик
5	Опрацювання магістерської роботи - другий розділ	3.06. 2022	Вик
6	Опрацювання магістерської роботи - третій розділ	5.09. 2021	Вик
7	Опрацювання магістерської роботи - четвертий розділ	3.10. 2022	Вик
8	Опрацювання та підготовка ілюстративного матеріалу	7.11. 2022	Вик
9	Оформлення магістерської роботи - текстової і графічної частини	18.11. 2022	Вик
10	Попередній захист магістерської роботи	25.11. 2022	Вик
11	Захист магістерської роботи на засіданні ЕК	7.12. 2022	Вик

Студент _____ О.В. Пахар
Підпис Ініціали, прізвище

Керівник проекту (роботи) _____ М.М. Касянчук
Підпис Ініціали, прізвище

Тема дип
інформаційної б
Автор робо
Керівник р
Пояснюва
Перелік кл
форуми, джере
Мета роб
вразливостей,
методів та ін
інтернет-ресу
Реалізова
вразливостей
тематичних р
нових вразли
оцінити свої
відповідних
самим підви
від реалізаці

30.11

АНОТАЦІЯ

Тема дипломної роботи: «Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів».

Автор роботи: студент групи КБм – 21 – 1 Пахар О.В.

Керівник роботи: д.т.н., доц. Касянчук М.М.

Пояснювальна записка: 80 с., 24 рисунки, 10 таблиць, 30 джерел.

Перелік ключових слів: моделі, алгоритми, потік повідомлень, тематичні форуми, джерела повідомлень, канали зв'язку.

Мета роботи - підвищення ефективності методів та засобів виявлення вразливостей, загроз безпеки інформації на основі запропонованих алгоритмів, методів та інформаційно-аналітичної системи аналізу потоку повідомлень інтернет-ресурсів.

Реалізований в інформаційно-аналітичній системі алгоритм прогнозування вразливостей та загроз безпеки інформації на основі аналізу потоку даних тематичних ресурсів дозволяє автоматизувати інформаційний процес виявлення нових вразливостей, загроз, надає фахівцям інформаційної безпеки можливість оцінити своєчасно ступінь захищеності ресурсів та при необхідності вжити відповідних заходів щодо нейтралізації можливих загроз та вразливостей, тим самим підвищити інформаційну безпеку обчислювальних комп'ютерних систем від реалізації нових мережевих комп'ютерних атак.

30.11.22

ANNOTATION

Theme of thesis: " The method of forecasting information security vulnerabilitiesbased on data from Internet resources".

The author of the work: a student of the group KBm - 20 - 1 Zhilevich M.I..Head of work: Ph.D, associate Professor Kasyanchuk M.M.

Explanatory note: 80 p., 24 figures, 10 tables, 30 sources.

List of keywords: models, algorithms, message flow, thematic forums, message sources, communication channels.

The purpose of the work - increasing the effectiveness of methods and means of detecting vulnerabilities, threats to information security based on the proposed algorithms, methods and information-analytical system of analyzing the flow ofmessages of Internet resources.

The algorithm for forecasting vulnerabilities and threats to information security implemented in the information and analytical system, based on the analysis of the data flow of thematic resources, allows automating the information process of detecting new vulnerabilities and threats, provides information security specialists with the opportunity to assess the degree of security of resources in a timelymanner and, if necessary, take appropriate measures to neutralize possible threats andvulnerabilities, thereby increasing the information security of computing computer systems against the implementation of new network computer attacks.

30.11.22

ЗМІСТ

ВСТУП

На сучасному етапі на більшість сфер діяльності суспільства зростає вплив глобальних інформаційних технологій. Відзначаються, при цьому, високі темпи розвитку світових єдиних телекомунікаційного та інформаційного просторів, сформувалися в суспільстві нові соціальні групи, виявляється значний вплив на сформований історично спосіб життя людей. На тлі стрімкого розвитку інформаційних технологій відзначається, зростання активності різноманітності комп'ютерних атак, здійснюваних і запланованих із застосуванням сучасних новітніх технологій.

На сучасному етапі, проблеми інформаційної безпеки розвитку суспільства у більшості сфер їх діяльності виходять на передній план. Це пов'язано зі значним зростанням кількості реалізованих проектів інформатизації. Більшість реалізованих проектів інформатизації спрямовані на побудову єдиного телекомунікаційного та інформаційного простору з метою оптимізації процесів обробки різноманітної інформації великих об’ємів, наприклад забезпечення оперативного доступу до інформації, надійного зберігання даних для користувачів інформаційного обміну

Актуальними та пріоритетними на сучасному етапі є задачі аналізу, класифікації виявлення існуючих механізмів реалізації атак та загроз інформаційної безпеки, які можуть призвести до отримання несанкціонованого доступу до конфіденційної інформації, порушення функціонування інформаційних систем. Таким чином, постає задача визначення заходів протидії атакам та загрозам, усунення вразливостей, оцінки заданої можливої шкоди, розробка нормативно-правової бази, механізмів захисту та критеріїв безпеки. Важливість даних проблем пов'язана з наступними основними факторами: зростанням різноманітності та кількості засобів комп’ютерної техніки та сфер людської діяльності їх застосування; високим рівнем довіри до інформаційно-пошукових систем обробки та управління даними; зростанням числа користувачів

інформаційного простору взаємодії; накопиченням великих об’ємів різнотипної інформації, інтенсивним обміном потоком даних в мережі між користувачами, з використанням широкого спектра механізмів доступу до конфіденційних ресурсів, інформаційних процесів; промисловим шпигунством та конкурентною боротьбою у сфері інформаційних послуг суспільства; недостатньою кількістю, на сучасному етапі, фахівців високої кваліфікації в області інформаційної безпеки, ринковими відношеннями в області розробки програмного забезпечення, обслуговування, розповсюдження, виробництва обчислювальної комп'ютерної техніки для реалізації інформаційної безпеки; різноманіттям атак, загроз і різнотипних каналів отримання несанкціонованого доступу до конфіденційних ресурсів та диференціацією негативних наслідків.

Таким чином виникає потреба у проведенні захисту комп'ютерних систем та інформаційних ресурсів та від блокування, несанкціонованого доступу до даних, знищення, та інших злочинних, небажаних, різноманіття та кількість яких постійно зростає. За оцінками, проведеними експертними організаціями, збитки в інформаційній сфері від злочинів в мережі Інтернет щорічно оцінюються в мільярди доларів.

Аналіз проведеного дослідження поточного стану в області інформаційної безпеки показує, що темпи розвитку інформаційних та комп'ютерних технологій значно випереджають процес створення програмно-апаратного забезпечення в області інформаційної безпеки. Пріоритетними, в даній ситуації, є задача аналізу, класифікації, виявлення діючих механізмів та засобів проведення атак і загроз інформаційній безпеці системи, які можуть призвести до отримання несанкціонованого доступу до конфіденційних даних, порушення функціонування інформаційної системи, визначення заходів протидії атакам та загрозам, оцінка заданої шкоди, розробка нормативно-правової бази, механізмів захисту та критеріїв інформаційної безпеки системи протидії.

На сьогодні не існує єдиного підходу до вирішення проблеми захищенності інформаційно-пошукових систем, стосовно предметних областей: розробниками

програмно-апаратного захисту інформації пропонуються відповідні компоненти на вирішення конкретних задач; забезпечення надійного захисту інформаційних ресурсів потребує реалізації відповідних технічних та організаційних заходів в комплексі, що супроводжуються розробкою відповідної документації [1,2].

Таким чином, аналіз проведеного дослідження вказує на необхідність вирішення наступних задач для забезпечення інформаційної безпеки: формування основ для опису процесів реалізації та виникнення атак, загроз, вразливостей інформаційної безпеки системи в умовах невизначеності та непередбачуваності їх прояву; розробка відповідних засобів забезпечення захисту конфіденційної інформації на основі проведеного дослідження та класифікації вразливостей, загроз; визначення загальних підходів до створення інформаційних систем забезпечення захисту конфіденційних даних, механізмів управління захистом на різних рівнях діяльності суспільства.

Одним із підходів вирішення наведених задач є застосування існуючих систем виявлення комп'ютерних атак, для захисту інформації [1, 3-5]. В аналітичних оглядах компаній, у сфері інтернет-технологій та захисту інформації, таких як Positive Technologies, Trustware, Kaspersky Labs, Symantec, наводяться висновки, що в останні роки на інформаційно-пошукові системи, про зростання кількості загроз, а також трансформації засобів, які використовуються нелигітимними кореспондентами, у повноцінну інформаційну зброю [3-7].

Більшість сучасних програмно-апаратних систем виявлення комп'ютерних загроз та атак працюють із використанням підходів сигнатурного аналізу та фіксації інтернет-мережевих аномалій. Дані підходи мають недоліки, пов'язані із використанням потужних обчислювальних ресурсів на їх реалізацію, а також, при цьому, при виявленні нових комп'ютерних загроз мають низьку ефективність [8].

Основними джерелами надходження знань про вразливості та атаки інформаційної безпеки є бази даних та знань, створювані державними, українськими та зарубіжними комерційними структурами. Наповнення інформаційних баз даних здійснюється із залученням дослідних авторитетних

центрів експертним шляхом. Разом з тим, інформація, що міститься в базах даних та знань вразливостей та загроз не є повною. Таким чином, актуальним залишається задача виявлення доступних інформаційних ресурсів, про комп'ютерні загрози, віруси, вразливості, а також можливість доступу до результатів досліджень компаній з виявлення загроз інформаційної безпеки систем протидії. Одним із джерел надходження інформації про вразливості та загрози інформаційної безпеки є інтернет-ресурси (інформаційні соціальні ресурси, також анонімні, які відносяться до інформаційної безпеки), обумовлено популярністю спеціалізованих інтернет-ресурсів, хто цікавиться відповідними предметними областями. Події, що відбуваються в відповідних предметних областях, є предметом для обговорення учасників дискусійних тематичних інтернет-майданчиків. Даний фактор дозволяє прогнозувати виникнення вразливостей, атак, загроз безпеки інформації, ґрунтуючись на проведенні аналізу потоку повідомлень тематичних інтернет-ресурсів. Як один із підходів вирішення задачі, магістерської роботи розглянуто можливість використання інформаційних систем нечіткого логічного виводу, вхідними даними яких є результати проведеного аналізу інформації тематичних інтернет-ресурсів. Фахівець безпеки інформації, зможе оцінити ступінь інформаційної небезпеки ресурсів, на основі отриманих результатів прогнозування виникнення вразливості, атаки, загрози, оцінити коректність моделі загроз безпеці інформації та задіяти протидію щодо нейтралізації вразливостей.

В результаті аналізу проведеного дослідження в області безпеки інформації, виявлено невирішенні питаня, стосовно автоматизації інформаційних процесів прогнозування вразливостей та загроз безпеки інформації.

Таким чином, актуальною залишається задача проектування та розробки методу, системи прогнозування, виявлення вразливостей, загроз безпеки інформації.

Об'єкт дослідження. Потоки повідомлень тематичних інтернет-ресурсів, що містять відомості про вразливості та загрози безпеки інформації.

Предмет дослідження. Алгоритми, методи прогнозування вразливостей, загроз безпеки інформації, підходи до їх реалізації з використанням автоматизованого аналізу даних інтернет-ресурсів.

Мета магістерської роботи - підвищення ефективності методів та засобів виявлення вразливостей, загроз безпеки інформації на основі запропонованих алгоритмів, методів та інформаційно-аналітичної системи аналізу потоку повідомлень інтернет-ресурсів.

Для досягнення поставленої мети в магістерській роботі вирішені наступні задачі: аналіз сучасних вразливостей та загроз інформаційної безпеки та засобів захисту інформації; запропонована модель бази даних інтернет-ресурсів, потоку даних тематичних ресурсів з метою виявлення вразливостей, загроз безпеки інформації; розробка алгоритму прогнозування вразливостей, загроз інформаційної безпеки, на основі отриманих результатів обробки інформації тематичних ресурсів та алгоритму фільтрації потоку текстових повідомлень інтернет-ресурсів, які містять інформацію про вразливості, загрози безпеки інформації, та статистична оцінка їх критеріїв; розробка інформаційно-аналітичної системи автоматизації проведення аналізу потоку даних тематичних інтернет-ресурсів та прогнозування появи нових вразливостей, загроз безпеки інформації з використанням нечіткого логічного виводу.

Методи дослідження. Для вирішення задач у магістерській роботі застосовувалися методи: логічного виводу, системного аналізу, пізнання та пошуку, теорії нечітких множин, інформаційного та функціонального моделювання, об'єктно-орієнтованого програмування, математичної статистики та логіки, семантичного аналізу тексту.

Положення, що виносяться на захист:

1. Моделі даних та потоку повідомлень тематичних ресурсів, дозволяють здійснювати статистичний та семантичний аналіз даних для прогнозування вразливостей, загроз безпеки інформації.

2. Алгоритм прогнозування вразливостей та загроз безпеки інформації на основі отриманих результатів аналізу потоку даних тематичних ресурсів, алгоритм семантичної фільтрації потоку повідомлень тематичних форумів про вразливості та загрози безпеки інформації та критерії їх оцінки.

3. Інформаційно-аналітична система автоматизації проведення аналізу потоку даних тематичних інтернет-ресурсів та прогнозування появи нових вразливостей, загроз безпеки інформації з використанням нечіткого логічного виводу

Наукова новизна:

1. Модель потоку повідомлень та бази даних, тематичного ресурсу, призначена для прогнозування вразливостей та загроз безпеки інформації, відрізняється можливістю опрацьовувати різнотипні дані, що застосовуються для організації дискусійних інформаційних тематичних ресурсів, а також можливістю статистичного аналізу та семантичної фільтрації повідомлень, дозволяє прогнозувати вразливості та загрози, враховуючи, при цьому, їхню тематичну приналежність.

2. Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів, заснований на нечіткому логічному виводу, семантичному та статистичному аналізі, відрізняється можливістю виявлення вразливостей та загроз до їх реалізації, дозволяє описувати закономірності інформаційного процесу наповнення тематичних ресурсів новими текстовими повідомленнями, що відображається на якості прогнозування.

Практична цінність. Реалізований в інформаційно-аналітичній системі алгоритм прогнозування вразливостей та загроз безпеки інформації на основі аналізу потоку даних тематичних ресурсів дозволяє автоматизувати інформаційний процес виявлення нових вразливостей, загроз, надає фахівцям інформаційної безпеки можливість оцінити своєчасно ступінь захищеності ресурсів та при необхідності вжити відповідних заходів щодо нейтралізації можливих загроз та вразливостей, тим самим підвищити інформаційну безпеку

обчислювальних комп'ютерних систем від реалізації нових мережевих комп'ютерних атак.

Обгрунтованість та достовірність результатів дослідження забезпечується коректним використанням математичного апарату, детальним аналізом стану досліджень в заданій предметній області, підтверджується, отриманими узгодженістю результатів при експериментальних дослідженнях, апробацією основних положень магістерської роботи на наукових конференціях, публікацією основних результатів дослідження, у провідних наукових виданнях.

Особистий внесок. Дослідження, проведені автором при виконанні магістерської роботи та викладені в роботі, в процесі наукової діяльності. Результати магістерської роботи, які виносяться на захист, отримані особисто автором, використаний в роботі запозичений матеріал, позначений посиланнями.

Апробація роботи. За темою дипломної роботи ОКР «Магістр» опубліковано одна фахова стаття, одна теза доповідей.

Структура і обсяг роботи. Дипломна робота ОКР «Магістр» складається зі вступу, основної частини, що містить 4 розділи, висновків і списку використаних джерел. Загальний обсяг роботи - 80 сторінок. Робота містить 24 рисунки та 10 таблиць. Список використаної літератури включає 30 бібліографічних джерела.

1 ДОСЛІДЖЕННЯ СУЧАСНИХ ВРАЗЛИВОСТЕЙ ТА ЗАГРОЗ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ТА ЗАСОБІВ ЗАХИСТУ ІНФОРМАЦІЇ

1.1 Класифікація вразливостей та загроз інформаційної безпеки, характерних для інтернет-ресурсів

Ключовими елементами забезпечення на відповідному рівні захисту інформації є аналіз, класифікація, визначення вразливостей та загроз безпеки даних. В основі формування вимог та проведення аналізу ризиків та вразливостей до систем інформаційного захисту відносять: перелік існуючих вразливостей та загроз, модель нелегітимного кореспондента, оцінка ймовірностей реалізації проведення загроз [4, 6].

Більшість існуючих моделей безпеки інформації, на сучасному етапі, ґрунтуються на забезпеченні конфіденційності, доступності, цілісності задіяної інформації [9]. Вразливості мережних інформаційних систем, як правило, є наслідком внесених в систему помилок. Помилки, що є причиною формування вразливостей, поділяються в свою чергу, на помилки реалізації та помилки адміністрування.

До помилок реалізації інформаційних систем слід віднести: помилки синхронізації, даний вид помилок, зумовлений використанням проміжних часових вікон між операціями обробки потоку даних; помилки перевірки умов виконання, нездатність інформаційної системи обробити виняткові ситуації, внаслідок визначення некоректної умови обробки потоку даних; помилки перевірки вхідних даних системи, подібні помилки призводять програмне забезпечення до вразливостей переповнення буфера.

До помилок адміністрування слід віднести: помилки оточення, прикладами цього роду помилок є помилки командного інтерпретатора; помилки конфігурування; помилки пов'язані з некоректною обробкою змінних зовнішнього середовища.

Виявлення зазначених помилок є безперервним процесом, здійснюється на всіх етапах життєвого циклу інформаційної системи: проектування, розробки, тестування, експлуатації програмного забезпечення.

До основних типів загроз безпеки інформації та інформаційних систем слід віднести [10]: стихійні лиха та аварії (пожежі, повені, урагани, землетруси); збої та відмови в роботі технічних складових та програмно-апаратного обладнання мережних інформаційних систем; наслідки помилок при розробці та проектуванні програмно-апаратних складових інформаційних систем (структур даних, технології обробки інформації, програмного забезпечення, апаратних засобів); цілеспрямовані дії порушників та зловмисників (нелегітимних кореспондентів); помилки експлуатації (користувачів, операторів, іншого персоналу).

В результаті аналізу проведеного дослідження потоку даних тематичних ресурсів зроблено висновок - опис більшості вразливостей та загроз безпеки інформації можливо отримати з потоку повідомлень тематичних учасників інтернет-ресурсів (хакерських форумів). Виняток, при цьому, становлять складні та рідкісні, в реалізації вразливості та загрози, що вимагають, в свою чергу, спеціалізованого устаткування та експертних знань.

Класифікація вразливостей та загроз за видом інформації, що захищається представлена на рис. 1.1.

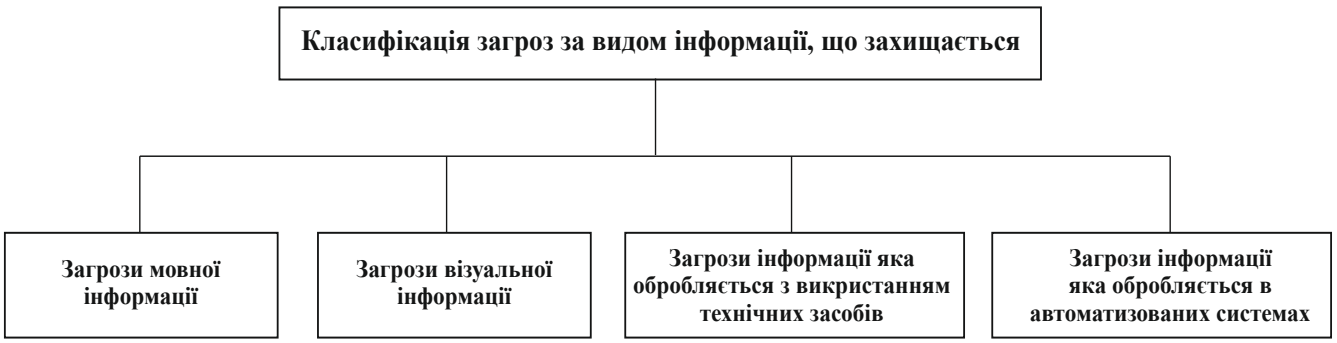


Рисунок 1.1 - Класифікація загроз за видом інформації, що захищається

На рис. 1.2 наведена класифікація вразливостей та загроз інформаційної безпеки за способами реалізації.

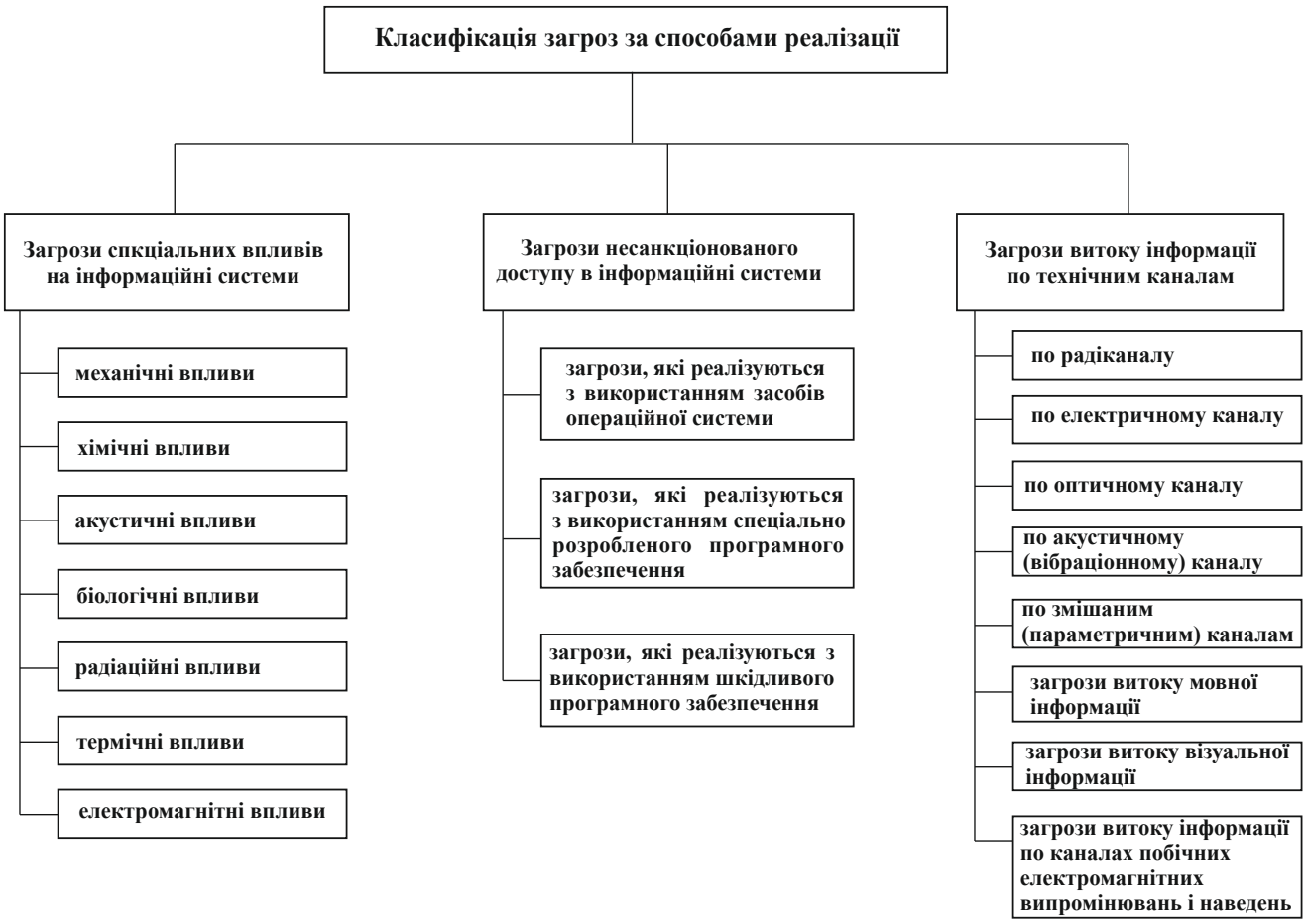


Рисунок 1.2 - Класифікація загроз інформаційної безпеки за способами реалізації

На рис. 1.3 наведена класифікація вразливостей та загроз інформаційної безпеки по виду джерел.

Рисунок 1.3 - Класифікація загроз інформаційної безпеки по виду джерел

1.2 Дослідження та класифікація нелегітимних кореспондентів інформаційної безпеки

Важливою складовою проведення успішного аналізу ризиків та визначення вимог до складу та параметрів систем безпеки інформації державних інформаційно-пошукових систем, є розробка гіпотетичної моделі потенційного нелегітимного кореспондента [11]. Класифікація порушників безпеки інформаційних систем наведена на рис. 1.4.

Рисунок 1.4 - Класифікація нелегітимних кореспондентів інформаційної безпеки

Визначення конкретних значень характеристик можливостей нелегітимних кореспондентів пов'язана, як правило, зі значним ступенем суб'єктивізму. Моделі нелегітимних кореспондентів допускається представлення, побудованих з

врахуванням механізмів обробки інформації та особливостей характерних відповідній предметній області, у вигляді переліку декількох можливих варіантів її представлення. Кожен вид, при цьому, потенційного нелегітимного кореспондента описується наведеними вище відповідними значеннями характеристик.

Аналіз проведеного дослідження потоку текстових повідомлень тематичних інтернет-ресурсів показав, що створювані учасниками дані містять корисні відомості про засоби та можливості реалізації вразливостей та загроз, якими на сьогодні володіють нелегітимні коресподенти безпеки інформації. Таким чином, їх обробка дозволить створити, із застосуванням сучасних аналітичних методів, сучасні ефективні засоби протидії атакам та знизити рівень вразливостей та загроз інформаційної небезпеки.

Аналіз існуючих, на теперішній час, вразливостей і загроз безпеки інформації, свідчить про те, що задачі захисту інформації забезпечення максимального рівня захищеності та досягнення відповідних цілей вимагає комплексного підходу та застосування доступних методів та засобів захисту конфіденційних даних. З цієї причини одним із ключових принципів розробки концепцій, що лежать в основі захисту конкретних засобів забезпечення безпеки інформації є комплексність.

Для досягнення поставлених цілей захисту інформації на об'єктах захисту необхідно провести роботи за наступними напрямами (рис. 1.5) [9, 10]. Процес забезпечення безпеки захисту інформації має носити безперервний та комплексний характер, здійснюватися на всіх етапах функціонування та створення автоматизованих засобів обробки потоку даних. У зазначених умовах процес захисту даних має ґрунтуватися та реалізовуватися на концептуальному підході та орієнтуватися на промислове виробництво відповідних засобів захисту. Для створення відповідних механізмів захисту, забезпечення їх ефективної та надійної роботи необхідно залучити фахівці високої кваліфікації у галузі інформаційної безпеки [3-5]. Основною метою забезпечення захисту інформації є усунення

виявлення, нейтралізація інтернет - джерел негативних впливів на конфіденційні дані (інформацію), а також умов та їх причин. Перераховані джерела створюють загрози безпеці конфіденційним даним.

Рисунок 1.5 – Цілі захисту інформації

Найбільш повно сутність забезпечення захисту інформації відображають її методи та цілі. До найбільш поширених основних методів захисту конфіденційних даних відносяться [9-11]: виявлення загроз, в результаті визначення конкретних несанкціонованих дій нелегітимних кореспондентів та реальних загроз інформаційній безпеці; попередження, шляхом застосування запобіжних заходів, відомих погроз, для забезпечення безпеки інформації для нейтралізації їх виникнення; ліквідація наслідків несанкціонованих дій, загроз та відновлення обробки інформації в штатному режимі; виявлення нових загроз безпеці під час постійного контролю та аналізу виникненням можливих, реальних загроз, а також своєчасне прийняття необхідних запобіжних заходів; ліквідація загроз, за

допомогою використання локалізації умов за яких можливе заподіяння несанкціонованих дій супротивником.

Попередження несанкціонованих дій та потенційних загроз супротивником може забезпечуватися різними засобами і заходами, починаючи зі створення умов, за яких користувачі інформаційних обчислювальних систем приділяють достатню необхідну увагу питанням інформаційної безпеки, закінчуючи створенням ешелонованої, глибокої системи захисту апаратними, програмними, криптографічними та фізичними засобами.

Виявлення загроз безпеки інформації можливе шляхом проведення заходів накопичення, збору та аналітичної обробки даних про підготовку заходів щодо подолання засобів безпеки інформації, підготовку несанкціонованих дій нелегітимних кореспондентів. Виявлення загроз - визначення конкретних джерел та їх загроз, які завдають певну шкоду конфіденційній інформації. До вказаних дій можна віднести - виявлення фактів несанкціонованого доступу до систем зберігання та обробки інформації, розголошення конфіденційної інформації. Локалізація та припинення загроз безпеці інформації спрямовані на усунення конкретних дій супротивника та актуальних погроз.

Наведені засоби застосовуються для захисту конфіденційних даних від несанкціонованих дій нелегітимних кореспондентів з метою забезпечення: збереження доступності та цілісності інформації; дотримання конфіденційності інформаційних ресурсів; недопущення до конфіденційної інформації несанкціонованого доступу; забезпечення авторських прав; запобігання розголошення та витоку конфіденційної інформації. Залежно від виду інформації, (комерційна таємниця, службова, державна) регламентується здійснення її захист та організація. Захист інформації може бути визначено як діяльність власника інформаційних ресурсів або уповноважених ним осіб, спрямований на: запобігання витоку та втраті інформації; забезпечення прав управління, розпорядження, володіння конфіденційною інформацією; збереження інформаційних ресурсів відповідно до вимог, встановлених нормативними та

законодавчими актами; збереження цілісності, повноти, достовірності інформації, масивів даних та програмних засобів обробки ресурсів.

Результати проведеного аналізу змісту популярних тем інтернет-ресурсів показали, що аналітична обробка повідомлень має базуватися на методах, які враховують особливу термінологічну базу, невизначеності, прихованість інформації, а також інші особливості знань у галузі безпеки інформації.

1.3 Дослідження систем прогнозування та виявлення атак, джерела даних вразливостей безпеки інформації

Дослідження з виявлення мережевих атак, ведуться вже понад чверть століття, спрямованих проти інформаційних систем та інформаційно-обчислювальних мереж. На теперішній час визначено основні ознаки мережевих атак, адаптовані та розроблені для практичного використання інструменти виявлення дій подолання систем захисту конфіденційних даних на логічному та фізичному рівнях. Представлені на ринку системи інформаційної безпеки виявлення мережевих атак (Cisco, RealSecure, Snort ISS), у складі даних систем відсутні ефективні засоби виявлення та попередження характерних ознак підготовки інтернет атак та відповідним чином реагування на них [11].

На даному етапі системи виявлення мережевих атак – це апаратно-програмні та програмні засоби, функціональні можливості даних засобів дозволяють автоматизувати процеси накопичення, контролю, збору подій програмно - апаратних систем та обчислювальних мереж. На підставі проведеного аналізу цих даних проводиться виявлення ознак порушення інформаційної безпеки. Зазначені системи сигналізують про виявлення ознак мережевих атак. На теперішній день існуючі системи виявлення комп’ютерних атак, вирішують одну, як правило, часну задачу, пов'язану із захистом конфіденційних даних від зовнішніх зловмисників, які намагаються обійти засоби захисту інформаційних ресурсів,

шляхом зовнішнього впливу на корпоративні локальні мережі, при цьому залишають невирішеною задачею захисту даних від внутрішніх загроз. Методи для виявлення атак, що застосовуються в сучасних системах захисту, у частині побудови формальних моделей атаки мають суттєві недоліки, оскільки мережеві атаки проводяться на різних рівнях інформаційно - обчислювальної системи та мають різну природу походження. Мережеві атаки характеризуються не лише якісними, а також кількісними ознаками. Атаки типу SQL-ін'єкцій, міжсайтовий скриптинг не мають сигнатур, для ефективно їх виявлення, але пов'язані з однотипними послідовностями дій порушника, які при досягненні певної кількості дій, трактують дії порушника, як атаку. Оскільки порушник має можливості у своєму розпорядженні проведення шкідливих дій на різні рівні інформаційно – обчислювальних системи, тому модель атаки має враховувати якісні та кількісні ознаки, що характеризують відповідні можливі атаки, а також охоплювати всі можливі взаємозв'язки подій на різних рівнях інформаційної системи.

Таким чином, кількість різних типів та способів отримання несанкціонованого доступу до конфіденційних даних інформаційних систем, істотно збільшилося, зросла значимість систем протидії та виявлення мережевих атак. Необхідність інтеграції систем протидії атакам до систем безпеки корпорації пов'язана з доступністю, на теперішній час, великої кількості різноманітних інформаційних ресурсів, спеціалізованої літератури що містять опис новітніх методів та підходів виявлення атак на конфіденційні дані інформаційних систем. Класифікація систем протидії виявлення атак наведено на рис. 1.6.

Рисунок 1.6 – Класифікація систем протидії виявлення комп'ютерних атак

Процедура опису впливів та атак, представляє опис набору параметрів, які підлягають подальшому аналізу та контролю. Більшість сучасних систем безпеки інформації реалізують пошук аномалій у властивостях функціонування комп'ютерних мереж, методи сигнатурного аналізу. Вже сформовано методологічний базис, в області безпеки конфіденційних даних, для проведення самостійних досліджень, про що говорять проведені дослідження робіт провідних закордонних та вітчизняних вчених. Основна увага дослідників приділялась питанням моделювання та формального опису систем протидії розмежування доступу, захист даних інформаційних систем від несанкціонованого доступу, антивірусний захист, створення операційних систем з вбудованими елементами захисту.

Водночас залишається широкий спектр практичних завдань невирішеним стосовно забезпечення безпеки даних інформаційних систем. Необхідність проведення моніторингу спеціалізованих досліджень, джерел даних про можливі загрози інформаційній безпеці та вразливості систем формулюється без зазначення конкретних способів реалізації.

Найбільш використовувані методи аналізу вразливостей та загроз безпеки інформації наступні: факторний аналіз - виявленням факторів, які ведуть до реалізації вразливостей, загроз з певною ймовірністю (зростання вірусної активності, наявність необхідних порушникам засобів реалізації загроз, вразливості інформаційно - обчислювальної системи); статистичний аналіз - базується на проведенні аналізу накопичуваних відомостей про інциденти безпеки даних інформаційної системи (частота виникнення загроз, причини, джерела); пряма експертна оцінка, визначають список, характеризуючих вразливостей та загроз, параметрів коефіцієнти їх важливості, заснована на знаннях експертів в області безпеки інформації.

Класифікація методів прогнозування вразливостей та загроз безпеки даних інформаційної системи наведено на рис. 1.7. До переваг прогнозування вразливостей та загроз інформаційної безпеки статистичними методами можливо

віднести універсальність зв'язку з відсутністю необхідності у знаннях про можливі вразливості та атаки, можливість адаптації статистичних та математичних апаратів до об'єктів, що використовуються.

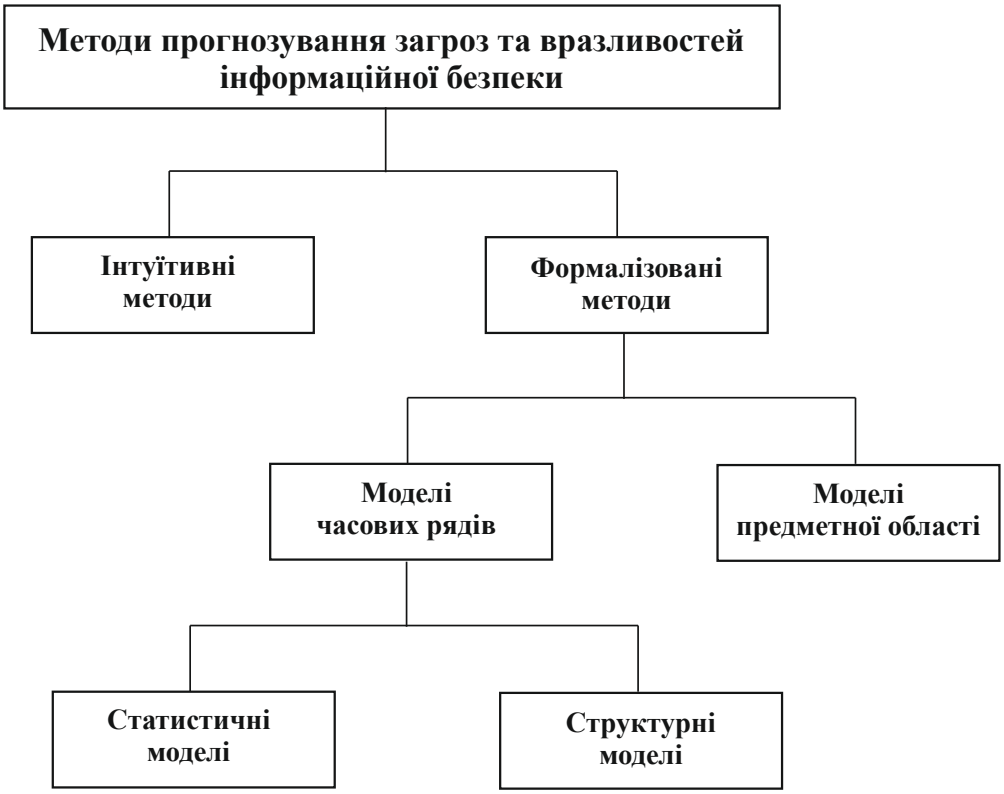


Рисунок 1.7 – Класифікація методів прогнозування вразливостей та загроз безпеки даних інформаційної системи

За відсутності статистичних даних доцільно застосовувати метод експертних оцінок, експертами даються відповіді про майбутній та поточний стан інформаційних об'єктів системи, що характеризуються деякими властивостями та параметрами. До методів прогнозування, в основі яких лежать експертні оцінки, відносять: прогнозування експертними групами - передбачає проведення дискусій, спрямованих на знаходження рішення проблеми єдиного правильного. Метод відрізняється наявністю можливості експертів представляти власні ідеї, а також критикувати чужі рішення. Перевагами методу - простота реалізації, менша ймовірність виникнення помилок; метод Дельфі - ґрунтується на «мозковому штурмі» групи висококваліфікованих спеціалістів. Суть методу полягає в організації математичної обробки та збору експертних оцінок, метод використовується при

необхідності прийняття швидкого рішення задачі; метод колективної експертної оцінки - передбачає узагальнення отриманих результатів роботи експертів в області безпеки інформаційної системи. Для досягнення різнобічного аналізу досліджуваної задачі, об'єктивного результату незалежні оцінки експертів обробляються індивідуальні.

Методи прогнозування вразливостей та загроз безпеки інформаційної системи застосовуються, як правило, в автоматизованих системах для проведення аналізу найбільш ймовірних та можливих загроз (умов, сукупності факторів, суб'єктів, які створюють небезпеку інформаційної системи для нормального її існування та функціонування), оцінки масштабу та ступеня небезпеки; розроблення концепцій забезпечення безпеки конфіденційних даних, конкретної системи - визначають заходи щодо зниження, припинення небажаних наслідків рівня безпеки системи; прогнозування розвитку найімовірніших шляхів позаштатної ситуації.

Таким чином, виходячи з проведеного аналізу даних найпопулярніших тем інтернет-ресурсів, зроблено висновок про те, що при використанні даних для вирішення задачі прогнозування вразливостей та загроз безпеки інформаційної системи доцільно використовувати статистичні методи – враховують наявні невизначеності у потоці повідомлень, термінологічну базу, особливості викладу науковцями знань в області безпеки інформації.

На теперішній час відома значна кількість загроз та вразливостей безпеки інформації. Проводиться структуризація та накопичення експертними організаціями доступної інформації про загрози та вразливості для потреб аудиторів безпеки інформації, розробників сучасних інформаційних технологій, залучених до процесу підтримки та створення інформаційних технологій. Структуризація загроз та вразливостей призвела до появи декілька баз даних, а також систем оцінки небезпеки атак. Існуючі бази даних загроз та вразливостей безпеки інформації наведені в табл. 1.1[9]. Інформація про загрози та вразливості, може істотно різнитися в різних базах даних. Саме тому постає необхідність

аналізувати всі наявні бази даних атак, щоб отримати вірогідну та якнайповнішу інформацію про загрози та вразливості.

Таблиця 1.1 – Бази даних загроз та вразливостей безпеки інформації

№ п/п	Назва	Пояснення
1	Common Vulnerabilities and Exposures (загальні впливи та вразливості)	Створена організацією MITRE у 2001 р.
2	National Vulnerabilities Database (Національна база даних загроз та вразливостей США)	Розроблена організацією NIST (National Institute of Technology and Standards, Національний інститут технологій та стандартів США)
3	BugTraq	База даних підтримується відкритою спільнотою
4	Secunia	Комерційна база даних загроз та вразливостей
5	Open Source Vulnerabilities Database (Відкрита база даних загроз та вразливостей)	Відкрита база даних загроз та вразливостей, підтримується групою активістів, створена в 2003 р.

На відміну від інтернет-ресурсів тематичних, де події безпеки інформації стають темами через нетривалий час для обговорення користувачами, даним в базах даних характерні часові затримки - опублікування описів вразливостей та загроз до декількох діб з моменту їх виявлення групою експертами. Таким чином, у зв'язку із зазначеними обставинами, оперативний аналіз повідомлень тематичних інтернет-ресурсів може сприяти, в результаті, скороченню термінів отримання необхідних відомостей про виникаючі загрози та вразливості захисту конфіденційних даних інформаційно – обчислювальних систем.

1.4 Постановка задачі

Проведено класифікацію та аналіз можливих вразливостей та загроз безпеки конфіденційних даних інформаційної системи, зроблено висновок, що опис більшості з атак міститься у тематичних інтернет-ресурсів в повідомленнях користувачів. Аналіз даних інтернет-ресурсів показав, що створювані повідомлення користувачами містять необхідні корисні відомості про засоби та можливості реалізації вразливостей та загроз, якими володіють зловмисники безпеці інформації. Обробка даних про вразливостей з застосуванням сучасних підходів та аналітичних методів дозволить забезпечити зниження рівня небезпеки вразливостей та загроз, запропонувати ефективні засоби протидії можливим атакам.

Розглянуто сучасні засоби та методи захисту конфіденційних даних, результати проведеного аналізу інформації інтернет-ресурсів можуть підвищити ефективність захисту даних, оскільки інтернет-ресурси, містять відомості про вразливості та загрози несанкціонованого доступу, спотворення, знищення конфіденційних даних, можливих технічних каналах витоку інформації, а також про шкідливе програмне забезпечення, актуальні вразливості та загрози інформаційної безпеки, засоби комп'ютерної розвідки.

Проведено аналіз принципів роботи систем прогнозування та виявлення мережевих атак та навадені їх недоліки. Для баз даних експертних організацій характерні часові затримки обробки та опублікування описів вразливостей та загроз до кількох діб з моменту виявлення їх групами експертів.

Результати аналізу даних інтернет-ресурсів можливо застосовувати в якості ефективного інструменту для реалізації протидії підготовки мережевих атак на інформаційно - обчислювальні системи.

Проведено аналіз дослідження методів визначення вразливостей та загроз інформаційній безпеці та моделей вразливостей, загроз безпеки інформації. Вказано на необхідність моніторингу джерел даних, спеціалізованих досліджень

про вразливості та загрози інформаційній безпеці, в даній ситуації не визначено конкретного способу реалізації.

Зроблено висновок про актуальність. Задача підвищення ефективності методів виявлення нових вразливостей та загроз конфіденційним даним інформаційних систем на основі розробки комплексів програм та алгоритмів є актуальною, дозволить здійснювати аналіз та виявлення інформаційних джерел, які містять інформацію про вразливості, шкідливе програмне забезпечення, комп'ютерні атаки. Обґрунтована можливість проведення аналізу тематичних інтернет-ресурсів як джерела виявлення вразливостей та загроз інформаційній безпеці.

2 МОДЕЛІ ТА АЛГОРИТМИ ПРОГНОЗУВАННЯ ВРАЗЛИВОСТЕЙ ТА ЗАГРОЗ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ОСНОВІ ІНТЕРНЕТ-РЕСУРСІВ

2.1 Прогнозування вразливостей та загроз інформаційної безпеки, особливості функціонування інтернет-ресурсів

На теперішній час в Інтернет мережі функціонує велика кількість інтернет-майданчиків, форумів (спеціалізованих інформаційних ресурсів), які використовуються учасниками мережі для обговорення питань механізмів та способів несанкціонованого доступу до конфіденційних даних, безпеки інформації Частина зареєстрованих користувачів цікавляться відомостями про захист та безпеку інформації, інші – способами здійснення атак на інформаційно - обчислювальні системи в мережі. Таким чином, форуми можуть розглядатися як джерела інформації про вразливості, шкідливе програмне забезпечення, комп'ютерні атаки.

На інтернет-ресурсах переважна більшість тем, які обговорюються присвячені висвітленню наступних питань: програмування, з метою реалізації вразливостей та загроз безпеки інформації; програмне забезпечення, що використовується для організації та проведення комп'ютерних атак; шахрайство з використанням сучасних інформаційних технологій; поширення та створення шкідливого програмного забезпечення; забезпечення сеансу анонімності при здійсненні протиправних дій із застосуванням сучасних інформаційних технологій; переведення в готівку викрадених коштів, протиправні операції з банківськими картками; захист інформації.

Перераховані теми, на даний час, відповідають актуальним загрозам безпеки конфіденційним даним [6, 8, 12], що надає можливість розглядати тематичні інтернет-ресурси як джерела повідомлень для проведення аналізу та виявлення вразливостей і загроз. Події, що відбуваються у конкретній предметній області, знаходять свій відбиток на відповідних дискусійних інтернет-майданчиках. Серед

тематичних учасників інтернет-ресурсів присутні учасники, які володіють відомостями про вразливості та загрози безпеці інформації, також, потенційні зловмисники (зацікавлені в подоланні механізмів та засобів захисту конфіденційних даних). Користувачі Інтернет мережі обмінюються наявними знаннями по актуальним темам з використанням форумів.

Зазначені фактори надають можливість прогнозувати вразливості та загрози інформаційної безпеки даних, ґрунтуючись на проведеному аналізі повідомлень тематичних інтернет-ресурсів, використовуючи, при цьому, закономірності, характерні для процесу обговорення вразливостей та загроз. В загальному вигляді процес аналізу тематичних інтернет-ресурсів та їх інформаційного наповнення наведено на рис. 2.1.

Для повідомлень інтернет-форуму, доступна інформація, про автора, час його створення, приналежність до відповідного форуму та до теми форуму, кількості повідомлень по темі форуму, рейтинг автора. Наведена структура повідомлень дозволяє проводити статистичний та семантичний аналіз інформації форуму. В результаті проведення семантичного аналізу інформації форумів можливо здійснення фільтрації тих даних, які не мають відношення до заданої предметної області вразливостей та загроз безпеці інформації.

Таким чином, на наступному кроці проведення аналізу виключаються дані, що не містять інформації про інформаційну безпеки, відповідно, досліджується інформація, що відноситься до вразливостей і загроз.

На теперішній час, для ефективного опису предметної галузі застосовується онтологія. При використанні даного підходу для опису предметної галузі, предметна область підлягає опису у вигляді сукупності понять, враховуючи організацію та існуючі властивості, зв'язки між ними. Онтологічні механізми та методи дозволяють обчислювати відстань (близкість) повідомлень форумів до термінів предметної області, заданої онтологією. Повідомлення, що мають нульове значення коефіцієнта близькості (відстані) до термінів онтології, не мають відношення до предметної області, що аналізується [12].

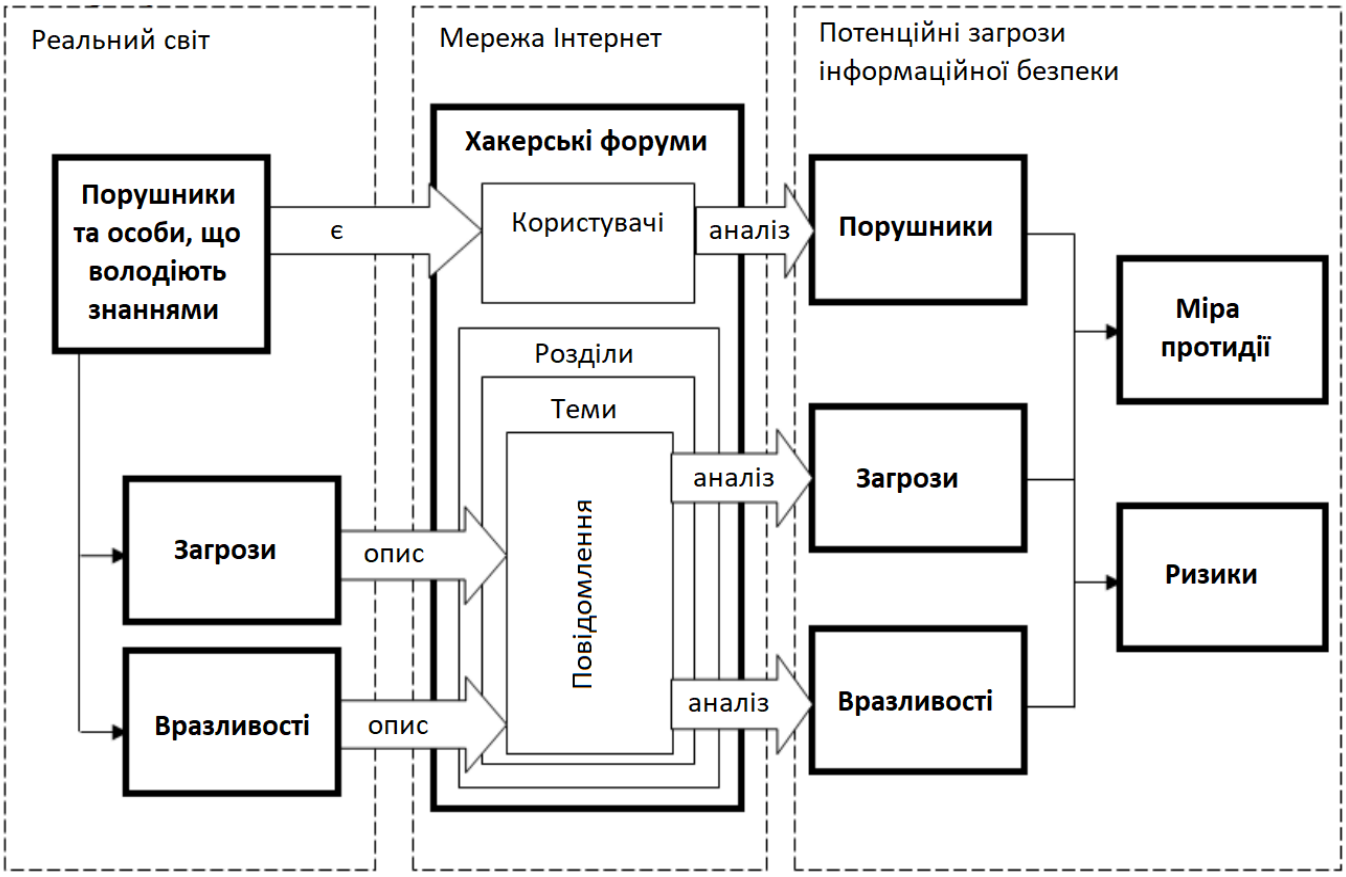


Рисунок 2.1 –Інформаційне наповнення тематичних інтернет-ресурсів

Для відповідного функціонування тематичних інтернет-ресурсів характерна закономірність, яка полягає в наступному: з появою вразливості чи загрози безпеці інформації, користувач форуму, якому відомо про загрозу, створює, відповідним чином, нову тему на інтернет форумі та залишає інформацію. Інші користувачі інтернет форуму залишають у новій створеній темі повідомлення, в яких спростовують чи доповнюють попередні повідомлення. Таким чином, в залежності від важливості інформації, що обговорюється на тематичному інтернет-ресурсі, проводиться оцінка внутрішнього рейтингу користувачів повідомлень. При високій значущості теми інтернет форуму, відповідний високий рейтинг користувачів повідомлень, також закономірно відповідно збільшення частоти появи інформації у темі інтернет форуму, де обговорюється важливі повідомлення, особливо на початковій стадія проведення дискусії.

Таким чином, вказані закономірності можуть бути задані та описані у вигляді відповідних правил нечітких продукцій, що застосовуються в інформаційних системах логічного нечіткого виводу.

Для прогнозування вразливостей та загроз безпеці конфіденційних даних можуть використовуватись результати проведеного аналізу повідомлень тем інтернет-ресурсів. Для вирішення даної задачі необхідно провести статистичний аналіз потоку даних інтернет - форуму та застосувати, при цьому, системи логічного нечіткого виводу. Учасники тематичних інтернет - форумів можуть створювати повідомлення, які не відносяться до предметної області, що аналізується, доцільно в даній ситуації застосовувати, для виключення їх з числа аналізованих, методи семантическього аналізу. Таким чином, вхідними даними в системі нечіткого виводу можуть бути використанні статистичні параметри, що характеризують інформаційний процес обговорення вразливостей та загроз інформаційної безпеки. Нечіткі правила системи нечіткого виводу описують закономірності зміни потоку інформації інтернет-ресурсів, правила розміщенні в базі нечітких продукцій. Обґрунтованість використання нечітких моделей в системі протидії, пов'язана зі значним ступенем присутньої невизначеності в інформації що підлягає аналізу, складності предметної області та неповноти інформації інтернет – форумів [12, 14].

Таким чином грунтуючись на результатах прогнозування, отриманих при виникненні раніше невідомих вразливостей та загроз інформаційної безпеки, спеціаліст, який здійснює захист інформації підприємства, може оцінити ступінь небезпеки атак та вжити необхідних заходів щодо усунення можливих загроз та вразливостей, переглянути в даній ситуації, моделі загроз інформаційної безпеки системи протидії.

2.2 Алгоритм прогнозування вразливостей та загроз інформаційної безпеки на основі даних інтернет-ресурсів

Розглянуті особливості функціонування форумів тематичних інтернет-ресурсів дозволяють здійснювати системою протидії прогнозування виникнення вразливостей та загроз безпеки конфіденційних даних. Для вирішення даної задачі необхідно проведення аналізу інтернет повідомлень, створюваних учасниками форумів тематичних інтернет-ресурсів, відповідно до представленого алгоритму на рис. 2.2. Вхідними параметрами запропонованого алгоритму (рис. 2.2) є: список форумів тематичних інтернет-ресурсів; онтологія вразливостей та загроз безпеки інформації; система логічного нечіткого виводу.

Запропонований алгоритм (рис.2.2) передбачає виконання наступних кроків:

1. Пошук нових форумів тематичних інтернет-ресурсів та додавання виявлених до наявного списку форумів.

2. Пошук нових термінів предметної області у наявних тематичних інтернет-ресурсів вразливостей та загроз безпеки конфіденційних даних, додавання нововиявлених термінів в онтологію.

3. Збір потоку повідомлень тематичних інтернет-ресурсів.

4. Проведення семантичної фільтрація потоку повідомлень тематичних інтернет- ресурсів із використанням запропонованих онтологічних методів.

5. Додавання інформації, що пройшла етап семантичної фільтрації інтернет-ресурсів в базу даних прецедентів.

6. Статистичний аналіз потоку повідомлень, що зберігаються в базі даних прецендентів.

7. Логічний нечіткий вивід про виникнення вразливостей та загроз безпеки конфіденційних даних.

8. Підготовка відповідного звіту про виявлену вразливість чи загрозу безпеки інформації.

Результатом роботи запропонованого алгоритму прогнозування вразливостей та загроз інформаційної безпеки є звіти про виявлених вразливостей, загроз інформаційної безпеки конфіденційних даних, до їх складу можуть також включатися відомості, що відображають отриманні результати аналізу текстових повідомлень, на підставі яких отримано висновок про виникнення вразливостей та загроз. Такими відомостями, в даному випадку можуть бути: частота створення участниками на форумах тематичних інтернет-ресурсів повідомлень, що відносяться до предметної області вразливостей та загроз інформаційної безпеки даних, в період проведення аналізу; частотна характеристика термінів вразливостей та загроз інформаційної безпеки даних, присутніх у повідомленнях інтернет-ресурсів в період проведення аналізу; середній рейтинг користувачів повідомлень, що відносяться до предметної галузі вразливостей та загроз інформаційної безпеки даних, в період проведення аналізу; список присутніх у повідомленнях користувачів термінів онтології вразливостей та загроз інформаційної безпеки даних, що дозволяє класифікувати прогнозовані вразливості та загрози; добірка текстів тематичних інтернет-ресурсів, містять терміни вразливостей та загроз інформаційної безпеки даних, створені на форумах тематичних інтернет-ресурсів в період проведення аналізу.

Запропонований алгоритм прогнозування вразливостей та загроз безпеки інформації відрізняється можливістю на ранніх етапах виявлення вразливостей та загроз, їх практичної реалізації, ґрунтується на проведення аналізу потоку повідомлень форумів тематичних інтернет-ресурсів, що в даній ситуації дозволяє спеціалістам з інформаційної безпеки приймати адекватні та своєчасні заходи щодо захисту конфіденційних даних організації.

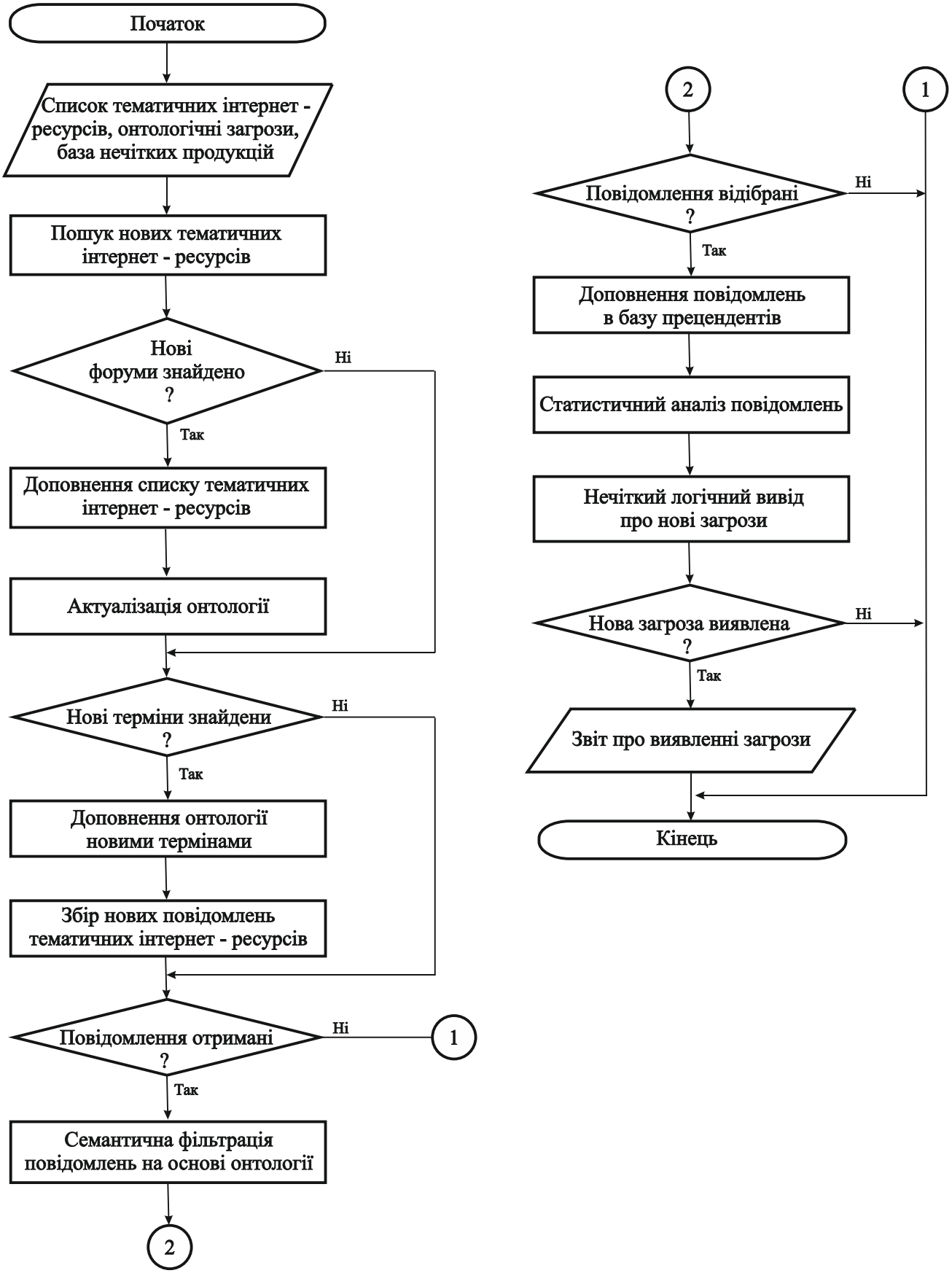


Рисунок 2.2 – Алгоритм прогнозування
вразливостей та загроз інформаційної безпеки

2.3 Модель потоку текстових повідомлень та бази даних інтернет-форуму

При створенні форумів інтернет-ресурсів найбільшою популярністю користуються наступні програмні платформи: Vanilla; Invision Power Board (IPB); vBulletin; PunBB; Simple Machines Forum (SMF); XenForo; phpBB.

При реалізації наведених програмних платформ, використовуються структури бази даних, які значно різняться. У той же час, у кожній базі даних, записи даних зберігаються відповідним чином, так що дозволяють для текстових повідомлень визначати їхню приналежність до конкретного інтернет - форуму, рейтингу автора, автора, темі форуму, часу створення, а також кількості повідомлень відповідної теми форуму. У зв'язку з цим, враховуючи наведену інформацію, сформовано модель інтернет-форумів (структуру бази даних) (рис. 2.3). Запропонована модель бази даних тематичного інтернет-ресурсу, відрізняється від існуючих, універсальністю, що надає можливість аналізувати та досліджувати повідомлення інтернет-форумів, реалізованих на базі найбільш популярних сучасних програмних платформ для створення відповідних дискусійних тематичних інформаційних форумів. Застосовуючи на практиці запропоновану модель бази даних, при прогнозуванні вразливостей та загроз інформаційної безпеки конфіденційних даних, надасть можливість аналізувати основну більшість існуючих тематичних інтернет-форумів, незалежно від задіяної конкретної програмної платформи, використовуваної для її реалізації.

Кожне повідомлення інтернет мережі є окремо структурою, що складається з пов'язаних між собою елементів (рис. 2.4).

Потоком текстових повідомлень інтернет мережі є множина текстових повідомлень тематичних інтернет-форумів, створюваних учасниками форуму. Враховуючи те, що моделювання мережевого потоку текстових повідомлень тематичних інтернет-форумів здійснюється з метою наступного прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних, при побудові моделі тематичного інтернет- форуму, необхідно, в даній ситуації

передбачити можливість здійснення семантичного та статистичного аналізу текстових повідомлень, враховуючи належність даних до конкретного форуму, кількість повідомлень теми форуму, теми форуму, автора, часу створення, рейтингу автора.

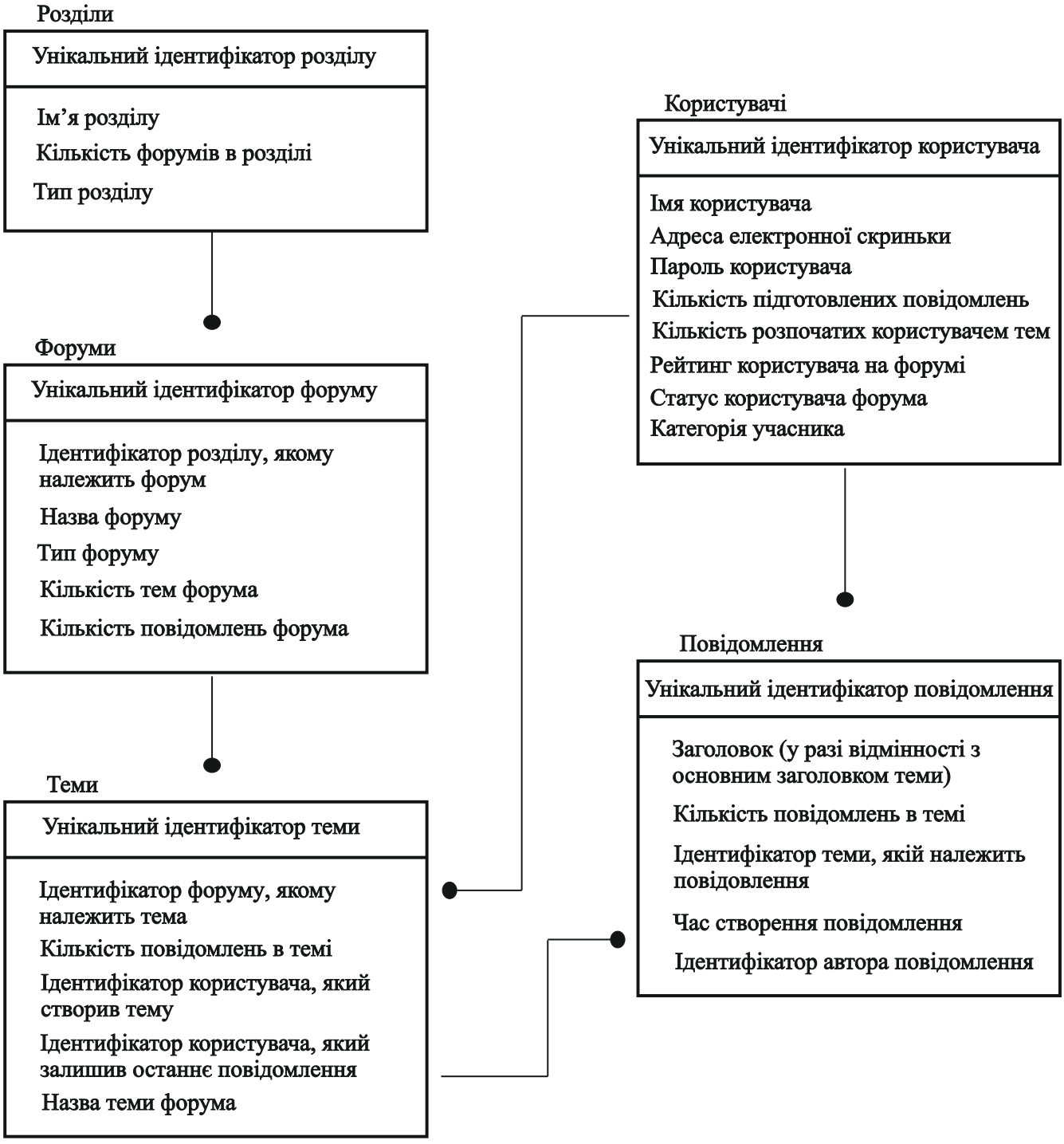


Рисунок 2.3 – Логічна модель бази даних тематичних інтернет-ресурсів

Рисунок 2.4 – Структура повідомлення тематичного інтернет-форуму

Онтологія, на сьогоднішній день є найбільш ефективним інструментом для опису конкретної предметної області. Суть онтологічного підходу полягає у представленні предметної області у вигляді організованої множини понять, враховуючи, також, існуючі зв'язки між ними та їх властивості [23]. Модель інтернет потоку текстових повідомлень, у загальному вигляді, яка має відношення до предметної області, заданої онтологією, представляється трійкою (2.1):

$$S_\tau = (M, O, T), \tag{2.1}$$

де S_τ – потік текстових повідомлень інтернет мережі у поточний час τ; M - множина інтернет повідомлень у потоку даних; O - онтологія даної предметної області; $T = \{1, ..., \tau\}$ - множина часових інтервалів, в які велися спостереження за інтернет потоком повідомлень (днів, місяців, годин).

$$O = (E, R, F), \tag{2.2}$$

де E - множина термінів заданої предметної області; R - множина взаємозв’язків між термінами даної предметної області; F - множина заданих на відношеннях та термінах онтології функцій аксіоматизації (інтерпретації).

Кожне текстове повідомлення $d \in M$ може бути представлене наступним чином:

$$d \in (s, t, F_d, A), \tag{2.3}$$

де s - текст інтернет повідомлення, $t=\{1,...,\tau\}$ - момент часу створення текстового повідомлення; $F_d=\{w_1,...,w_k\}$ – вектор, що представляє текстове повідомлення заданої предметної області, заданою відповідною онтологією O, k – кількість в онтології O термінів, координати $w_i(i=1,...,k)$ – ваги термінів у текстовому повідомленні, A – рейтинг автора текстового повідомлення.

При розрахунку ваги термінів використовується модель Term frequency – Inverse document frequency (TF–IDF), відповідно до якої вага терміна текстового повідомлення прямо пропорційна частоті входження терміна в інтернет повідомлення і обернено пропорційна кількості текстових повідомлень, у яких зустрічається термін (2.4):

$$w_i = F_i \cdot \log\left(\frac{D}{DF_i}\right), \tag{2.4}$$

де w_i – вага i - терміна у текстовому повідомленні; F_i – частота i - терміна у інтернет повідомленні; D – загальна кількість текстових повідомлень; DF_i – кількість текстових повідомлень, у яких зустрічається i - термін.

Розглянута модель не враховує того, що текстові повідомлення тематичних форумів можуть мати різний розмір, у зв'язку з чим, вага терміна і відповідно частота будуть зменшуватися зі зростанням розміру текстових повідомлень. Враховуючи дану ситуацію необхідно проводити нормування ваг термінів у інтернет повідомленні, діленням їх на довжину вектора-повідомлення (еклідову норму) (2.5):

$$w_i^* = \frac{w_i}{\overline{d}} = \frac{w_i}{\sqrt{\sum_{i=1}^{k} w_i}}, \tag{2.5}$$

Запропонованав модель потоку текстових повідомлень тематичних форумів, які відносяться до даної предметної області, заданої онтологією, відрізняється від існуючих, можливістю проводити статистичний аналіз та семантичну фільтрацію інтернет повідомлень, враховуючи, при цьому, приналежність до конкретного

автора, форуму, рейтингу автора, тему форуму, часу створення, а також кількості повідомлень конкретної теми форуму, дозволяє здійснювати аналіз та дослідження інтернет повідомлень тематичних інформаційних ресурсів.

2.4 Фільтрація потоку повідомлень системи прогнозування інформаційної безпеки

При використанні онтологічних методів для опису та обробки предметної області, необхідно представити її у вигляді організованої структури сукупності термінів (понять), враховуючи існуючі властивості та зв'язки між ними. У задачах, що передбачають подальшу обробку, предметної області, розроблених онтологій, найчастіше застосовується для їх представлення формат OWL (Ontology Web Language) – мова опису онтологій для семантичного павутиння.

У термінознавстві та лексикографії застосовуються алгоритми, що ґрунтуються на статистичних та лінгвістичних методах, для видалення термінів [16,17].

При використанні статистичних методів основним критерієм є ступінь термінологічності, визначається у відповідності до числових закономірностей, характерними для нетермінів і термінів. При використанні лінгвістичних методів терміни відбираються за лінгвістичними ознаками та певними граматичними лексичними шаблонами [19, 20].

При використанні онтологічного підходу знання про відповідні предметні області (онтологія) зберігаються у вигляді (2.6):

$$O=\left(E,R,F\right), \tag{2.6}$$

де E - множина термінів заданої предметної області; R - множина взаємозв’язків між термінами даної предметної області (2.7):

$$R\subset\left\{R_{inc},R_{add},R_{term},R_{lem},R_{NC}\right\}, \tag{2.7}$$

де R_{inc} – множина вбудованих відношень між об'єктами («є Підкласом»); R_{add} – множина вбудованих відношень між об'єктами, дозволяють розширювати набір відповідних об'єктів предметної області, що аналізується, шляхом об’єднання лем області, пов'язаних між собою об'єктів («має Відношення», «є Частиною»); R_{term} – відношення «є Терміном», визначається експертним шляхом і носить допоміжний характер. Приймає логічний тип значення (в залежності від того, наскільки об'єкт характерний для предметної галузі, що аналізується). Прикладне застосування R_{term} знаходить при вирішенні задач видалення, з використанням тезаурусного критерію термінологічності термінів; R_{lem} – відношення «має Лемму», дана властивість приймає значення рядкового типу, яке виходить в результаті лемування, приведенні найменування об'єкта до початкової форми; R_{NC} – множина відношень між тематичними об'єктами, описують особливості взаємодії між собою об'єктів даної предметної області (властивості «є Елементом», «є Типом вірусів»); F – множина заданих на термінах та відношеннях онтології предметної області функцій інтерпретації (аксіоматизації).

Задача фільтрації текстових повідомлень, які не належать до предметної області, для якої проводиться аналіз, може бути вирішена із використанням семантичної метрики "нетермін/термін". Для використання семантичної метрики необхідно попередньо розробити відповідну онтологію предметної області у форматі OWL. Далі для кожного тематичного повідомлення, що надходить в інтернет мережі, розраховується значення конфіцієнта ступеня близькості термінів включених до онтології, в результаті виділяються тематичні повідомлення, що виключно відносяться до розглядуваної предметної області [8 – 11].

Значення коефіцієнта ступеня близькості текстового повідомлення до всіх термінів предметної області, який розраховується при використанні семантичної метрики «не термін/термін», приймає значення в діапазоні від 0 до 1 (текстове повідомлення відноситься до певного терміну, чим ближче значення коефіцієнта k_{Ont} до 1). Для вирішення задач відбору текстових повідомлень, що належать до

до даної предметної області, для якої проводиться аналіз, заданої у вигляді онтології, використовуються два критерії: вкладених зв'язків та тезаурусний критерій. Тезаурус – словник термінів на природній мові, де явно вказуються між термінами відношення предметної області, здебільшого застосовується для вирішення задач інформаційного пошуку, онтологія - ускладнена версія тезаурусу.

Застосування тезаурусного підходу до вирішення задачі фільтрації текстових повідомлень полягає в пошуку лем, що містяться в тематичних повідомленнях, що надходять інтернет мережі, серед термінів онтології даної предметної області. Використання тезаурусного підходу для відповідного класу онтології предметної області визначається властивість «має Лемму», шляхом приведення до початкової форми (лемування) найменування об'єкта області, що аналізується.

Для розрахунку коефіцієта ступеня близькості тематичних повідомлень до термінів даної предметної області відповідно до тезаурусного критерію, необхідно виконати послідовність дій: провести оцінку коефіцієнта ступеня близькості текстового повідомлення, що надходить до кожного об'єкта онтології, для якої проводиться аналіз; визначити опорний об'єкт заданої онтології, який найбільш близько асоціюється з тематичним повідомленням, що надходить в інтернет мережі.

Розрахунок коефіцієнта ступеня близькості текстового повідомлення термінам заданої предметної області з використанням тезаурусного критерію представлено на рис. 2.5.

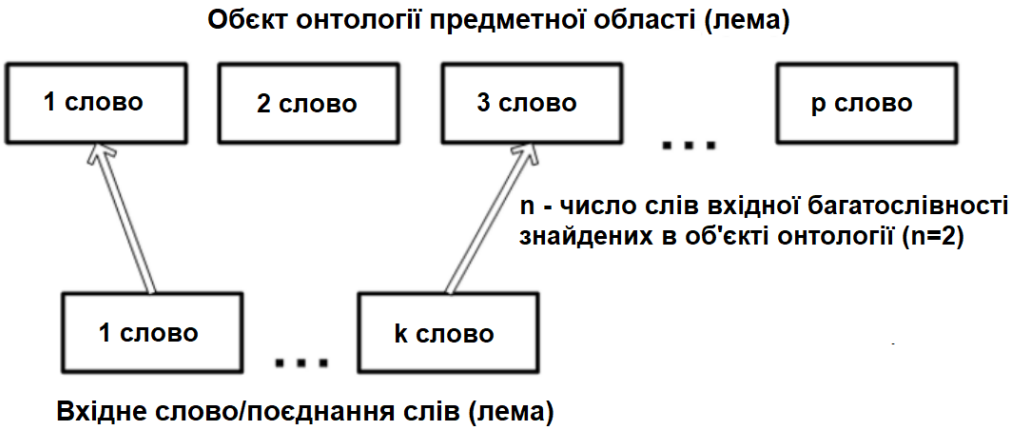


Рисунок 2.5 – Визначення опорного об'єкта онтології

Коефіцієнт ступеня близькості опорного об'єкта онтології до тематичного повідомлення, що надходить в інтернет мережі розраховується наступним чином (2.8):

$$k_t = \max\left(\frac{n_i}{p_i}\right)_{i=1}^{m}, \qquad (2.8)$$

де m – загальна кількість об'єктів заданої онтології; n_i – кількість слів у лемі тематичного повідомлення, присутніх у лемі i–го об'єкта заданої онтології предметної області; p_i – кількість слів у лемі i–го об'єкта онтології предметної області.

У випадку коли значення коефіцієнта k_t отримано однакове для декількох об'єктів онтології заданої предметної області, опорним приймається об'єкт онтології, для якого значення n_i набуває максимальної величини. Якщо, при цьому, існує декілька об'єктів онтології, для яких значення n_i і k_t однакові, то в даній ситуації всі об'єкти вважаються опорними і для кожного об'єкта проводиться аналіз по онтологічного критерію.

Відповідно до тезаурусного критерію коефіцієнт ступіня близькості тематичного повідомлення термінам даної предметної області розраховується за наступним чином (2.9):

$$k_{Ont} = \frac{k_t}{c+1}, \qquad (2.9)$$

де k_t - коефіцієнт ступеня близькості, розрахований на першому етапі проведення аналізу (обчислюється за (2.8)); c - число відношень між об'єктами, які пов'язують опорний об'єкт онтології предметної області з об'єктами, що мають значення властивості «є Терміном». Якщо опорний об'єкт онтології предметної області є терміном, то c=0. Тезаурусний критерій схематично представлений на рис. 2.6.

Рисунок 2.6 – Схема тезаурусного критерію

Застосування метрики "не термін/термін" для коефіцієнта оцінки ступеня близькості тематичних повідомлень до термінів заданої предметної області передбачається рух графом, при цьому об'єкти класів онтології предметної області є вузлами графа. Якщо у опорного об'єкта онтології предметної області властивість «є Терміном» хибно, і він при цьому не пов'язаний з іншими об'єктами, або у всіх пов'язаних з ним об'єктів онтології значення властивості «є Терміном» хибно, то в даному випадку проводиться пошук інших опорних об'єктів онтології і знову ж таки проводиться оцінка. При цьому тематичне повідомлення не відноситься до предметної області ($k_{Ont}=0$), коли опорні об'єкти онтології відсутні або для всіх опорних об'єктів предметної області характерна розглянута ситуація.

Критерій вкладених зв'язків онтології ґрунтується на тому, що крім оцінки коефіцієнта ступеня термінологічності кожного текстового повідомлення, метрика «не термін/термін» дозволяє проводити фільтрацію тематичних повідомлень шляхом зіставлення леми повідомлення та поєднаннями лем об'єктів онтології предметної області, пов'язаних відношеннями.

Таким чином, тематичне повідомлення вважається таким, що відноситься до заданої предметної області, якщо його лема співпадає з об'єднанням лем об'єктів даної онтології, пов'язаних між собою, при цьому односпрямованими відношеннями. Особливість запропонованого методу пов'язана з тим, що об'єкти онтології необхідно представляти переважно однослів'ями, що мають, в даному випадку максимальну кількість відношень із іншими об'єктами онтології. Для практичного використання методу визначальними є відношення, що дозволяють формувати природним чином словосполучення.

2.5 Висновки

Розроблено алгоритм прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних, що відрізняється від відомих, можливістю виявлення вразливостей та загроз на ранніх етапах їх практичної реалізації, і оснований на проведення аналізу потоку текстових повідомлень мережі тематичних інтернет-ресурсів, що, в свою чергу дозволяє спеціалістам з інформаційної безпеки приймати своєчасні адекватні контрзаходи щодо захисту кофіденційних даних.

Запропоновано інформаційну модель бази даних форуму тематичного інтернет-ресурсу, відрізняється від відомих, універсальністю, дозволяє аналізувати та досліджувати потік даних інтернет-форумів, реалізованих на базі популярних програмних платформ для розробки дискусійних інформаційних тематичних ресурсів.

Запропоновано модель потоку текстових повідомлень тематичних інтернет форумів, які відносяться до заданої предметної області, даної онтології, відрізняється від відомих, можливістю проводити статистичний аналіз та семантичну фільтрацію повідомлень, враховуючи належність до автора, рейтингу автора, форуму, часу створення, кількості повідомлень, темі форуму, дозволяє здійснювати аналіз та дослідження повідомлень тематичних інтернет- ресурсів.

3 МЕТОД ПРОГНОЗУВАННЯ ВРАЗЛИВОСТЕЙ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ОСНОВІ ДАНИХ ІНТЕРНЕТ-РЕСУРСІВ

3.1 Алгоритм фільтрації потоку повідомлень та статистичного аналізу інформаційної безпеки

На підставі описаних вище особливостей функціонування тематичних інтернет-ресурсів, методів семантичної фільтрації текстових повідомлень та послідовність проведення аналізу створюваних учасниками форуму повідомлень в період проведення аналізу тематичних повідомлень може бути представлена наведеним алгоритмом рис. 3.1.

Запропонований алгоритм фільтрації потоку повідомлень та статистичного аналізу передбачає фільтрацію тематичних повідомлень, що не відносяться до заданої предметної області, що задана відповідною онтологією, а також підрахунок кількості текстових повідомлень, що пройшли етап фільтрації потоку даних, та визначення середнього рейтингу авторів текстових повідомлень.

Вхідними параметрами алгоритму фільтрації потоку повідомлень та статистичного аналізу інформаційної безпеки є: D_τ – множина текстових повідомлень тематичних інтернет-ресурсів, створених в період проведення аналізу потоку даних; O – онтологія предметної області вразливостей та загроз інформаційної безпеки конфіденційних даних.

Основні кроки алгоритму проведення аналізу потоку текстових повідомлень наступні:

1. Обнулення значень K_τ – кількості тематичних повідомлень про вразливості та загрози інформаційної безпеки конфіденційних даних та A_τ – середнього рейтингу авторів тематичних повідомлень створених у період часу проведення аналізу τ;

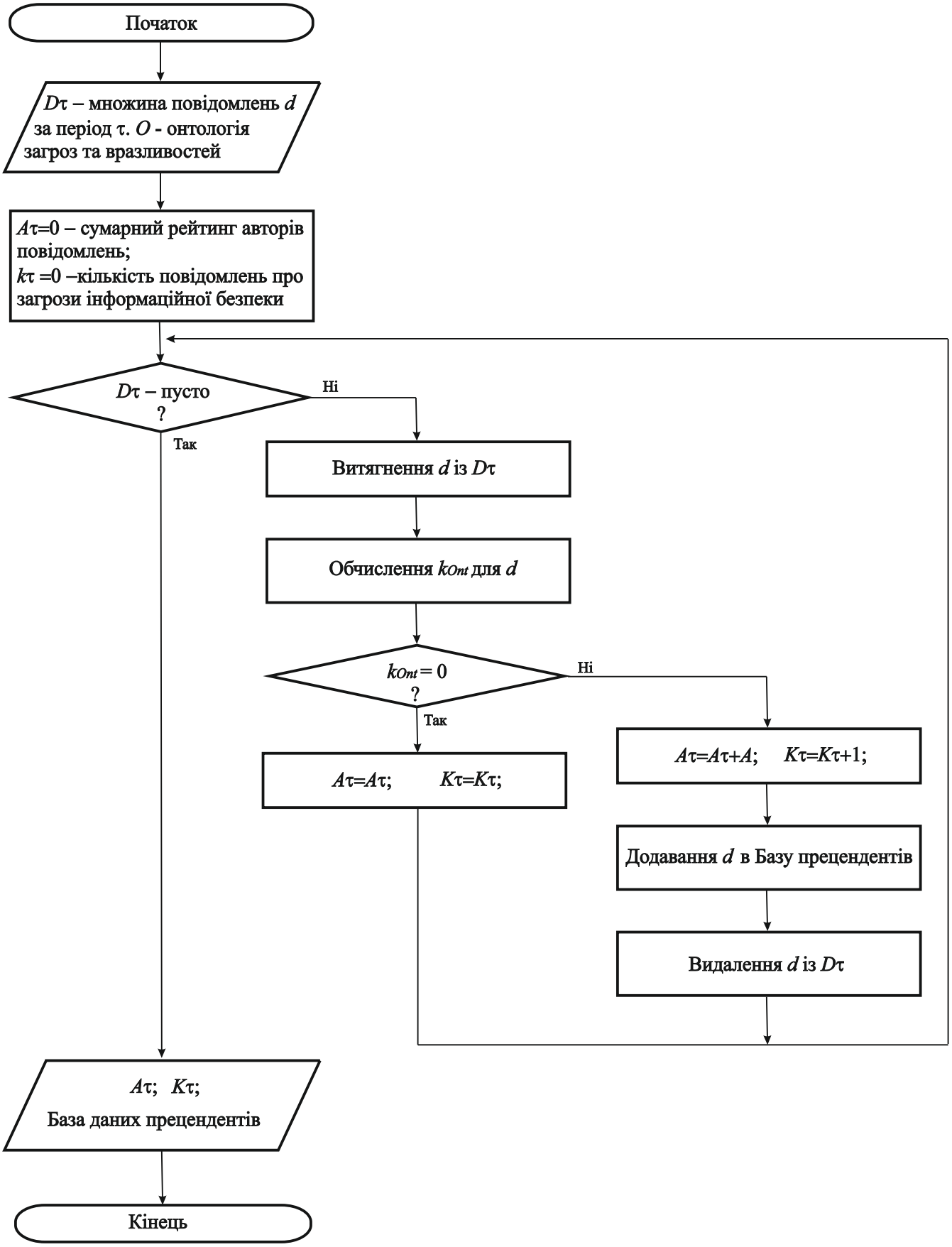


Рисунок 3.1 – Алгоритм аналізу потоку текстових повідомлень

2. Обчислення для кожного текстового повідомлення коефіцієнта k_{Ont} – близькості до термінів предметної області O заданої онтології;

3. Додавання тематичних повідомлень множини D_τ, для яких виконується нерівність $k_{Ont} > 0$, до бази даних прецедентів для їх подальшого використання для формування відповідних звітів про прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних;

4. Обчислення K_τ – кількості повідомлень множини D_τ, для яких виконується нерівність $k_{Ont} > 0$;

5. Обчислення A_τ – середнього рейтингу авторів тематичних повідомлень множини D_τ, для яких виконується нерівність $k_{Ont} > 0$.

Таким чином, результатом роботи запропонованого алгоритму аналізу потоку текстових повідомлень є визначення статистичних показників, що характеризують потік тематичних повідомлень в період проведення аналізу потоку даних: K_τ – кількість текстових повідомлень, що містять терміни вразливостей та загроз інформаційної безпеки з онтології конфіденційним даним; A_τ – середній рейтинг авторів тематичних повідомлень, що містять терміни вразливостей та загроз інформаційної безпеки з онтології конфіденційним даним; поповнення бази даних прецедентів текстовими повідомленнями, що містять терміни вразливостей та загроз інформаційної безпеки з онтології конфіденційним даним.

Отриманні результати застосування запропонованого алгоритму аналізу потоку текстових повідомлень можуть бути використані як значення вхідних параметрів у системі логічного нечіткого виводу та при формуванні звітів про прогнозування вразливостей та загроз інформаційній безпеці організації.

Розроблений алгоритм аналізу тематичних інтернет-ресурсів потоку текстових повідомлень дозволяє обчислювати статистичні параметри, здійснювати семантичну фільтрацію текстових повідомлень, а також результати алгоритму можуть бути використанні для побудови системи логічного нечіткого виводу для прогнозування подій предметної області для якої проводиться аналіз.

3.2 Метод прогнозування вразливостей та загроз інформаційної безпеки

Розглянутий вище матеріал вказує на важливість при забезпеченні захисту інформаційно - обчислювальних систем задачі щодо підтримки в актуальному стані моделі загроз інформаційної безпеки конфіденційних даних. При цьому, фахівцю, який забезпечує безпеку даних інформаційно - обчислювальних систем, необхідно своєчасно приймати адекватні рішення щодо необхідності перегляду моделі інформаційної безпеки конфіденційних даних, у разі виявлення вразливостей або виникнення загроз.

Існуючі, на теперішній час, методи підтримки прийняття рішень надають наступні можливості: надавати критеріальні оцінки параметрам та ранжувати критерії, значущими для заданої задачі (надає можливість оцінити надані варіанти рішень); формалізувати процес на основі наявних даних знаходження рішення (процес генерації варіантів розв'язку); використовувати формальні процедури прийнятих рішень прогнозування наслідків; використовувати під час прийняття колективних рішень формалізовані процедури узгодження; вибирати кращий варіант, що призводить до розв'язання поставленої задачі.

Основні задачі, які вирішуються методами підтримки прийняття рішення: вибір альтернативи; генерація варіантів рішення (альтернатив).

Критерій підтримки прийняття рішень - функція, що виражає переваги особи, яка приймає відповідне рішення, визнає правило, за яким вибирається оптимальний чи прийнятний варіант рішення задачі. Існує безліч критеріїв підтримки прийняття рішень, що використовуються в залежності від умов поставленої задачі [16, 17].

Нечіткі множини застосовуються при вирішенні поставленої задачі при необхідності описувати нечіткі знання та поняття, а також в подальшому проводити операції з цими знаннями і поняттями та формувати нечіткі виводи.

Обґрунтованість застосування нечітких моделей при вирішенні поставленої задачі пов'язана зі значним ступенем присутності невизначеності, по причині

складності предметної області та неповноти наданої інформації, а також наявністю відповідних відомостей про систему якісного характеру [16, 17].

Основною перевагою використання нечітких систем є їх універсальність, будь-яку безперервну функцію можна представити із заданою точністю нечіткою моделлю [17, 18]. Інформаційні системи, що побудовані на логічній нечіткій логіці, дозволяють синтезувати модель об'єкта предметної області, на основі евристичної інформації, а також інформації отриманої експертним шляхом або в результаті проведення експерименту. До недоліків логічних нечітких систем відносять низьку швидкість при великій кількості керуючих правил їх роботи, відсутність, на теперішній час, алгоритмів, що дозволяють здійснювати синтез стійких моделей [15, 16].

Побудова логічних нечітких систем при вирішенні певних відповідних задач, на відміну від використання класичних методів, нерідко передбачає введення суб'єктивного характеру додаткових аксіом. У зв'язку з цим, процес створення логічних нечітких моделей притаманні елементи творчості [8, 11].

Як правило, методи логічного нечіткого виводу застосовуються для вирішення задач, пов'язаних, насамперед з апроксимацією функцій, класифікацією та розпізнаванням образів, управлінням та моделюванням нелінійними об'єктами, прийняття адекватних рішень в умовах невизначеності [23, 25].

Центральне місце в системах логічного нечіткого моделювання займає нечіткий вивід, який є відповідною процедурою або алгоритмом для отримання логічних нечітких виводів, ґрунтуючись на застосуванні операцій нечіткої логіки та нечітких передумов.

У загальному вигляді структура системи логічного нечіткого виводу та послідовність реалізованих системою етапів представлена на рис. 3.2.

Продукційне правило для системи нечітко логічного виводу, відповідно до існуючих на теперішній час методик побудови бази даних правил логічної нечіткої системи, представляється наступним чином (3.1):

$$ЯКЩО(u_1 \epsilon A_1) I ... I (u_n \epsilon A_n) ТО (y \epsilon Q_j), \quad (3.1)$$

де $u_1,...,u_n$ – нечіткі змінні логічної нечіткої системи з n входами; $A_1,...,A_n$ – нечіткі множини, що відповідають нечітким змінним $u_1,...,u_n$; y – нечітка вихідна логічна змінна; Q_j – нечітка множина, що відповідає нечіткій логічній змінній y.

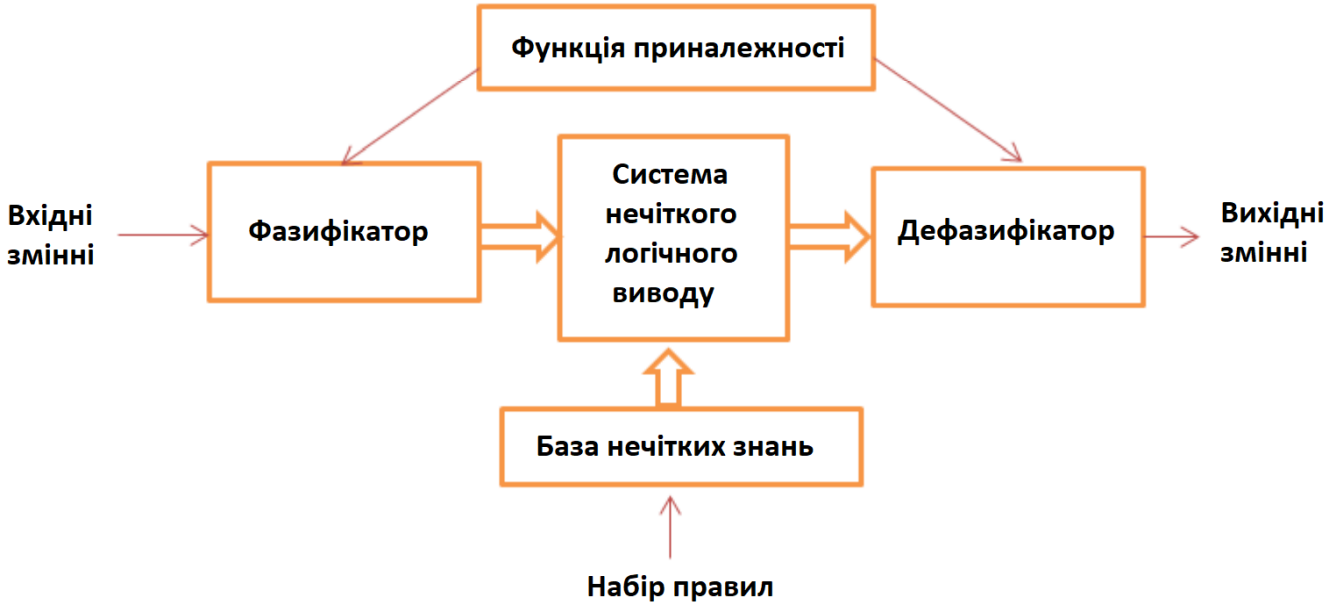


Рисунок 3.2 – Структура системи нечіткого логічного виводу

Фазифікацією вхідних змінних називається - процес перетворення чітких значень вхідних параметрів у відповідні їм нечіткі множини. В залежності від виду функцій приналежності, реалізуються процеси фазифікації наступні: гаусова, трикутна, одноелементна[10, 13]. В результаті одноелементної, наприклад, фазифікації чіткого числа u_i для i – го входу нечіткої системи створюється A_i – нечітка множина з функцією приналежності «Сінглтон» [12, 13] (3.2):

$$\mu_{A_i}(x) = \begin{cases} 1, x = u_i \\ 0, інакше \end{cases} \quad (3.2)$$

Значення коефіцієнтів ступенів приналежності підумов нечітких логічних продукцій обчислюються як результат перетину нечітких множин системи A_i, отриманих шляхом фазифікації вхідних параметрів u_i та нечітких множин системи A_i із відповідних правил бази даних нечітких продукцій. При перетині

нечітких множин системи застосовується, так звана Т-норма. Її часним випадком є операція отримання значення мінімуму (3.3):

$$A_i(u_i) = A_i(u_i) \wedge A_i(u_i), \tag{3.3}$$

де A_i - нечітка множина системи, яка визначена для i - ї підумови деякого продукційного правила (3.1); A_i – нечітка множина системи, яка отримана в результаті фазифікації чіткого значення змінної для i - го входу системи; A_i – нечітка множина логічної системи, що відповідає i – й умові деякого продукційного правила бази даних.

Процедури агрегування умов, акумулювання та активізації підзаключень правил нечітких продукцій логічної системи, а також операція дефазифікації залежить від вибору відповідного алгоритму нечіткого логічного виводу [15, 17].

На теперішній час найбільш затребувані алгоритми нечіткого логічного виводу Ларсена, Такагі-Сугено, Цукамото та Мамдані. Найбільшою популярністю при вирішенні прикладних завдач користуються алгоритми Мамдані і Такаги-Сугено [16, 17].

Аналіз проведених досліджень оцінки ефективності наведених алгоритмів нечіткого логічного виводу показав, що їх застосування залежить від специфіки задачі, яку необхідно вирішувати з їх використанням. Застосування алгоритму Мамдані дозволяє, при цьому, уникнути великих об’ємів обчислювальних операцій. З урахуванням даної особливості пов'язана його популярність при вирішенні практичних задач нечіткого логічного моделювання.

Таким чином, беручи до уваги, що алгоритм Мамдані для вирішення нечіткої задачі використовує менші обчислювальні ресурси і реалізації нечіткого виводу, то для вирішення задачі прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних обрано алгоритм Мамдані.

Метод прогнозування вразливостей та загроз інформаційної безпеки включає наступні етапи:

1. Етап формування правил логічних нечітких продукцій у вигляді :

$$ЯКЩО\left(u_1 \epsilon A_1\right) I \ldots I\left(u_n \epsilon A_n\right) TO\left(y \epsilon Q_j\right)$$

2. Етап фазифікації вхідних параметрів за формулою:

$$\mu_{A_i}(x)=\begin{cases}1,\ x=u_i\\ 0,\ інакше\end{cases}$$

3. Етап обчислення коефіцієнтів ступенів приналежності підумов відповідно до правил логічних нечітких продукцій за формулою:

$$A_i\left(u_i\right)=A_i\left(u_i\right)\wedge A_i\left(u_i\right),$$

4. Етап агрегування умов, відповідно до правил нечітких логічних продукцій. Визначення значень коефіцієнтів ступенів приналежності передумов кожного продукційного правила. При перетині нечітких логічних множин використовується метод Т-норма, часним випадком є операція мінімуму:

$$a_j=A_1\left(u_1\right)\wedge A_2\left(u_2\right)\wedge \ldots \wedge A_n\left(u_n\right),$$

де a_j - ступінь приналежності передумови для j - го правила; $A_1\left(u_1\right)\wedge A_2\left(u_2\right)\wedge \ldots \wedge A_n\left(u_n\right)$ - нечіткі логічні множини для n підумов j-го правила. При цьому використовуються і вважаються активними для подальших розрахунків ті продукційні правила, для яких значення коефіцієнтів ступенів приналежності передумов не є нулем.

5. Етап активізації нечітких виводів у правилах логічних нечітких продукцій. Здійснюється, дана операція, із застосуванням операції мінімуму. Для вихідних параметрів визначаються «усічені» функції приналежності, розглядаються лише активні правила логічних нечітких продукцій.

$$\overline{Q}_i\left(y\right)=a_j\wedge Q_i\left(y\right),$$

де a_j - значення коефіцієнта ступеня приналежності передумови j - го правила продукції, $Q_i\left(y\right)$- нечітка множина виводів j - го продукційного правила, $\overline{Q}_i\left(y\right)$ - «усічена» нечітка множина виводів j - го продукційного правила.

6. Етап акумуляції виводів правил логічних нечітких продукцій. Здійснюється об'єднанням знайдених «усічених» логічних функцій приналежності та отриманням для вихідного параметру підсумкової логічної нечіткої множини. Для об'єднання логічних нечітких множин застосовується метод S-норма, окремим випадком застосування якого є операція максимуму:

$$\overline{Q}(y) = \overline{Q}_1(y) \vee \overline{Q}_2(y) \vee ... \vee \overline{Q}_j,$$

де $\overline{Q}(y)$ – логічна нечітка множина, що відповідає результату роботи логічної нечіткої системи; $\overline{Q}_1(y) \vee \overline{Q}_2(y) \vee ... \vee \overline{Q}_j$ – «усічені» нечіткі логічні множини, що відповідають виводам продукційним активним правилам.

7. Етап дефазифікації. Отриманий нечіткий результат логічного виводу приводиться до чіткого представлення, із застосуванням методу центра ваги.

$$y = \frac{\sum_{j=1}^{R} b_j \int \mu_{\overline{Q}_j(y)} dy}{\sum_{j=1}^{R} \int \mu_{\overline{Q}_j(y)} dy}, \qquad (3.4)$$

де y - чітке значення результату виходу логічної нечіткої системи; b_j - центри функцій приналежності відповідних термів онтології вихідної нечіткої змінної y для j - го правила продукції; R – кількість правил логічних нечітких продукцій; $\int \mu_{\overline{Q}_j(y)} dy$ – величина площі під усіченою нечіткою множиною $\overline{Q}_j$ для j-го правила продукції.

Для прискоренного проведення обчислень застосовується дискретна форма:

$$y = \frac{\sum_{j=1}^{R} a_j b_j}{\sum_{j=1}^{R} a_j} \qquad (3.5)$$

3.3 Система нечіткого логічного виводу вразливостей та загроз інформаційної безпеки

При побудові системи логічного нечіткого виводу про виникнення вразливостей та загроз інформаційної безпеки конфіденційних даних на основі проведеного аналізу потоку текстових повідомлень тематичних інтернет-ресурсів як вхідними параметрами можуть виступати наступні показники статистичного аналізу - середній рейтинг авторів текстових повідомлень, створених у період часу проведення аналізу, частота виникнення нових текстових повідомлень, що містять терміни вразливостей та загроз.

Емпіричні знання, описаних закономірностей зміни частоти виникнення текстових повідомлень тематичних інтернет-ресурсів та рейтингу їх авторів, залежно від значущості інформації мережі, що обговорюється на форумі, можна представити у вигляді набору евристичних правил, представлених у табл. 3.1.

Таблиця 3.1 - Евристичні правила функціонування форумів тематичних інтернет- ресурсів

№ п/п	Правило
1	Якщо частота появи текстових повідомлень на форумі дуже висока, рівень рейтингу авторів високий - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже висока.
2	Якщо частота появи текстових повідомлень на форумі висока, рівень рейтингу авторів високий - ймовірність виникнення вразливості або загрози інформаційної безпеки висока.
3	Якщо частота появи текстових повідомлень на форумі середня, рівень рейтингу авторів високий - ймовірність виникнення вразливості або загрози інформаційної безпеки середня.
4	Якщо частота появи текстових повідомлень на форумі низька, рівень рейтингу авторів високий - ймовірність виникнення вразливості або загрози інформаційної безпеки низька.

Продовження табл. 3.1

5	Якщо частота появи текстових повідомлень на форумі дуже низький, рівень рейтингу авторів високий - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже низький.
6	Якщо частота появи текстових повідомлень на форумі дуже високий, рівень рейтингу авторів середній - ймовірність виникнення вразливості або загрози інформаційної безпеки висока.
7	Якщо частота появи текстових повідомлень на форумі висока, рівень рейтингу авторів середній - ймовірність виникнення вразливості або загрози інформаційної безпеки середня.
8	Якщо частота появи текстових повідомлень на форумі середня, рівень рейтингу авторів середній - ймовірність виникнення вразливості або загрози інформаційної безпеки низька.
9	Якщо частота появи текстових повідомлень на форумі низька, рівень рейтингу авторів середній - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже низька.
10	Якщо частота появи текстових повідомлень на форумі дуже низька, рівень рейтингу авторів середній - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже низька.
11	Якщо частота появи текстових повідомлень на форумі дуже висока, рівень рейтингу авторів низький - ймовірність виникнення вразливості або загрози інформаційної безпеки середня.
12	Якщо частота появи текстових повідомлень на форумі висока, рівень рейтингу авторів низький - ймовірність виникнення вразливості або загрози інформаційної безпеки низька.
13	Якщо частота появи текстових повідомлень на форумі середня, рівень рейтингу авторів низький - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже низька.
14	Якщо частота появи текстових повідомлень на форумі низька, рівень рейтингу авторів низький - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже низька.
15	Якщо частота появи текстових повідомлень на форумі дуже низька, рівень рейтингу авторів низький - ймовірність виникнення вразливості або загрози інформаційної безпеки дуже низька.

На підставі наведених правил в табл. 3.1 побудовано базу даних правил системи логічного нечіткого виводу, як одна з вхідних лінгвістичних змінних нечіткої системи використовується частота появи нових текстових повідомлень: β_1 – «частота появи текстових повідомлень», друга лінгвістична змінна β_2 – «рівень рейтингу авторів повідомлень». Ймовірність виникнення вразливості чи загрози інформаційної безпеки конфіденційних даних буде вихідна лінгвістична змінна: β_3 – «імовірність виникнення вразливості чи загрози інформаційної безпеки конфіденційних даних». Для скороченого запису продукційних правил доцільно застосовувати символічні позначення термів, що наведено у табл. 3.2.

Таблиця 3.2 - Символічні позначення термів

№ п/п	Терм	Позначення
1	Дуже високий	ДВ
2	Високий	В
3	Середній	С
4	Низький	Н
5	Дуже низький	ДН

З урахуванням введених позначень та змінних, проектована система логічного нечіткого виводу міститиме 15 правил логічних нечітких продукцій, представлених у табл. 3.3. В якості, як терм-множини першої лінгвістичної змінної нечіткої системи буде використовуватися множина $T_1 = \{$«*дуже висока*», «*висока*», «*середня*», «*низька*», «*дуже низька*»$\}$, або символічному представленні $T_1 = \{ДВ, В, С, Н, ДН\}$ з функціями приналежності $\mu_{T_1}(x)$, наведеними на рис. 3.3. Для другої логічної лінгвістичної змінної терм-множиною використовується множина $T_2 = \{$«*висока*», «*середня*», «*низька*»$\}$, або в символічному представленні $T_2 = \{В, С, Н\}$ з функціями приналежності $\mu_{T_2}(x)$, наведеними на рис. 3.4. Терм - множиною логічної вихідної лінгвістичної змінної

використовується множина $T_3 = \{$«*дуже висока*», «*висока*», «*середня*», «*низька*», «*дуже низька*»$\}$, або символічному представленні $T_3 = \{ДВ, В, С, Н, ДН\}$ з функціями приналежності $\mu_{T_3}(x)$, наведеними на рис. 3.5.

Таблиця 3.3 - Правила логчних нечітких продукцій

Правило_1	*ЯКЩО* «β_1 *є ДВ*»	*І* «β_2 *є В*»	*ТО* «β_3 *є ДВ*»
Правило_2	*ЯКЩО* «β_1 *є В*»	*І* «β_2 *є В*»	*ТО* «β_3 *є В*»
Правило_3	*ЯКЩО* «β_1 *є С*»	*І* «β_2 *є В*»	*ТО* «β_3 *є С*»
Правило_4	*ЯКЩО* «β_1 *є Н*»	*І* «β_2 *є В*»	*ТО* «β_3 *є Н*»
Правило_5	*ЯКЩО* «β_1 *є ДН*»	*І* «β_2 *є В*»	*ТО* «β_3 *є ДН*»
Правило_6	*ЯКЩО* «β_1 *є ДВ*»	*І* «β_2 *є С*»	*ТО* «β_3 *є В*»
Правило_7	*ЯКЩО* «β_1 *є В*»	*І* «β_2 *є С*»	*ТО* «β_3 *є С*»
Правило_8	*ЯКЩО* «β_1 *є С*»	*І* «β_2 *є С*»	*ТО* «β_3 *є Н*»
Правило_9	*ЯКЩО* «β_1 *є Н*»	*І* «β_2 *є С*»	*ТО* «β_3 *є ДН*»
Правило_10	*ЯКЩО* «β_1 *є ДН*»	*І* «β_2 *є С*»	*ТО* «β_3 *є ДН*»
Правило_11	*ЯКЩО* «β_1 *є ДВ*»	*І* «β_2 *є Н*»	*ТО* «β_3 *є С*»
Правило_12	*ЯКЩО* «β_1 *є В*»	*І* «β_2 *є Н*»	*ТО* «β_3 *є Н*»
Правило_13	*ЯКЩО* «β_1 *є С*»	*І* «β_2 *є Н*»	*ТО* «β_3 *є ДН*»
Правило_14	*ЯКЩО* «β_1 *є Н*»	*І* «β_2 *є Н*»	*ТО* «β_3 *є ДН*»
Правило_15	*ЯКЩО* «β_1 *є ДН*»	*І* «β_2 *є Н*»	*ТО* «β_3 *є ДН*»

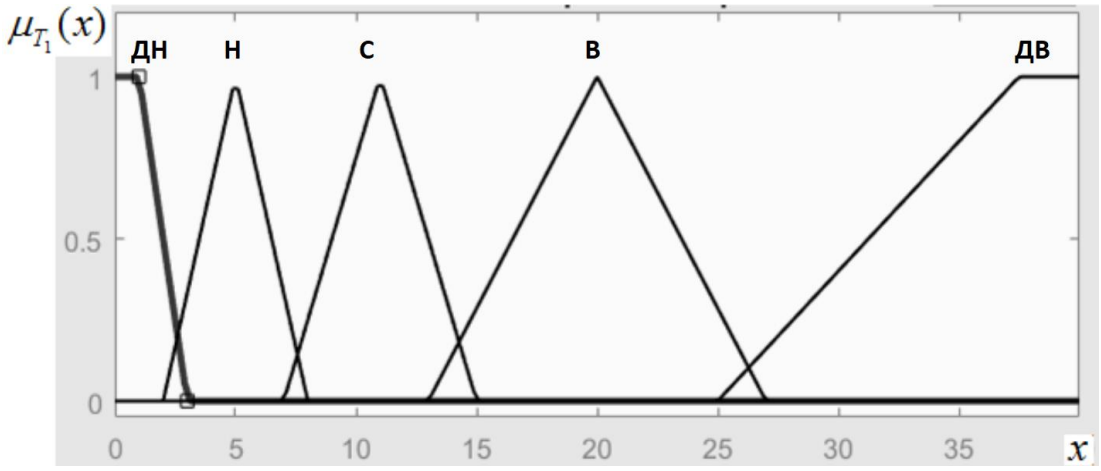


Рисунок 3.3 – Графік функцій приналежності термів для вхідної логічної змінної «Частота появи текстових повідомлень»

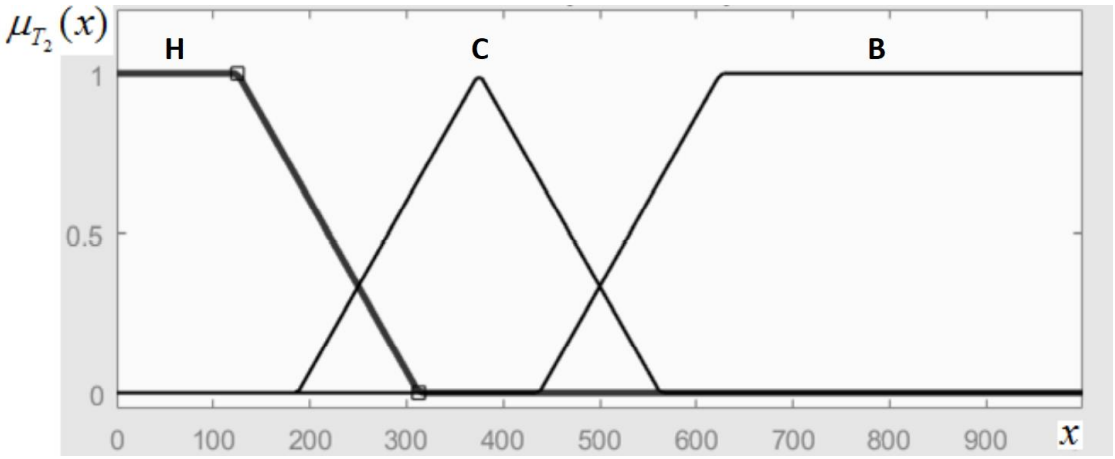


Рисунок 3.4 – Графік функцій приналежності термів для вхідної логічної змінної «Рівень рейтингу авторів повідомлень»

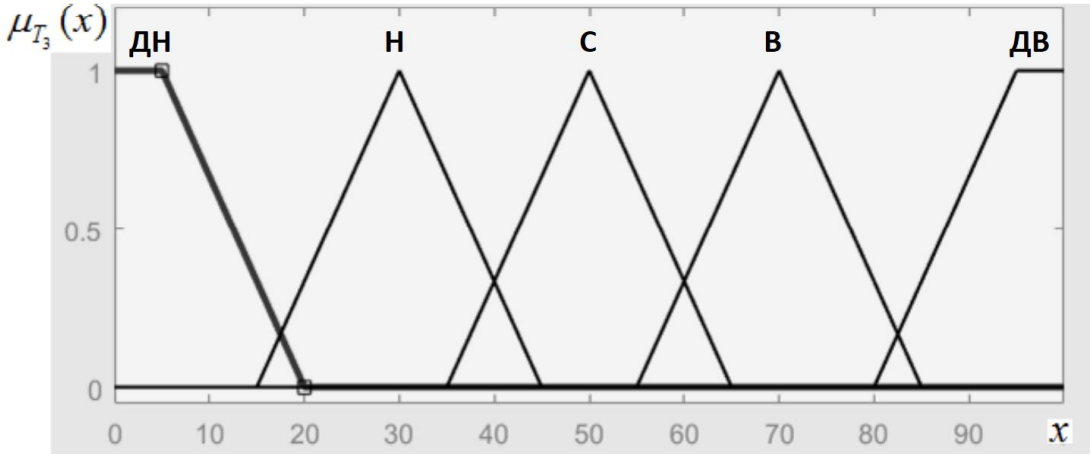


Рисунок 3.5 – Графік функцій належності термів для вихідний логічної змінної «Ймовірність виникнення вразливості чи загрози інформаційної безпеки даних»

Таким чином, при цьому, частота появи текстових повідомлень вимірюється в одиницях на добу, середній рівень рейтингу авторів тематичних повідомлень – в одиницях, ймовірність виникнення вразливості чи загрози – у відсотках.

Запропонована база даних правил нечітких продукцій та функції приналежності для системи логічного нечіткого вививоду про виникнення вразливостей та загроз інформаційної безпеки конфіденційних даних, ґрунтується на проведенні аналізу потоку даних тематичних форумів інтернет-ресурсів, відрізняються від наявних, можливістю адаптивного опису закономірностей процесу наповнення інтернет - форумів новими текстовими повідомленнями, шляхом застосування додаткових вхідних параметрів системи логічного нечіткого

виводу та модифікації функцій приналежності, що дозволяє покращити якість прогнозування можливих вразливостей та загроз.

3.4 Висновки

Запропоновано алгоритм проведення аналізу потоку текстових повідомлень тематичних форумів інтернет-ресурсів, що відрізняється від наявних, можливістю здійснювати обчислювати статистичні параметри, семантичну фільтрацію текстових повідомлень для побудови системи логічного нечіткого виводу для прогнозування подій під час проведення аналізу заданої предметної області.

Запропонована база даних правил логічних нечітких продукцій та функції приналежності системи логічного нечіткого виводу про виникнення вразливостей та загроз інформаційної безпеки конфіденційних даних, ґрунтуючись на проведенню аналізу потоку даних форумів тематичних інтернет-ресурсів, відрізняються, від наявних можливістю адаптивного опису закономірностей процесу наповнення інтернет - форумів новими текстовими повідомленнями, шляхом модифікації функцій приналежності, застосуванням додаткових вхідних параметрів системи логічного нечіткого виводу, що дозволяє покращити якість прогнозування виявлення вразливостей та загроз.

4 ІНФОРМАЦІЙНО-АНАЛІТИЧНА СИСТЕМИ ПРОГНОЗУВАННЯ ВРАЗЛИВОСТЕЙ ТА ЗАГРОЗ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

4.1 Засоби моделювання нечітких інформаційно-аналітичної систем та морфологічного аналізу потоку повідомлень

При проектуванні інформаційно-аналітичної системи прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних були використанні: СУБД MySQL, засіб для логічного нечіткого виводу Logic Fuzzy Designer, редактор онтологій Protégé, засіб морфологічного аналізу тексту повідомлень Mystem.

Проектування нечіткої системи проведено на основі візуального моделювання систем за допомогою мови UML із застосуванням програмної платформи StarUML - описані користувачі та їх функції. логічної та фізичної ER-моделі бази даних. Логічне моделювання нечіткої системи реалізовано у вигляді UML-діаграм: діяльності, послідовності дій, класів. Фізичне моделювання нечіткої системи представлено в UML-діаграм розгортання та компонентів. Наведено опис алгоритму роботи системи та модулів. Запропоновано базу знань у форматі онтології вразливостей та загроз інформаційної безпеки конфіденційних даних, також формування набору логічних правил нечітких продукцій, на основі яких здійснюється логічний нечіткий вивод.

На теперішній час, з метою автоматизації широкого спектру вирішення задач застосовуються експертні системи та інформаційні технології. Аналіз проведених досліджень дозволяє зробити висновок, що для вирішення задачі магістерського дослідження з розробки інформаційно-аналітичної нечіткої системи для логічного нечіткого виводу про появу вразливостей та загроз інформаційної безпеки, автоматизації проведення аналізу потоку повідомлень тематичних інтернет-ресурсів, про доцільність використання експертних систем прогнозування. Для вирішення задач прогнозування вразливостей та загроз інформаційної безпеки

конфіденційних даних на основі потоку тематичних повідомлень інтернет-ресурсів, з використанням запропонованих алгоритмів та методу, можуть використовуватися експертні системи прогнозування гібридного типу, призначені для використання на інформаційно - обчислювальній техніці загального призначення.

При прогнозуванні вразливостей та загроз інформаційної безпеки конфіденційним даним на основі отриманих повідомлень тематичних інтернет-форумів використання онтології предметної області відіграє ключову позицію. Успіх проведення аналізу повідомлень форуму залежить від способу побудови онтології предметної області. Робота з онтологіями передбачає використання методологій та методів їх побудови, із застосуванням прикладних інструментів та спеціалізованих мов програмування, також вирішення задач, пов'язаних з забезпеченням життєвого циклу та їхнього розробкою.

В якості основні програмних інструментів для розробки онтологій предметної області виступають редактори онтологій. Основна їхня функція - забезпечення можливостей формалізації знань про предметну область, для якої проводиться аналіз у заданому форматі онтологічної структури. До основних функцій сучасних редакторів онтології відносяться: імпорт онтології із зовнішніх форматів та експорт у потрібний формат; інтерактивна розробка онтології; редагування метаданих онтології (версії формалізації, загального опису, простору імен); робота з елементами онтології: видалення, редагування, створення відношень онтології, аксіом, об'єктів, класів. Функціональні можливості редакторів онтології можуть бути розширені, підключенням додаткових модулів та плагінів для візуалізації онтологій, несуперечності, перевірки логічної цілісності. Сучасні редактори онтологій розрізняються реалізованих у них наборами функцій, форматами представлення та зберігання даних, можливостями проведення модифікації вхідного коду. Результати порівняння характеристик та параметрів сучасних та популярних редакторів онтологій наведені в табл. 4.1.

Таблиця 4.1 - Таблиця характеристик редакторів онтології

№ п/п	Редактор онтології	Модель	Мова ПЗ	Мова представлення	Зберігання онтологій	Розширення
1	Protégé	Local	Java	OKBC	Файли, СУБД	Плагіни
2	Ontolingua	Web	Lisp	Ontolingua	Файли, СУБД	-
3	OntoStudio	Local	Java	OXML	Файли, СУБД	Плагіни
4	OntoEdit	Local	Java	OXML	Файли	Плагіни
5	Oiled	Local	Java	DAWL+OIL	Файли	-
6	OntoSauns	Web	Lisp	LOOM	Файли	-
7	WebOnto	Web	Lisp, Java	OCML	Файли	-
8	WebODE	Web	Java	-	СУБД	Сервер додатків

Найбільш популярним редактором онтологій на теперішній час - редактор Protégé. Архітектура редактора Protégé легко розширюється, вільно поширюється, підтримка модулів розширення, має відкритий вихідний код. Редактор Protégé використовується для розробки бази знань (онтології) вразливостей та загроз інформаційної безпеки конфіденційних даних, застосовується для проведення обчислення результатів експериментів, які є основою для проведення оцінки ефективності запропонованих алгоритмів та моделей, в основу яких покладено тезаурус інформаційної безпеки, класифікацію вразливостей і загроз.

Моделювання нечітких логічних систем засобами нечіткої логіки, на теперішній час із використанням більше 40 інформаційних систсем, які мають необхідні відповідні функціональні можливості. Порівняння нечітких логічних систем їх основних характеристик та параметрів найбільш наведено в табл. 4.2.

Таблиця 4.2 - Порівняння основних характеристик та параметрів програмних систем логічного нечіткого виводу

Програмне забезпечення	Графічна система	Кількість функцій	Інтерфейс	Мова програмування
MathCAD	2D, 3D анімації	до 300	Простий	Математична мова розрахунків
Excel	2D, 3D	Більше 300	Простий	Visual Basic
MatLAB	2D, 3D	Більше 300	Середній	Visual Basic, C++, Java
Mathematica	2D, 3D	Більше 700	Складний	Власна мова програмування
MathConnex	2D, 3D	Більше 1000	Складний	Власна мова програмування

Програмні інструменти нечіткого логічного моделювання, що входять до складу програмної платформи MatLab користуються популярністю при вирішенні задач, пов'язаних з застосуванням та розробкою логічних нечітких моделей. Система MatLab використана як один із програмних інструментів для проведення експериментальної оцінки розроблених алгоритмів прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних. Для реаалізації логічного нечіткого моделювання в MatLab використаний компонент Fuzzy Logic Toolbox, функції якого дозволяють реалізовувати операції нечіткого виводу та нечіткої логіки. Для реалізації алгоритму семантичної фільтрації текстових повідомлень тематичних форумів інтернет-ресурсів проведено аналіз сучасних програмних інструментів, які забезпечують функціями семантичного та морфологічного аналізу текстових повідомлень, порівняльна характеристика яких наведена в табл. 4.3. Морфологічні процесори виконують відповідні функції лематизації словоформ. Можливості проведення морфологічного синтезу реалізовані у двох розглянутих процесорів, у багатьох задачх комп'ютерної лінгвістики морфологічного синтезу дана функція вкрай важлива.

Таблиця 4.3 - Порівняльна характеристика програмних засобів морфологічного аналізу текстстових повідомлень

№ п/п	Система	Відкриті вихідні коди	Швидкість слів в сек.	Підключення словників	Обєм словника тис. слів
1	AOT	Так	60-90 тис.	Ні	160
2	MyStem	Ні	100-120 тис.	Так	>250
3	TreeTagger	Ні	20-25 тис.	Так	210
4	Pymorphy2	Так	80-100 тис.	Ні	250

Вихідні коди двох із наведених процесорів (табл. 4.3) є закритими, у зв'язку з чим програмні засоби поширюються виключно у вигляді бінарних файлів. Також словник процесора TreeTagger доступний у вигляді бінарного файлу, закритим є словник системи MyStem. Швидкість обробки слів у наведених програмних платформ є достатньо високою. Для роботи з обмеженими предметними областями, особливо важливою задачею є можливість підключення словника даних. Дана функція реалізована у процесорі MyStem. Кожна морфологічна програмна система використовує власну систему морфологічних тегів, таким чином, у зв'язку з цим порівняти результати роботи даних процесорів на однакових потоках текстів достатньо складно. На рис. 4.1 наведено етапи нечіткого виводу роботи інформаційно-аналітичної системи прогнозування вразливостей та загроз інформаційної безпеки.

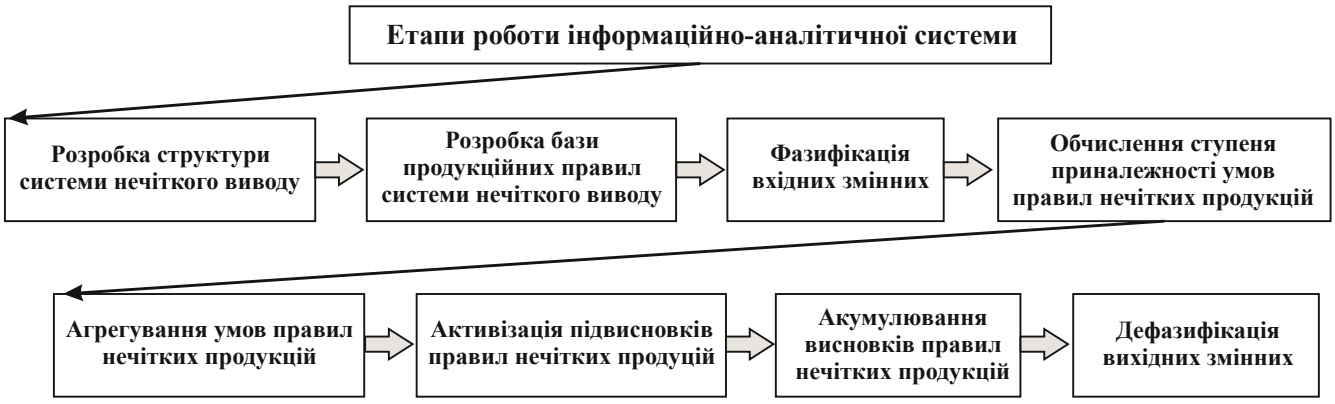


Рисунок 4.1 - Етапи нечіткого виводу роботи інформаційно-аналітичної системи

4.2 Інформаційно-аналітична система прогнозування вразливостей та загроз інформаційної безпеки

На теперішній час існують програмні інструменти, які надають функції збору текстових повідомлень, що розміщуються на тематичних інтернет-сервісах. Їх застосування в практичному використанню дозволяє реалізувати функції нечіткої інформаційно-аналітичної системи формування інтернет - потоку повідомлень різних дискусійних тематичних інтернет-ресурсів. Так як, форуми тематичних інтернет-ресурсів представляють сховища неформалізованих даних, щодо інформаційної безпеки та технологій, містять нечіткі поняття та знання (відсутні будь-які формати викладу текстових повідомлень, учасниками застосовується специфічний сленг, який є у користувачів інтернет-дискусій), доцільно, в даній ситуації використовувати для роботи з даними форумами механізми нечіткої логіки. Обґрунтованість застосування нечітких логічних моделей пов'язана зі значною часткою невизначеності потоку повідомлень, обумовленої складністю предметної області та неповнотою інформації. Як інструмент для досягнення поставленої задачі пропонується використовувати нечітку інформаційно-аналітичну систему прогнозування вразливостей та загроз інформаційної безпеки, що надає функціональні можливості, які представленні на рис. 4.2.

Ефективність роботи інформаційно-аналітичної системи прогнозування вразливостей та загроз інформаційної безпеки в значній мірі залежить від якості використовуваної у ній бази продкуційних правил (бази знань). База знань є сполучною ланкою між ключовими модулями системи та сховищем даних. До бази продкуційних правил включено список тематичних форумів та онтологію предметної області

Організацію процесу аналізу потоку текстових повідомлень тематичних форумів інтернет-ресурсів та прогнозування вразливостей та загроз інформаційної безпеки конфіденційних даних відображає структура нечіткої інформаційно-аналітичної системи, яка представлена на рис. 4.3. Стрілками темного кольору

позначені потоки текстових повідомлень інтернет-ресурсів в процесі пошуку джерел даних тематичних форумів, предметної області що представляє інтерес. Світлими стрілками позначений потік повідомлень тематичних форумів інтернет-ресурсів у процесі прогнозування вразливостей та загроз інформаційної безпеки конфіденційним даним.

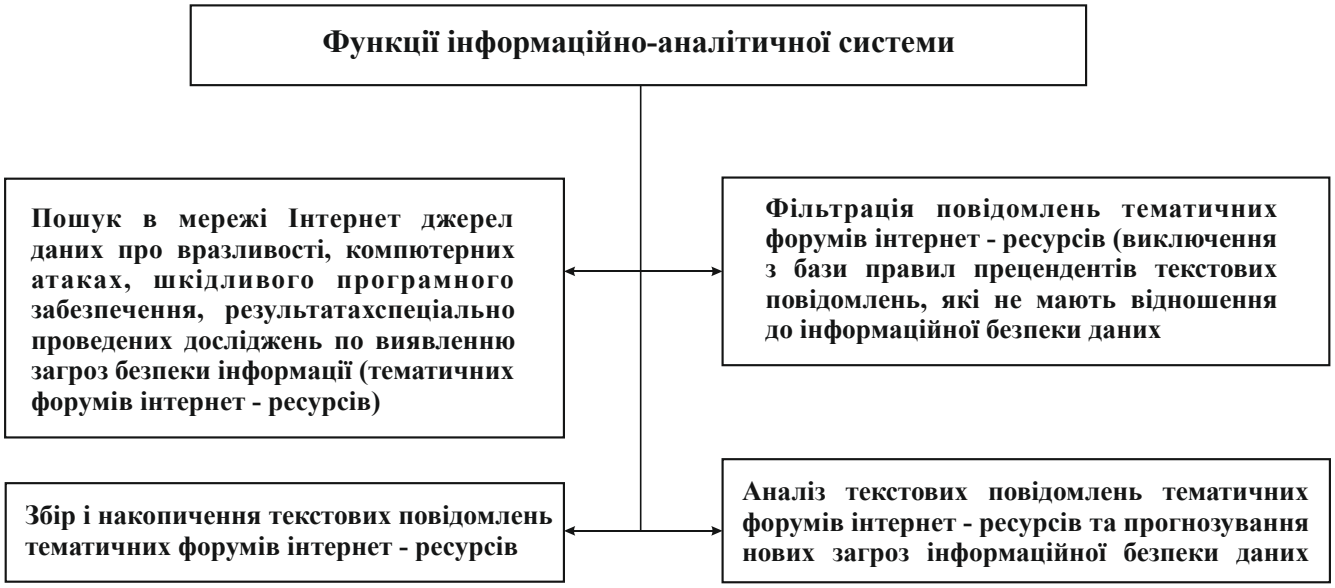


Рисунок 4.2 – Функції інформаційно-аналітичної системи аналізу потоку повідомлень тематичних інтернет-форумів

Список тематичних джерел форумів містить адреси інтернет- ресурсів, на яких розміщуються текстові публікації про шкідливе програмне забезпечення, вразливості та комп'ютерні атаки.

На початковому етапі роботи інформаційно-аналітичної системи список формується експертним шляхом, із загальної кількості форумів тематичних інтернет – ресурсів виділяються ті, тематика яких дозволяє інтернет - інформацію віднести до хакерських (інформація містить результати спеціалізованих досліджень з виявлення вразливостей та загроз інформаційної безпеки конфіденційних даних, повідомлення про комп'ютерні атаки, вразливості, шкідливе програмне забезпечення). Автоматизоване виявлення нових форумів тематичних інтернет-ресурсів, в даній ситуації, можливо, шляхом проведення аналізу різноманітних форумів інтернет-ресурсів, з використанням

запропонованих критеріїв відбору текстових повідомлень, що належать до заданої предметної області, для якої проводиться аналіз з використання онтології.

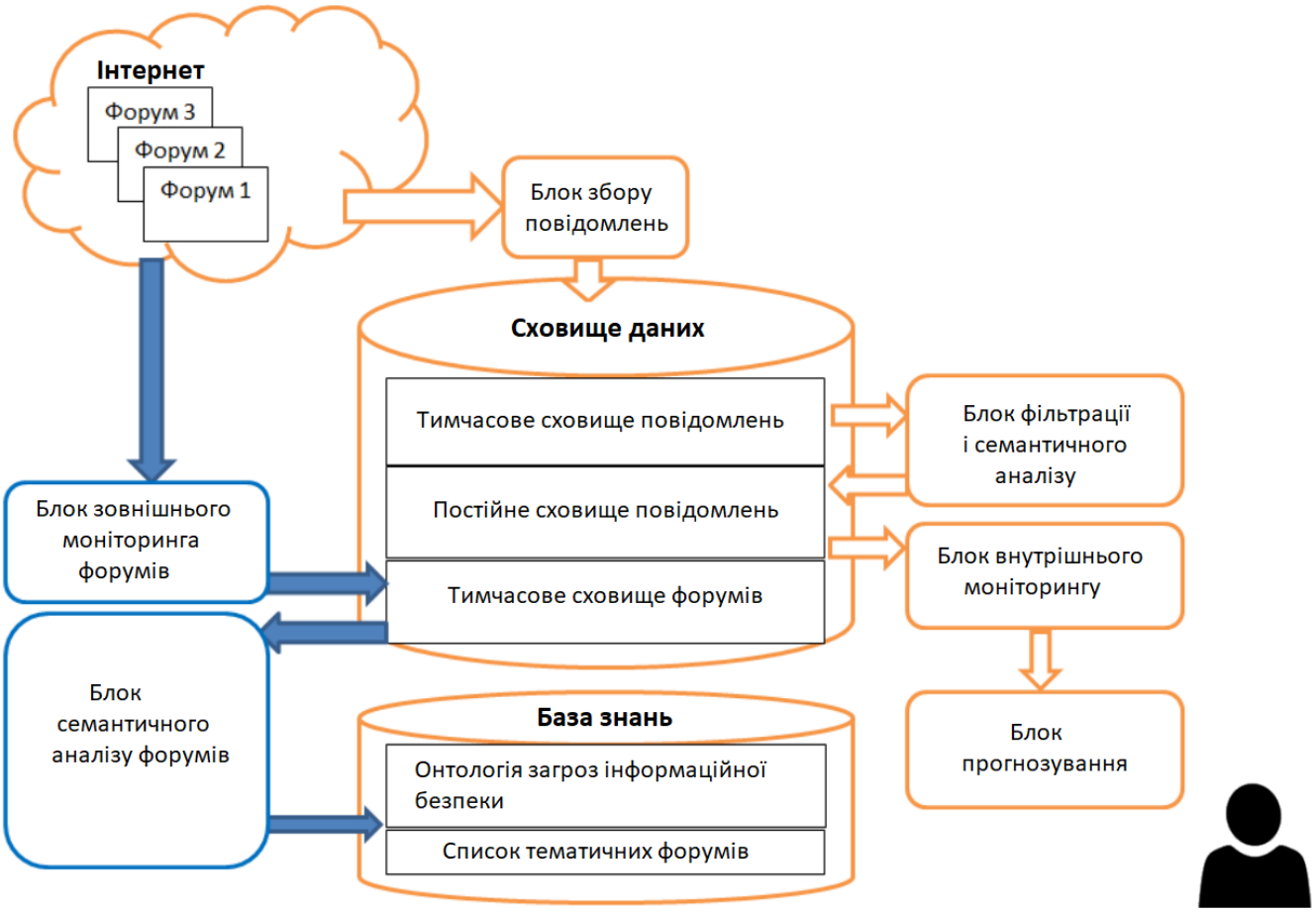


Рисунок 4.3 - Структура нечіткої інформаційно-аналітичної системи

На теперішній час для вирішення задачі проектування та розробки інформаційно-обчислювальних систем використовується універсальна мова моделювання UML - дозволяє реалізувати об'єктно-орієнтований підхід до проектування систем, будувати моделі систем із зазначенням їх основних якостей.

Для розробки концептуальної моделі інформаційно-обчислювальних систем застосовуються моделі бізнес-об'єктів: діаграми діяльності, варіантів використання, послідовностей дій. Під час проектування логічної моделі інформаційно-обчислювальної системи вимоги до системи формуються на основі застосування діаграм варіантів використання. На етапі попереднього проектування інформаційної системи використовуються діаграми послідовностей, станів, класів. На етапі проектування фізичної моделі інформаційної системи детальне

проектування виконується із використанням діаграм класів, компонентів та розгортання. Для опису функціонального призначення інформаційної системи застосовуються діаграми UML варіанти використання. Діаграми є концептуальним представленням інформаційної системи (вхідними моделями) у процесі розробки та проектування. В залежності від розв'язуваних задач, роботу з інформаційною системою, можуть здійснювати користувачі двох типів:

1. Експерт, для роботи доступні чотири модулі інформаційної системи: редактор онтології (Protégé), використовується для формалізації накопичуваних експертних знань в форматі онтології, що відносяться до заданої предметної області; редактор функцій приналежності вхідних та вихідних параметрів, правил логічних нечітких продукцій (Fuzzy Logic Designer), використовується для здійснення логічного нечіткого виводу про вразливості та загрози інформаційної безпеки конфіденційних даних; підсистема розширення ядра онтології, для вилучення термінології з текстів предметної області, експерту надається можливість оцінювати вилучення на термінологічність і вносити їх до онтології предметної області, розширюючи базу знань правил продукцій; редактор списку тематичних форумів інтернет-ресурсів, використовується для формування потоку текстових повідомлень, що аналізуються.

2. Спеціаліст з інформаційної безпеки - необхідно прийняти рішення про достатність заходів, для здійснення захисту конфіденційних даних. Користувачу надається доступ до підсистеми логічного нечіткого виводу про виникнення вразливостей та загроз безпеки інформації, в основі лежить проведення аналізу текстових повідомлень тематичних форумів інтернет-ресурсів. Під час отримання виводу про виникнення вразливості чи загрози безпеки інформації, користувачу надана можливість оцінити актуальність безпеки для інформаційно-обчислювальної системи, та прийняти відповідні запобіжні заходи усунення негативних чинників. Нечітка інформаційно-аналітична система розробляється для обробки потоку інтернет повідомлень тематичних форумів інтернет-ресурсів, фільтрації текстових повідомлень на основі онтології заданої предметної області,

проведення статистичного аналізу, сематичного аналізу потоку даних, логічного нечіткого виводу, базується на результатах потоку текстових повідомлень.

При використанні об'єктно-орієнтованого підходу розробки системи, центральне місце займає розробка моделей, представлених у вигляді діаграми класів. Діаграми класів розглядаються на початкових етапах розробки та моделювання інформаційної системи. Для представлення інформаційно-обчислювальної аналітичної системи задіяно сім класів: Автори повідомлень, Форуми, Онтологія, Теми, Розділи, Інформаційні ресурси, Повідомлення. На рис. 4.4. наведено діаграму класів інформаційно-аналітичної системи.

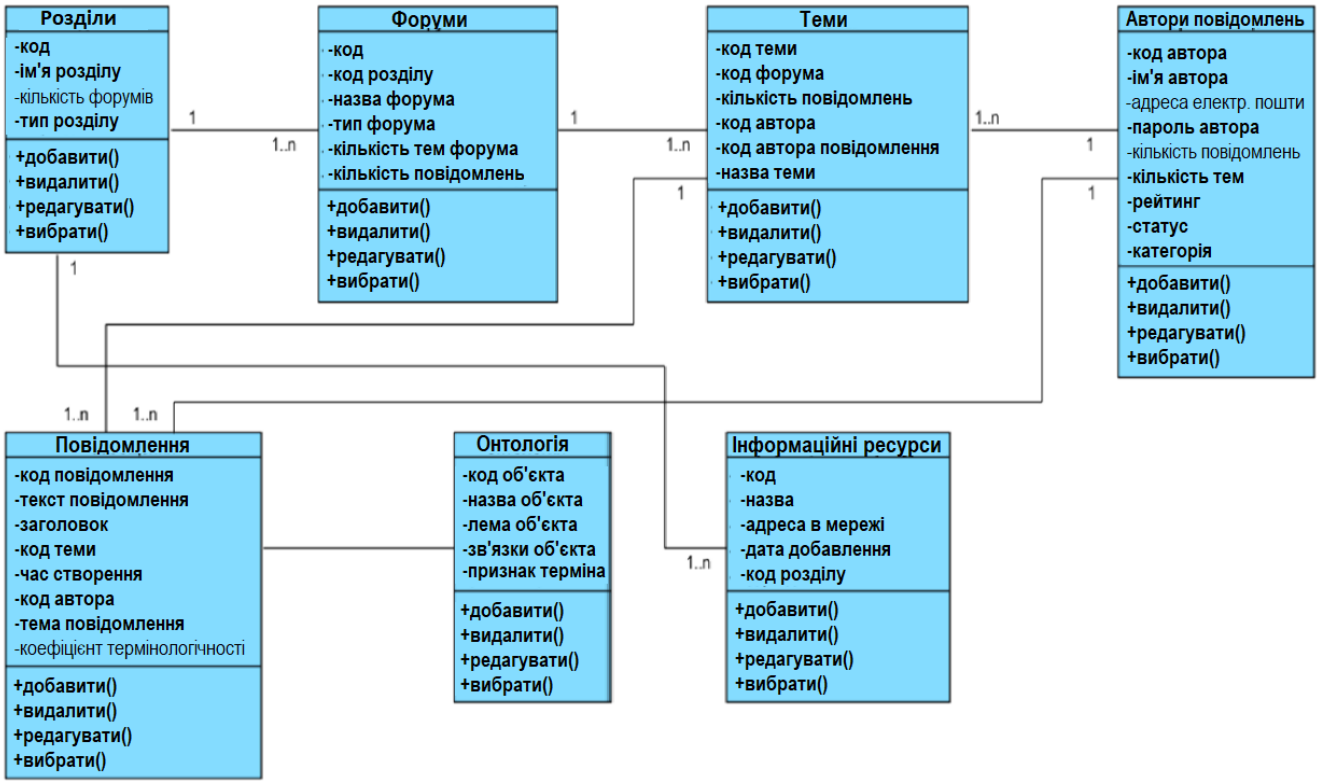


Рисунок 4.4 – Діаграма класів нечіткої інформаційно-аналітичної системи

Для опису взаємодії інформаційної системи об'єктів використанні діаграми послідовності дій. Діаграми описують послідовності, у яких об'єкти діаграми отримують та надсилають повідомлення на протязі часу. Діаграма діяльності інформаційно-обчислювальної аналітичної системи, наведена на рис. 4.5. Діаграми послідовності визначають основні повідомлення, на які реагують об'єкти, компоненти, відображають динамічну складову інформаційно системи.

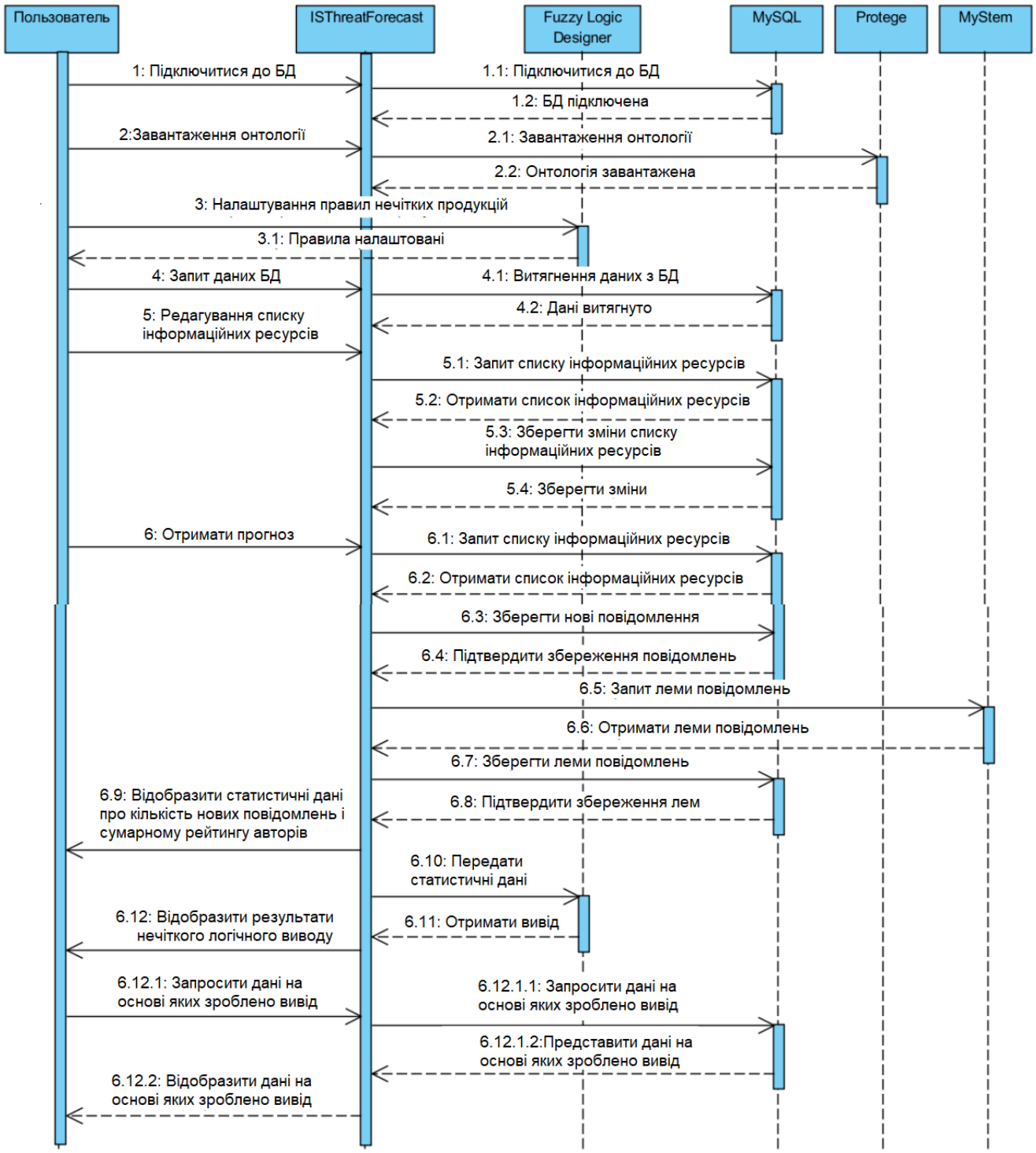


Рисунок 4.5 – Діаграма діяльності інформаційно-аналітичної системи

Запропонована діаграма описує логічну модель, в якій як засіб для реалізації нечіткого логічного виводу використовується Fuzzy Logic Designer, MySQL, редактор онтології – Protégé, засоби семантичної обробки MyStem.

Розглянуті діаграми реалізують концептуальну сторону побудови моделі інформаційно-аналітичної системи, подання системи здійснюється на логічному

рівні. Для реалізації фізичної системи, потрібно реалізувати в матеріальні сутності всі елементи логічного представлення. Для фізичного представлення моделі використовується діаграма UML розгортання, відображається загальна топологія та конфігурація інформаційно-обчислювальної системи, а також розподіл за окремими вузлами компонентів. Вузлами діаграми інформаційно-обчислювальної системи є персональний комп'ютер користувача, сервер адміністратора. Вузли пов'язані суцільною лінією - наявність фізичного каналу обмінюватись інформацією, пунктирна лінія - вузли взаємодіють шляхом направлення різноманітних звернень та використання файлів обміну інформацією (рис. 4.6).

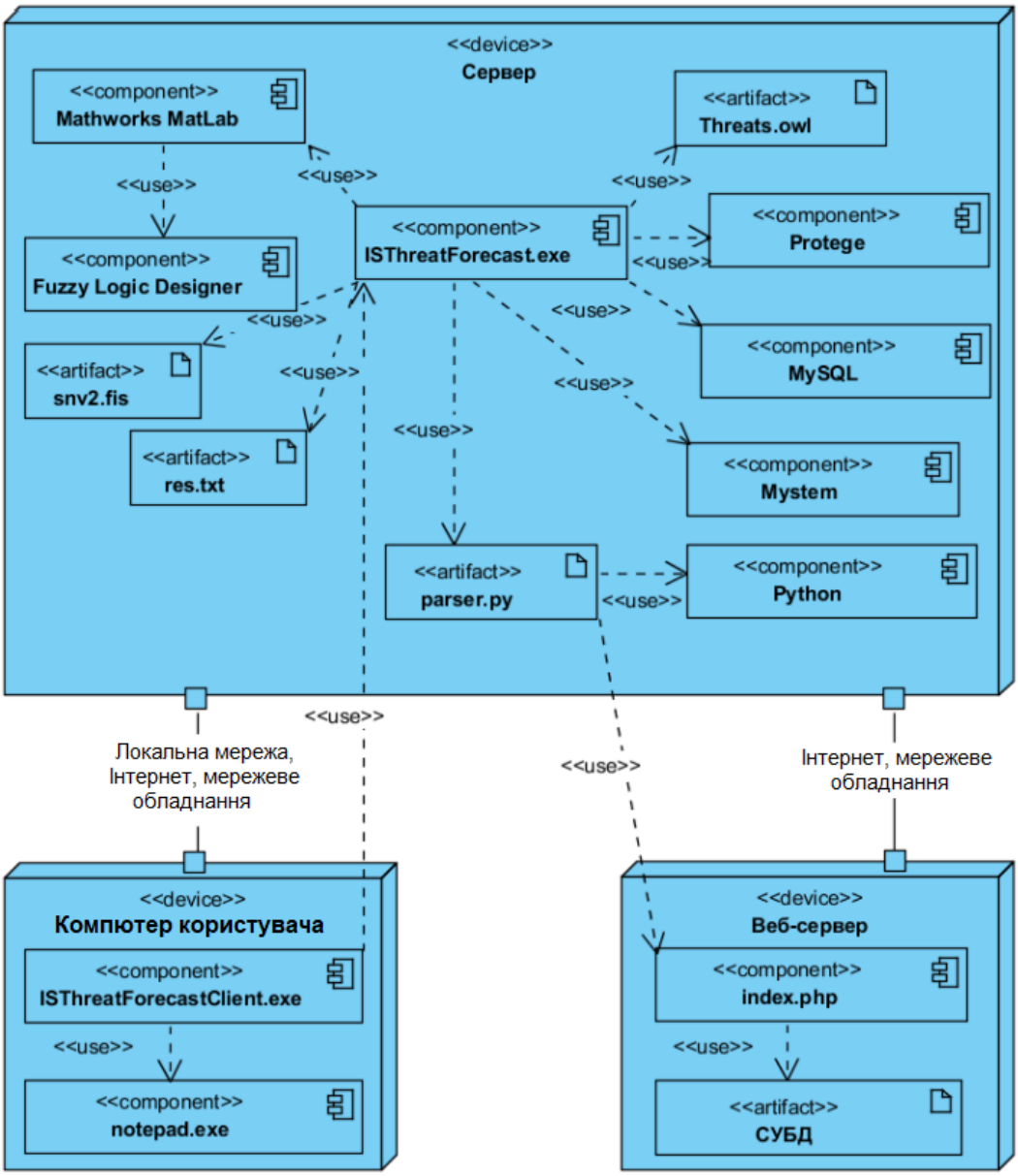


Рисунок 4.6 – Діаграма розгортання інформаційної системи

Інформаційно-аналітична система використовує онтологію інформаційної безпеки, побудовану на початковому етапі роботи системи, експертним шляхом на основі тезауруса. Розширення онтології здійснюється виявленням нових термінів у потоку текстових повідомлень хакерських термінів відповідно до запропонованих алгоритмів. Для реалізації онтології експертом використовується Protege, результат OWL-онтологія, яка завантажується до бази знань. Тематичні текстові повідомлення інтернет-ресурсів піддаються морфологічному аналізу з використанням Mystem, результати додаються до бази знань.

4.3 Оцінка методу прогнозування вразливостей та загроз інформаційної безпеки

Для оцінки ефективності та адекватності методу прогнозування вразливостей та загроз інформаційної безпеки, а також коректності роботи інформаційної системи, проведено експерименти, під час яких проводився аналіз потіку повідомлень форумів тематичних інтернет-ресурсів. В рамках експерименту проведені наступні дії:

1. На основі тезауруса інформаційної безпеки, класифікацій вразливостей і загроз безпеки інформації, отриманої експериментальним шляхом, побудована онтологія вразливостей і загроз інформаційної безпеки конфіденційних даних.

2. Сформовано набір правил логічних нечітких продукцій - відображають закономірності залежності кількості створюваних на форумі тематичних інтернет-ресурсів текстових повідомлень та середнього рейтингу авторів від ймовірності виникнення вразливостей та загроз інформаційної безпеки конфіденційних даних. Запропоновано базу правил продукцій та визначено функції приналежності для інформаційної системи логічного нечіткого виводу, засновану на використанні алгоритму Мамдані.

3. Здійснено збір текстових повідомлень, відібраних експертним шляхом тематичних форумів інтернет-ресурсів.

4. Результати отриманих ймовірностей виникнення вразливостей та загроз інформаційної безпеки конфіденційних даних співвіднесені зі значеннями кількості записів, включених до бази знань вразливостей та загроз. Проведено розрахунки ефективності запропонованого алгоритму безпеки інформації.

5. Проведено обчислення показників кількості текстових повідомлень форумів тематичних інтернет-ресурсів, середнього рейтингу авторів повідомлень. Отримані результати використані в якості вхідних параметрів інформаційної системи логічного нечіткого виводу, обчисленні значення ймовірності виникнення загроз та вразливостей
інформаційної безпеки;

6. Проведено статистичний та семантичний аналіз отриманих текстових повідомлень шляхом фільтрації даних, що не містять термінів заданої онтології інформаційної безпеки конфіденційних даних.

При вирішенні задачі оцінки якості прогнозування інформаційної системи, використовуються показники, що наведені в табл. 4.4:

Таблиця 4.4 - Показники якості прогнозування аналітичної системи

№ п/п	Назва, формула, опис
1	$MAPE$ – середня абсолютна процентна помилка системи прогнозування $$MAPE = \frac{1}{h}\sum_{i=1}^{h}\left|\frac{f_{T,i} - y_{T+i}}{y_{T+i}}\right| \cdot 100\%, \quad (4.1)$$ де h - довжина інтервалу, на якому проводиться прогнозування загроз; $f_{T,i}$ - прогнозне значення часового ряду, отримане в момент часу T на i кроків наперед; y_{T+i} - значення часового ряду в момент часу $T+i$
2	MAE - середня абсолютна помилка системи прогнозування: $$MAE = \frac{1}{h}\sum_{i=1}^{h}\left| f_{T,i} - y_{T+i}\right|, \quad (4.2)$$

продовження табл. 4.4

3	*RMSE* - квадратний корінь із середньої квадратичної помилки системи прогнозування: $$RMSE = \sqrt{\frac{1}{h}\sum_{i=1}^{h}\left(f_{T,i} - y_{T+i}\right)^2}, \quad (4.3)$$
Для оцінки якості прогнозування загроз зручніше використання середньої абсолютної процентної помилка (*MAPE*), вимірюється у відсотках від значення прогнозованого показника. Показник може бути використаний для порівняння якості прогнозування загроз, систем побудованих із застосуванням різних моделей, також в якості прогнозування конкретних моделей, для яких визначено рівень помилки прогнозування критичний.	
4	Відносна кількість випадків прогнозування до загального числа випадків: $$\eta = \frac{p}{p+q}, \quad (4.4)$$ де p – число випадків прогнозування, які підтверджені фактичними даними; q – число випадків, які не знайшли фактичного підтвердження.
5	Критерій Дарбіна-Уотсона: $$d = \frac{\sum_{i=2}^{m}\left(e_i - e_{i-1}\right)^2}{\sum_{i=2}^{m} e_i^{\,2}}, \quad (4.5)$$ де m - довжина часового ряду, e_i - помилка прогнозування загроз: $e_i = x_i - \bar{x}_i$

Важливим критерієм, що характеризує правильність використання моделі прогнозування, є перевірка на адекватність. Моделі, в яких залишкова компонента має властивості нормальності розподілу, незалежності, випадковості вважаються адекватними. Перевірка кореляції здійснюється із використанням критерію.

Відповідно до критерію Дарбіна-Уотсона, модель адекватна, якщо значення *d*, розраховане за формулою 4.5, близька до 2.

Для оцінки ефективності методу прогнозування вразливостей та загроз безпеки інформації на основі проведеного аналізу текстових повідомлень учасників тематичних форумів інтернет-ресурсів проведено експерименти з автоматизованого збору повідомлень інтернет-ресурсів (darkmoney.cc, rdot.org, grabberz.com, zloy.bz, nulled.io, hakepok.su, verified.cm, hashcrack.in). При обчисленні функцій приналежності вхідних параметрів використовувалися результати проведеного аналізу текстових повідомлень. На основі отриманих результатів із застосуванням запропонованого методу, проведено обчислювальні експерименти щодо формування логічного нечіткого виводу про виникнення вразливостей та загроз інформаційної безпеки конфіденційних даних.

Для оцінки отриманих результатів проведено розрахунки показників *MAPE*, *MAE*, *RMSE* (за формулами 4.1, 4.2, 4.3) для значень прогнозування інформаційно-аналітичної нечіткої системи про виникнення вразливостей та загроз безпеки інформації та кількості виявлених вразливостей та загроз, в період проведення аналізу, а також проведені розрахунки, згладжених часових рядів із інтервалом згладжування три та п’ять діб. Результати представлені у табл. 4.5.

Таблиця 4.5 - Показники якості прогнозування загроз

Показник	Експериментальні дані	Згладжування з періодом 3 доби	Згладжування з періодом 5 діб
MAPE (%)	94,21	147,04	119,73
MAE (%)	26,87	17,82	13,14
RMSE (%)	17,14	12,27	10,01

Для перевірки адекватності розробленої моделі, згідно до критерію Дарбіна-Уотсона (4.5) обчислено значення d = 2,383. Таким чином, відповідно із зазначеним критерієм, запропонована модель вважається адекватною, якщо значення критерію *d* близько 2.

Проведено розрахунок показника точності прогнозування загроз η (4.4), для довірчих інтервалів 10, 15, 20%. Результати обчислень наведено у табл. 4.6.

Таблиця 4.6 - Показника точності прогнозування загроз

Довірчий інтервал	Експериментальні дані	Згладжування з періодом 3 доби	Згладжування з періодом 5 діб
10%	0,517	0,189	0,114
15%	0,554	0,559	0,482
20%	0,683	0,695	0,696

Наведені показники дозволяють зробити висновок, що результати прогнозування інформаційно-аналітичної нечіткої системи в більшості випадків підтверджуються даними бази знань вразливостей та загроз.

Таким чином, спеціаліст із інформаційної безпеки, на основі отриманих результатів прогнозування вразливості або загрози, може оцінити ступінь небезпеки інформаційних ресурсів організації та вжити відповідних заходів щодо нейтралізації загроз та вразливостей.

Покращення якості прогнозування виникнення вразливостей та загроз безпеки інформації з використанням систем логічного нечіткого виводу може сприяти збільшення кількості вхідних змінних, використанню більш точних нечітких правил продукцій, також велике значення має визначення функцій приналежності вихідних та вхідних параметрів системи логічного нечіткого виводу, необхідно враховувати статистичні показники потоку текстових повідомлень форумів тематичних інтернет-ресурсів.

4.4 Висновки

Запропонованого схему структурну інформаційної системи для прогнозування вразливостей та загроз безпеки інформації. Для проведення логічного моделювання інформаційно системи побудовані UML–діаграми діяльності, послідовності дій, класів. Для фізичного моделювання системи розроблено UML-діаграми розгортання та компонентів.

Проведено аналіз систем нечіткого логічного виводу, сучасних засобів обробки великих об’ємів даних, засобів морфологічного аналізу тексту, редакторів онтологій.

Обґрунтовано можливість реалізації інформаційно-аналітичної системи прогнозування вразливостей та загроз безпеки інформації на основі аналізу текстових повідомлень тематичних інтернет-ресурсів з використанням настпних програмних продуктів: СУБД MySQL, редактора онтології – Protégé, системи нечіткого логічного виводу – Fuzzy Logic Designer, засобів морфологічного аналізу даних – Mystem.

Для проведення оцінки отриманих результатів обчисленні показники $MAPE$, MAE, $RMSE$ для значень прогнозування виникнення вразливостей та загроз інформаційної безпеки, а також розрахованим на їх основі згладжених часових рядів з періодом три і п’ять. Для перевірки адекватності моделі, згідно з критерієм Дарбіна-Уотсона, розраховано значення показника $d = 2{,}383$, значення якого вказує що запропонована модель адекватна.

ВИСНОВКИ

В результаті магістерського дослідження вирішена наукова задача, полягає в підвищенні ефективності засобів та методів виявлення вразливостей та загроз інформаційної безпеки конфіденційних даних на основі розробки інформаційно-аналітичної системи та алгоритмів для проведення аналізу потоку повідомлень тематичних форумів інтернет-ресурсів.

Для досягнення мети поставленої в магістерській роботі вирішено та сформульовано наступні задачі:

1. Розроблено модель бази знань тематичного форуму інтернет-ресурсу, призначена для прогнозування вразливостей та загроз безпеки інформації, відрізняється можливістю працювати з різнотипними потоками даних, різних програмних платформ, що застосовуються для реалізації дискусійних тематичних форумів інформаційних ресурсів, а також представлена модель потоку текстових повідомлень, що належать до заданої предметної області, описаної заданою онтологією, відрізняється від аналогів можливістю статистичного аналізу та семантичної фільтрації текстових повідомлень, дозволяє прогнозувати вразливості та загрози, враховуючи їхню приналежність до конкретного форуму, рейтингу автора, кількості повідомлень теми форуму, часу створення, темі форуму.

2. Розроблено алгоритм прогнозування вразливостей та загроз безпеки інформації, заснований на логічному нечіткому виводу, семантичному та статистичному аналізі, відрізняється від аналогів можливістю виявлення вразливостей та загроз до їх безпосередньої реалізації, а також гнучко дозволяє описувати закономірності процесу наповнення тематичних форумів інтернет-ресурсів новими текстовими повідомленнями, що в результаті сприяє покращенню якості прогнозування загроз.

3. Розроблено для вирішення задачі прогнозування вразливостей та загроз безпеки інформації алгоритм проведення аналізу потоку текстових повідомлень тематичних форумів інтернет-ресурсів, заснований на семантичному та

статистичному аналізі, відрізняється від наявних можливістю обчислювати статистичні параметри, здійснювати семантичну фільтрацію текстових повідомлень, для прогнозування подій системи нечіткого логічного виводу.

4. Запропоновано інформаційно-аналітичну систему прогнозування вразливостей та загроз безпеки інформації шляхом проведення автоматизованого аналізу текстових повідомлень форумів тематичних інтернет-ресурсів, реалізує запропоновані алгоритми, дозволяє прогнозувати вразливості та загрози, вживати адекватних заходів щодо захисту інформації. Отримані результати свідчать про ефективність запропонованого методу прогнозування вразливостей та загроз, а також коректній роботі розробленої інформаційно-аналітичної системи та можливості застосування на практиці.

За темою роботи опубліковано 1 теза та 1 наукова стаття.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Архангельский, В.И. Системы фуцци – управления./ В.И. Архангельский, И.Н. Богаенко, Г.Г. Грабовский – К.: Тэхника, 2012. – 208 с.

2. Гошубєв, О.В. Програмно-технічні засоби захисту інформації від комп'ютерних злочинів / О. В. Гошубєв–«Запоріз. ін-т муніцип. упр. і держ.», 2018. – 145

3. Горбулін, П.В. Проблеми захисту інформаційного простору України / М.М. Баченок, П.В. Горбулін – К.: Інтертехнологія, 2019. – 138 с.

4. Джулій, В.М. Інформаційно-ознакова модель шкідливої інформації в соціальних мережах/ В. М. Пічура, І.В. Муляр, О.О Зацепіна, В.М. Джулій,– Вимірювальна та обчислювальна техніка в технологічних процесах № 3 (2022)-73–78с.

5. Джулій, В.М. Модель потоку текстових повідомлень тематичних інтернет-ресурсів системи прогнозування інформаційної безпеки / В.М. Джулій, Н.С. Петляк, Ю.В. Хмельницький, О.В. Пахар - Вісник Хмельницького національного університету. Технічні науки. 2022. № . С. - .

6. Джулій, В.М. Метод класифікації додатків інтернет - трафіка комп'ютерних мереж в умовах невизначеності / В.М. Джулій, Л.В. Солодєєва, О.В. Мірошніченко, // Збірник наукових праць ВІКНУ імені Тараса Шевченка. – К.: ВІКНУ, 2022. – Вип. №74. – С. 73-82.

7. Довгий, С.О. Сучасні телекомунікації: управління, технології, мережі, регулювання, економіка / О.Я. Савченко, С.О. Довгий, П.П. Воробієнко – К.: Український видавничий центр, 2014. – 521 с.

8. Дроб'язко, В. С. Охорона баз даних : регіональні, міжнародні, національні аспекти / В. С. Дроб'язко – К. : Л.-Поліграф, 2018. – 132 с.

9. Дубчак, Л. Метод обробки нечітких даних на основі механізму Мамдані. Системи обробки інформації. 2016. Вип. 7(105). 131с.

10. Ермаков, А.М. Основы конфигурации корпоративных сетей Cisco. А.М. Ермаков – М.: ФГБОУ, 2015. — 458 с.

11. Зайченко, Ю.П. Основи проектування інтелектуальних систем: навчальний посібник / Ю.П. Зайченко. – К.: Слово, 2014. – 352 с.

12. Желдак, Т.А. Нечіткі множини в системах управління та прийняття рішень: Навчальний посібник./ Т.А. Желдак, Л.С. Коряшкіна - Дніпро: НТУ «Дніпровська політехніка» , 2020. — 320 с.

13. Кудінов, В.А. Основи протидії кіберзлочинності. / В. М. Смаглюк, В. Г. Хахановський, В.А. Кудінов. – К. : НАВС, 2016. – 104 с.

14. Кузьменко, Г.Н. Компьютерные сети и сетевые технологии/ Г. Н. Кузьменко– Л: Наука и техника, 2016. – 369 с.

15. Кутузов, О. И. Моделирование и оценка вероятностно-временных характеристик. Инфокоммуникационные сети. / Т. М. Татарникова, О. И. Кутузов - СПб. : РГГМУ, 2017. – 384 с

16. Лавров, Є. А. Математичні методи дослідження операцій. / В. В. Шендрик, Л. П. Перхун, Є. А. Лавров– Суми : СДУ, 2017. – 214 с.

17. Ленков, С.В. Модель безпеки поширення в інформаційно-телекомунікаційних мережах забороненої інформації / В.С. Орленко, С.В. Ленков, А.В. Атаманюк, О.В. Сєлюков, В.М.Джулій, // Збірник наукових праць ВІКНУ ім. Т. Шевченка. – К.: ВІКНУ, 2020. –№68. – С. 53-64.

18. Леоненков, А.В. Нечеткое моделирование средствами fuzzyTECH и MATLAB/ А. Леоненков. – СПб: БХВ-Петербург, 2013. – 736 с

19. Лук’янов, Б. В. Комп’ютерни аналіз даних / Б. В. Лук’янов – К. : Академія, 2017. – 345 с.

20. Матвійчук, А. В. Штучний інтелект в економіці: нечітка логіка, нейронні мережі./ Матвійчук, А. В. – К.: КНЕУ, 2011. – 439 с

21. Остапов, С. Е. Технології захисту інформації: навч. посіб. / С.П. Євсеєв, О.Г. Король, С.Е. Остапов – Харків : ХНЕУ, 2016. – 471 с.

22. Пахар, О.В. Класифікація загроз та вразливостей інформаційної безпеки, характерних для інтернет-ресурсів / В.М. Джулій, О.В. Пахар/ Тези доповідей XVIII

міжнародної наукової конференції молодих учених, аспірантів та студентів. / ред. кол. Д. Струнін – К., - 2022. –С.112

23. Олійник А. О. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 271 с.

24. Соціальні мережі – реальні загрози віртуального світу. [Електронний ресурс]. – Режим доступу : http://ogo.ua/ articles/view/011–02–23/26490.htm.

25. Субботін С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень: Навчальний посібник. / С. О. Субботін— Запоріжжя: ЗНТУ, 2017. — 341 с.

26. Сявавко, М. Математичне моделювання за умов невизначеності. / М. Сявавко, О.Рибицька — Львів: Українські технології, 2014. — 320 с.

27. Трубачев, П. А. Оценка безопасности сетевых информационных технологий / П.А. Трубачев, ред. В. А. Галатенко – СПб.: РГГМУ, 2015. – 358 с.

28. Штовба, С.Д. Проектирование нечетких систем в среде MATLAB / С. Штовба. – М: Горячая линия–Телеком, 2017. – 288 с.

29. Zadeh L. Real–Life Applications of Fuzzy Logic. Fuzzy logic now and then. / Zadeh L. - Hindawi, 2015. p. 125.

30. Zimmerman H. Fuzzy set theory and its applications. Fuzzy logic intoduction. / H. Zimmerman - Kluwer, 1991. p. 315.

ДОДАТОК А

(обовязковий)

Код (лістинг) програмних компонентів взаємодії системи протидії та її поточної онтології з базою даних системи MYSQL

```
using System;
using System.Collections.Generic;
using System.Data.SqlClient;
using System.Linq;
using System.Text;
using System.Threading.Tasks; using MySql.Data.MySqlClient;
namespace Onthology {
class DAO {
private Random random = new Random(); private MySqlConnection connection;
private String connectionInfo = "Database=onthology;Data Source=localhost;User Id=root;Password=Password";
//Інформація для підключення до БД public DAO()
{
CreateConnection(); TestConnection();
}
private void CreateConnection() {
connection = new MySqlConnection(connectionInfo); //Створення підключення до БД
}
public void TestConnection() // тестування створеного підключення {
connection.Open(); connection.Close();
}
public List<Category> LoadFromDB()  //завантаження інформації з БД {
String query1 = "select * from class";
MySqlCommand command = new MySqlCommand(query1, connection); MySqlDataReader reader = null;
List<Category> categories = new List<Category>(); List<Document> documents = new List<Document>(); try
{
connection.Open();
```

```
reader = command.ExecuteReader(); while (reader.Read())
{
if (!Boolean.Parse(reader[2].ToString())) {
Category category = new Category(); category.ID = reader[0].ToString(); category.Name = reader[1].ToString(); categories.Add(category);
} else {
Document document = new Document(); document.ID = reader[0].ToString(); document.Name = reader[1].ToString(); document.Content = reader[3].ToString(); documents.Add(document);
}
}
reader.Close(); connection.Close(); connection.Open();
String query5 = "select * from hierarchy";
command = new MySqlCommand(query5, connection); MySqlDataReader reader5 = command.ExecuteReader(); while (!reader5.IsClosed && reader5.Read())
{
String superId = reader5[1].ToString(); String subId = reader5[2].ToString();
Document doc = FindDocumentById(documents, subId); if(doc != null)
{
Category category = FindCategoryById(categories, superId); category.Documents.Add(doc);
doc.Category = category; }
else {
Category superClass = FindCategoryById(categories, superId); Category subClass = FindCategoryById(categories, subId); superClass.Subcategories.Add(subClass); subClass.Supercategories.Add(superClass);
}
} connection.Close();
return categories; }
catch (Exception e) {
System.Windows.Forms.MessageBox.Show(e.Message + "\r\nПомилка під час завантаження даних з бази");
return null; }
finally {
```

```
connection.Close(); }
}
private Category FindCategoryById(List<Category> list, String id) {
foreach(Category category in list) {
if (category.ID.Equals(id)) {
return category; }
}
return null;
}
private Document FindDocumentById(List<Document> list, String id) {
foreach (Document document in list) {
if (document.ID.Equals(id)) {
return document; }
}
return null;
}
public void SaveToDB(List<Category> categories) {
try
{
String query1 = "delete from class; delete from hierarchy;"; MySqlCommand command = new MySqlCommand(query1, connection); connection.Open();
command.ExecuteNonQuery(); }
catch (Exception e) {
System.Windows.Forms.MessageBox.Show(e.Message); }
finally {
connection.Close(); }
String newID = "1";
foreach (Category category in categories) {
try {
String query2 = "insert into class values (" + newID + ",'" + category.Name + "',"
+ "false" + ", " + "null); ";
MySqlCommand command2 = new MySqlCommand(query2, connection); connection.Open();
command2.ExecuteNonQuery(); category.ID = newID;
```

```
        newID = "" + (Int32.Parse(newID) + 1); }
        catch (Exception e) {
        System.Windows.Forms.MessageBox.Show(e.Message); }
        finally {
        connection.Close(); }
        }
        foreach (Category category in categories) {
        if (category.Supercategories.Count == 0) {
        AddRecursively(category); }
        } }
        private void AddRecursively(Category category) {
        int i = 0;
        foreach (Category subcategory in category.Subcategories) {
        try {
        String query2 = "insert into hierarchy values (" + random.Next(65545) + "," +
category.ID +
        ", " + subcategory.ID +"); ";
        MySqlCommand command2 = new MySqlCommand(query2, connection);
connection.Open();
        command2.ExecuteNonQuery(); AddRecursively(subcategory); i++;
        }
        catch (Exception e) {
        System.Windows.Forms.MessageBox.Show(e.Message); }
        finally {
        connection.Close(); }
        } }

} }
```

ДОДАТОК Б
(обовязковий)
Перелік наукових праць

КЛАСИФІКАЦІЯ ЗАГРОЗ ТА ВРАЗЛИВОСТЕЙ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ, ХАРАКТЕРНИХ ДЛЯ ІНТЕРНЕТ-РЕСУРСІВ

к.т.н. Джулій В.М. *(ХмНУ),*
Пахар О.В. *(ХмНУ)*

Ключовими елементами забезпечення захисту інформації є визначення, аналіз та класифікація загроз та вразливостей безпеки. На основі аналізу проведеного дослідження ризиків та формулювання вимог до систем захисту відносять: перелік існуючих загроз та вразливостей, оцінка ймовірностей реалізації загроз, модель порушника.

Більшість існуючих моделей інформаційної безпеки ґрунтуються на забезпеченні цілісності, доступності та конфіденційності інформації [1]. Вразливості інформаційних систем, як правило, є наслідком помилок. Помилки, що формують вразливості, поділяються на помилки адміністрування та помилки реалізації. До помилок реалізації відносять: помилки синхронізації - вид помилок, зумовлений існуванням тимчасових вікон між операціями обробки даних; помилки перевірки умов - нездатність програми обробити виняток, внаслідок некоректного визначення умови обробки даних; помилки перевірки вхідних даних - як правило, подібні помилки призводять до вразливостей переповнення буфера. До помилок адміністрування відносять: помилки конфігурування; помилки оточення. Прикладами помилок цього роду є помилки, пов'язані з некоректною обробкою змінних оточення, та помилки командного інтерпретатора.

Виявлення наведених помилок є безперервним процесом, що здійснюється на всіх етапах життєвого циклу інформаційної системи: розробки, тестування та експлуатації системи.

Як основні види загроз безпеці інформаційних систем та інформації виділяють [2]: аварії та стихійні лиха (повені, пожежі, землетруси, урагани, і т.д.); відмови та збої в роботі обладнання та технічних складових інформаційних систем; наслідки помилок проектування та розробки складових інформаційних систем (апаратних засобів, структур даних, технології обробки інформації, програм тощо); помилки експлуатації (операторів, користувачів та іншого персоналу); цілеспрямовані дії зловмисників та порушників.

В результаті аналізу даних тематичних інтернет-ресурсів зроблено висновок про те, що опис більшості загроз та вразливостей інформаційної безпеки може бути витягнуто з повідомлень тематичних користувачів інтернет-ресурсів, наприклад, хакерських форумів. Виняток становлять рідкісні, складні в реалізації загрози та вразливості, що вимагають експертних знань чи спеціалізованого устаткування.

ЛІТЕРАТУРА:

1. Бурячок В.Л. Інформаційний та кіберпростори: проблеми безпеки, методи та засоби боротьби: підручник. /В.Л. Бурячок, Г.М. Гулак, В.Б.Толубко - Київ: ТОВ «СІК ГРУП Україна», 2015. - 449 с.

2. Дудикевич В.Б. Забезпечення інформаційної безпеки держави: навч. посіб. / В.Б. Дудикевич, І.Р. Опірський, П.І. Гаранюк - Львів: Видавництво Львівської політехніки, 2017. 204 с.

УДК 004.056:621.397.3:004.942

Володимир ДЖУЛІЙ
Хмельницький національний університет
ORCID http://orcid.org/0000-0003-1878-4301
e-mail: dg2303@ukr.net
Наталія ПЕТЛЯК
Хмельницький національний університет
ORCID http://orcid.org/0000-0001-5971-4428
e-mail: npetlyak@khmnu.edu.ua
Юрій ХМЕЛЬНИЦЬКИЙ
Хмельницький національний університет
ORCID http://orcid.org/0000-0002-4005-5669
e-mail: getman-58@ukr.net,
Олександр ПАХАР
Хмельницький національний університет
e-mail:dg2303@ukr.net

МОДЕЛЬ ПОТОКУ ТЕКСТОВИХ ПОВІДОМЛЕНЬ ТЕМАТИЧНИХ ІНТЕРНЕТ-РЕСУРСІВ СИСТЕМИ ПРОГНОЗУВАННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Розглянута актуальна задача побудови моделі потоку текстових повідомлень тематичних інтернет-ресурсів системи прогнозування інформаційної безпеки та інформаційної модель бази даних форуму тематичного інтернет-ресурсу, відрізняється від відомих, універсальністю, дозволяє аналізувати та досліджувати потік даних інтернет-форумів, реалізованих на базі популярних програмних платформ для розробки дискусійних інформаційних тематичних ресурсів. Представлена модель потоку текстових повідомлень тематичних інтернет форумів, відрізняється від відомих, можливістю проводити статистичний аналіз та семантичну фільтрацію повідомлень, враховуючи належність до автора, рейтингу автора, форуму, часу створення, кількості повідомлень, темі форуму, дозволяє здійснювати аналіз та дослідження текстових повідомлень тематичних інтернет- ресурсів.

Ключові слова: моделі, алгоритми, модель потоку повідомлень, тематичні форуми, джерела повідомлень.

Volodymyr DZHULIY, Natalia PETLYAK,
Yuri Khmelnytskyi, Oleksandr PAKHAR
Khmelnytsky National University

TEXT MESSAGE FLOW MODEL OF THEMATIC INTERNET RESOURCES INFORMATION SECURITY FORECASTING SYSTEMS

Abstract. The tasks of analysis and classification of detection of existing mechanisms for the implementation of attacks and threats to information security, which can lead to unauthorized access to confidential information, disruption of the functioning of information systems, are relevant and prioritized at the current stage. The importance of the problems is connected with the following main factors: growth of variety and quantity of means of computer technology and spheres of human activity; a high level of trust in information and search systems for data processing and management; the growth of the number of users of the information space of interaction; accumulation of large volumes of various types of information, intensive exchange of data flow in the network between users, using a wide range of access mechanisms to confidential resources, information processes; industrial espionage and competitive struggle in the sphere of information services of society; insufficient number at the present stage of highly qualified specialists in the field of information security, market relations in the field of software development, maintenance, distribution, production of computing equipment for the implementation of information security.

The presented model of the flow of text messages of thematic Internet resources of the information security forecasting system and the information model of the forum database of the thematic Internet resource, which differs from

the known ones in its universality, allows to analyze and study the data flow of Internet forums implemented on the basis of popular software platforms for the development of discussion information thematic resources The model of the flow of text messages of thematic Internet forums allows statistical analysis and semantic filtering of messages, taking into account authorship, author rating, forum, time of creation, number of messages, forum topic, allows analysis and research of text messages of thematic Internet resources.

Solving the set tasks will allow: to improve the quality of decisions made in the process of identifying and counteracting malicious information; sort information objects of influence for the operator by priority; set the input data settings of the system of detection and countermeasures against the spread of malicious information in networks.

Keywords: models, algorithms, message flow model, thematic forums, message sources.

Постановка проблеми у загальному вигляді
та її зв'язок із важливими науковими чи практичними завданнями

На сучасному етапі, на більшість сфер діяльності суспільства зростає вплив глобальних інформаційних технологій. Відзначаються високі темпи розвитку світових єдиних телекомунікаційного та інформаційного просторів, сформувалися в суспільстві нові соціальні групи, виявляється значний вплив на сформований історично спосіб життя людей. На тлі стрімкого розвитку інформаційних технологій відзначається, зростання активності різноманітності комп'ютерних атак, здійснюваних і запланованих із застосуванням сучасних новітніх технологій [1].

Актуальними та пріоритетними на сучасному етапі є задачі класифікації та аналізу виявлення існуючих механізмів реалізації атак та загроз інформаційної безпеки, які можуть призвести до отримання несанкціонованого доступу до конфіденційної інформації, порушення функціонування інформаційних систем. В результаті, постає задача визначення заходів протидії атакам та загрозам, усунення вразливостей, оцінки заданої можливої шкоди, розробка нормативно-правової бази, механізмів захисту та критеріїв безпеки [2 - 4].

Важливість проблем пов'язана з наступними основними факторами: зростанням різноманітності та кількості засобів комп'ютерної техніки та сфер людської діяльності їх застосування; високим рівнем довіри до інформаційно-пошукових систем обробки та управління даними; зростанням числа користувачів інформаційного простору взаємодії; накопиченням великих об'ємів різнотипної інформації, інтенсивним обміном потоком даних в мережі між користувачами, з використанням широкого спектра механізмів доступу до конфіденційних ресурсів, інформаційних процесів; промисловим шпигунством та конкурентною боротьбою у сфері інформаційних послуг суспільства; недостатньою кількістю, на сучасному етапі, фахівців високої кваліфікації в області інформаційної безпеки, ринковими відношеннями в області розробки програмного забезпечення, обслуговування, розповсюдження, виробництва обчислювальної комп'ютерної техніки для реалізації інформаційної безпеки [5,6].

Постановка задачі

На сучасному етапі, проблеми інформаційної безпеки розвитку суспільства у більшості сфер їх діяльності виходять на передній план. Це пов'язано зі значним зростанням кількості реалізованих проектів інформатизації. Більшість проектів інформатизації спрямовані на побудову єдиного телекомунікаційного та інформаційного простору з метою оптимізації процесів обробки різнотипної інформації великих об'ємів, забезпечення оперативного доступу до інформації, надійного зберігання даних для користувачів інформаційного обміну. На даному етапі виникає потреба у проведенні захисту комп'ютерних систем та інформаційних ресурсів від блокування, несанкціонованого доступу до даних, знищення, та інших злочинних, небажаних дій, кількість яких постійно зростає, збитки в інформаційній сфері від злочинів в мережі Інтернет щорічно оцінюються в мільярди доларів [7 - 9].

Темпи розвитку інформаційних та комп'ютерних технологій значно випереджають процес створення програмно-апаратного забезпечення в області інформаційної безпеки. Пріоритетними, в даній ситуації, є задачі проведення аналізу, класифікації, виявлення діючих механізмів та засобів проведення атак і загроз інформаційній безпеці системи, які можуть призвести до отримання несанкціонованого доступу до конфіденційних даних,

порушення функціонування інформаційної системи, визначення заходів протидії атакам та загрозам, оцінка заданої шкоди, розробка нормативно-правової бази, механізмів захисту та критеріїв інформаційної безпеки системи протидії.

На сьогодні не існує єдиного підходу до вирішення проблеми захищенності інформаційних систем, стосовно предметних областей, розробниками програмного захисту інформації пропонуються відповідні компоненти на вирішення конкретних задач, забезпечення надійного захисту інформаційних ресурсів потребує реалізації відповідних організаційних та технічних заходів в комплексі, що супроводжуються розробкою відповідної документації [1,2]. Проведене дослідження вказує на необхідність вирішення наступних задач для забезпечення інформаційної безпеки: формування основ для опису процесів виникнення та реалізації загроз інформаційної безпеки системи в умовах невизначеності та непередбачуваності їх прояву; розробка відповідних засобів забезпечення захисту конфіденційної інформації на основі проведеного дослідження та класифікації вразливостей та загроз; визначення загальних підходів до створення інформаційно-аналітичних систем забезпечення захисту конфіденційних даних, механізмів управління захистом на різних рівнях діяльності суспільства [2 - 4].

Вирішення поставлених задач дозволить: підвищити якість прийнятих рішень у процесі виявлення та протидії шкідливій інформації; сортувати інформаційні об'єкти впливу для оператора по пріоритету; задати вхідні дані налаштування системи виявлення та протидії поширенню шкідливої інформації в мережах.

Основна частина

При створенні форумів інтернет-ресурсів найбільшою популярністю користуються настпні програмні платформи: Vanilla, Invision Power Board, vBulletin, PunBB, Simple Machines Forum, XenForo. При реалізації програмних платформ, використовуються бази даних структури, яких значно різняться. Записи даних дозволяють для текстових повідомлень визначати їхню приналежність до конкретного інтернет - форуму, автора, рейтингу автора, часу створення, а також кількості повідомлень відповідної теми форуму. Враховуючи наведену інформацію, запропоновано модель інтернет-форума (рис. 1).

Модель тематичного інтернет-ресурсу відрізняється від існуючих, універсальністю, що надає можливість проводити аналіз та досліджувати повідомлення інтернет-форумів, реалізованих на базі найбільш популярних програмних платформ для створення дискусійних тематичних інформаційних форумів. Застосування на практиці моделі інтернет-форуму, при прогнозуванні вразливостей та загроз інформаційної безпеки, надасть можливість проводити аналіз більшості існуючих тематичних інтернет-форумів, незалежно від задіяної програмної платформи, використовуваної для реалізації форуму [1,2,7].

Текстові повідомлення інтернет мережі є окремо структурою, що складається з пов'язаних між собою елементів (рис. 2). Потік даних в інтернет- мережі є множина текстових повідомлень тематичних інтернет-форумів, створюваних учасниками форуму. Враховуючи те, що моделювання мережевого потоку повідомлень тематичних інтернет-форумів здійснюється з метою наступного прогнозування вразливостей та загроз інформаційної безпеки, при побудові моделі тематичного інтернет- форуму, необхідно передбачити можливість здійснення семантичного та статистичного аналізу текстових повідомлень, враховуючи належність даних до конкретного форуму, кількості повідомлень теми форуму, автора, часу створення, рейтингу автора.

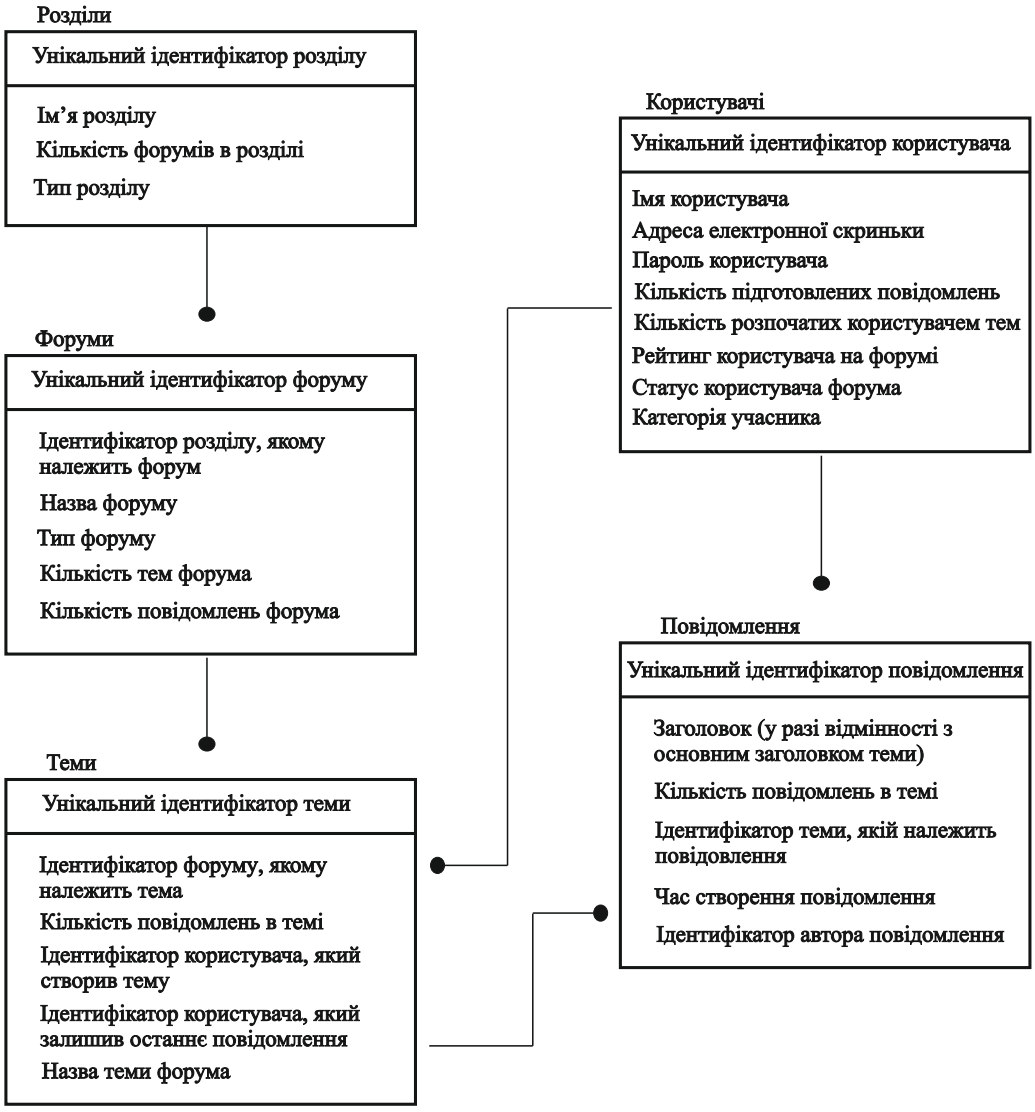


Рис. 1. Логічна модель бази даних тематичних інтернет-ресурсів

Онтологія є найбільш ефективним інструментом для опису конкретної предметної області. Суть онтологічного підходу полягає у представленні предметної області у вигляді організованої множини понять, враховуючи існуючі зв'язки між ними та їх властивості [3, 10]. Модель потоку текстових повідомлень, у загальному вигляді, яка має відношення до предметної області, заданої онтологією, представляється трійкою (1):

$$S_\tau = (M, O, T), \tag{1}$$

де S_τ – потік текстових повідомлень інтернет мережі у поточний час τ; M - множина інтернет повідомлень у потоці даних; O - онтологія даної предметної області; $T = \{1, ..., \tau\}$ - множина часових інтервалів, в які велися спостереження за інтернет потоком повідомлень (годин, днів, місяців).

$$O = (E, R, F), \tag{2}$$

де E - множина термінів заданої предметної області; R - множина взаємозв'язків між термінами даної предметної області; F - множина заданих на відношеннях та термінах онтології функцій аксіоматизації (інтерпретації). Кожне текстове повідомлення $d \in M$ може бути представлене наступним чином:

$$d \in (s, t, F_d, A), \tag{3}$$

де S - текст інтернет повідомлення, $t = \{1,...,\tau\}$ - момент часу створення текстового повідомлення; $F_d = \{w_1,...,w_k\}$ – вектор, що представляє текстове повідомлення заданої предметної області, заданою відповідною онтологією O, k – кількість в онтології термінів, координати $w_i\,(i = 1,...,k)$ – ваги термінів у текстовому повідомленні, A – рейтинг автора текстового повідомлення.

При розрахунку ваги термінів використовується модель TF–IDF (Term frequency – Inverse document frequency) [10], відповідно до якої вага терміна текстового повідомлення прямо пропорційна частоті входження терміна в інтернет повідомлення і обернено пропорційна кількості текстових повідомлень, у яких зустрічається термін (4):

$$w_i = F_i \cdot \log\left(\frac{D}{DF_i}\right), \tag{4}$$

де w_i – вага i - терміна у текстовому повідомленні; F_i – частота i - терміна у інтернет повідомленні; D – загальна кількість текстових повідомлень; DF_i – кількість текстових повідомлень, у яких зустрічається i - термін.

Рис. 2. Структура повідомлення тематичного інтернет-форуму

Розглянута модель не враховує того, що текстові повідомлення тематичних форумів можуть мати різний розмір, у зв'язку з чим, вага терміна і відповідно частота будуть зменшуватися зі зростанням розміру текстових повідомлень. Враховуючи дану ситуацію необхідно проводити нормування ваг термінів у інтернет повідомленні, діленням їх на довжину вектора-повідомлення (еклідову норму) (5):

$$w_i^* = \frac{w_i}{\overline{d}} = \frac{w_i}{\sqrt{\sum_{i=1}^{k} w_i}}, \tag{5}$$

При використанні онтологічних методів для опису та обробки предметної області, необхідно представити її у вигляді організованої структури сукупності термінів (понять), враховуючи існуючі властивості та зв'язки між ними. У задачах, що передбачають подальшу обробку, предметної області, розроблених онтологій, найчастіше застосовується для їх представлення формат OWL (Ontology Web Language) – мова опису онтологій для семантичного павутиння. У термінознавстві та лексикографії застосовуються алгоритми, що ґрунтуються на статистичних та лінгвістичних методах, для видалення термінів [10].

При використанні статистичних методів основним критерієм є ступінь термінологічності, визначається у відповідності до числових закономірностей, характерними для термінів і нетермінів. При використанні лінгвістичних методів терміни відбираються за лінгвістичними ознаками та певними граматичними лексичними шаблонами [6,10].

При використанні онтологічного підходу знання про відповідні предметні області (онтологія) зберігаються у вигляді (6):

$$O=(E,R,F), \tag{6}$$

де E - множина термінів заданої предметної області; R - множина взаємозв'язків між термінами даної предметної області (7):

$$R\subset\{R_{inc},R_{add},R_{term},R_{lem},R_{NC}\}, \tag{7}$$

де R_{inc} – множина вбудованих відношень між об'єктами («є Підкласом»); R_{add} – множина вбудованих відношень між об'єктами, дозволяють розширювати набір відповідних об'єктів предметної області, для якої проводиться аналіз, шляхом об'єднання лем області, пов'язаних між собою об'єктів («має Відношення», «є Частиною»); R_{term} – відношення «є Терміном», визначається експертним шляхом і носить допоміжний характер, приймає логічний тип значення (в залежності від того, наскільки об'єкт характерний для предметної області). Прикладне застосування R_{term} знаходить при вирішенні задач видалення, з використанням тезаурусного критерію термінологічності термінів; R_{lem} – відношення «має Лемму», дана властивість приймає значення рядкового типу, яке виходить в результаті лемування, приведенні найменування об'єкта до початкової форми; R_{NC} – множина відношень між тематичними об'єктами, описують особливості взаємодії між собою об'єктів даної предметної області (властивості «є Елементом», «є Типом вірусів»); F – множина заданих на термінах та відношеннях онтології предметної області функцій аксіоматизації (інтерпретації).

Задача фільтрації текстових повідомлень, які не належать до предметної області, для якої проводиться аналіз, може бути вирішена із використанням семантичної метрики «термін/не термін». Для використання семантичної метрики необхідно попередньо розробити відповідну онтологію предметної області у форматі OWL. Далі для кожного тематичного повідомлення, що надходять в інтернет мережі, розраховується значення ступеня близькості термінів включених до онтології, в результаті виділяються тематичні повідомлення, що виключно відносяться до даної предметної області [7 – 9].

Значення коефіцієнта ступеня близькості текстового повідомлення до всіх термінів предметної області, який розраховується при використанні семантичної метрики «термін/ не термін», приймає значення в діапазоні від 0 до 1 (текстове повідомлення відноситься до певного терміну, чим ближче значення коефіцієнта k_{Ont} до 1). Для вирішення задач відбору текстових повідомлень, до даної предметної області, заданої у вигляді онтології, використовуються два критерії: вкладених зв'язків та тезаурусний критерій. Застосування тезаурусного підходу до вирішення задачі фільтрації текстових повідомлень полягає в пошуку лем, що містяться в тематичних повідомленнях, надходять в інтернет мережі, серед термінів онтології даної предметної області. Використання тезаурусного підходу для відповідного класу онтології предметної області визначає властивість «має Лемму», шляхом приведення до початкової форми (лемування) найменування об'єкта області, що аналізується.

Для розрахунку коефіцієта ступеня близькості тематичних повідомлень до термінів даної предметної області відповідно до тезаурусного критерію, необхідно виконати послідовність дій: провести оцінку коефіцієнта ступеня близькості текстового повідомлення до кожного об'єкта онтології, для якої проводиться аналіз; визначити опорний об'єкт заданої онтології, який найбільш близько асоціюється з тематичним повідомленням, що надходить в інтернет мережі.

Розрахунок коефіцієнта ступеня близькості текстового повідомлення термінам заданої предметної області з використанням тезаурусного критерію представлено на рис. 3.

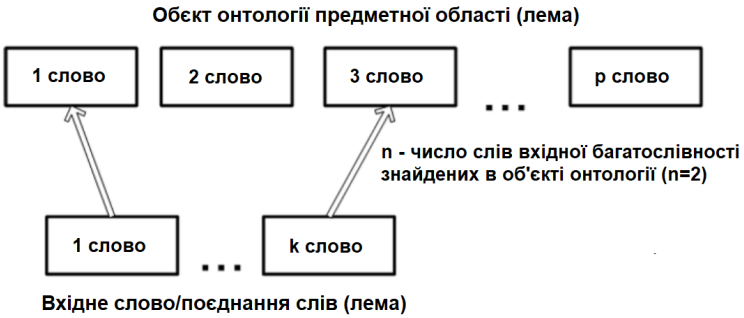


Рис. 3. Визначення опорного об'єкта онтології

Коефіцієнт ступеня близькості опорного об'єкта онтології до тематичного повідомлення, що надходить в інтернет мережі розраховується наступним чином (8):

$$k_t = \max\left(\frac{n_i}{p_i}\right)_{i=1}^{m}, \tag{8}$$

де m – загальна кількість об'єктів заданої онтології; n_i – кількість слів у лемі тематичного повідомлення, присутніх у лемі i–го об'єкта заданої онтології предметної області; p_i – кількість слів у лемі i–го об'єкта онтології предметної області.

У випадку коли значення коефіцієнта k_t отримано однакове для декількох об'єктів онтології заданої предметної області, опорним приймається об'єкт онтології, для якого значення n_i набуває максимальної величини. Якщо, при цьому, існує декілька об'єктів онтології, для яких значення n_i і k_t однакові, то в даній ситуації всі об'єкти вважаються опорними і для кожного об'єкта проводиться аналіз по онтологічного критерію.

Відповідно до тезаурусного критерію коефіцієнт ступіня близькості тематичного повідомлення термінам даної предметної області розраховується за наступним чином (9):

$$k_{Ont} = \frac{k_t}{c+1}, \tag{9}$$

де k_t - коефіцієнт ступеня близькості, розрахований на першому етапі проведення аналізу (обчислюється за (8)); c - число відношень між об'єктами, які пов'язують опорний об'єкт онтології предметної області з об'єктами, що мають значення властивості «є Терміном». Якщо опорний об'єкт онтології предметної області є терміном, то c=0. Тезаурусний критерій схематично представлений на рис. 4.

Застосування метрики "термін / не термін" для коефіцієнта оцінки ступеня близькості тематичних повідомлень до термінів заданої предметної області передбачається рух графом, при цьому об'єкти класів онтології предметної області є вузлами графа. Якщо у опорного об'єкта онтології предметної області властивість «є Терміном» хибно, і він при цьому не пов'язаний з іншими об'єктами, або у всіх пов'язаних з ним об'єктів онтології значення властивості «є
Терміном» хибно, то в даному випадку проводиться пошук інших опорних об'єктів онтології і знову ж таки проводиться оцінка. При цьому тематичне повідомлення не відноситься до предметної області ($k_{Ont} = 0$), коли опорні об'єкти онтології відсутні або для всіх опорних об'єктів предметної області характерна розглянута ситуація.

Критерій вкладених зв'язків онтології ґрунтується на тому, що крім оцінки коефіцієнта ступеня термінологічності кожного текстового повідомлення, метрика «термін/не термін» дозволяє проводити фільтрацію тематичних повідомлень шляхом зіставлення леми повідомлення та поєднаннями лем об'єктів онтології предметної області, пов'язаних відношеннями.

Рис. 4. Схема тезаурусного критерію

Таким чином, тематичне повідомлення вважається таким, що відноситься до заданої предметної області, якщо його лема співпадає з об'єднанням лем об'єктів даної онтології, пов'язаних між собою, при цьому односпрямованими відношеннями.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Запропонована інформаційна модель бази даних форуму тематичного інтернет-ресурсу, відрізняється від відомих, універсальністю, дозволяє аналізувати та досліджувати потік даних інтернет-форумів, реалізованих на базі популярних програмних платформ для розробки дискусійних інформаційних тематичних ресурсів.

Запропонована модель потоку текстових повідомлень тематичних інтернет форумів, які відносяться до заданої предметної області, даної онтології, відрізняється від відомих, можливістю проводити статистичний аналіз та семантичну фільтрацію повідомлень, враховуючи належність до автора, рейтингу автора, форуму, часу створення, кількості повідомлень, темі форуму, дозволяє здійснювати аналіз та дослідження текстових повідомлень тематичних інтернет-ресурсів. Особливість запропонованого підходу пов'язана з тим, що об'єкти онтології необхідно представляти переважно однослів'ями, що мають, в даному випадку максимальну кількість відношень із іншими об'єктами онтології. Для практичного використання методу визначальними є відношення, що дозволяють формувати природним чином словосполучення.

Літертура

1. Ленков, С.В. Модель безпеки поширення забороненої інформації в інформаційно-телекомунікаційних мережах / С.В. Ленков, В.М. Джулій, В.С. Орленко, О.В. Сєлюков, А.В. Атаманюк // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2020. – Вип. №68. – С. 53-64.

2. Джулій В.М. Інформаційно-ознакова модель джерела шкідливої інформації в соціальних мережах/ В.М. Джулій І.В. Муляр О.О Зацепіна В. М. Пічура – Вимі-рювальна та обчислювальна техніка в технологічних процесах № 3 (2022)- 73 – 78

3. Ленков, С.В. Методы и средства защиты информации. В 2-х томах /С.В. Ленков, Д.А. Перегудов, В.А. Хорошко –К: Арий, 2008.–464с

4. Остапов С. Е. Технології захисту інформації: навчальний посібник / С.Е. Остапов, С.П. Євсеєв, О.Г. Король–Харків : Вид-во ХНЕУ, 2016. – 476 с.

5. Лєнков, С.В. Аналіз існуючих методів та алгоритмів виявлення атак в бездротових мережах передачі даних / С.В. Лєнков, В.М. Джулій, Н.М. Берназ, С.О. Божук // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2017. – Вип. № 56. – С.124-132

6. Довгий, С.О. Сучасні телекомунікації: мережі, технології, економіка, управління, регулювання / С.О. Довгий, О.Я. Савченко, П.П. Воробієнко – К.: Український Видатничій Центр, 2012. – 520 с.

7. Джулій, В.М. Модель оцінки ймовірнісно-часових характеристик інформаційної взаємодії в мережі інтернет речей / В.М. Джулій, І.В. Муляр, О.В. Сєлюков, Б.М. Кізюн // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2019. – Вип. № 63. – С.96-106

8. Джулій, В.М., Кльоц Ю.П., Муляр І.В., Жилевич М.Л., Джулій А.В. Контроль додатків інтернет-трафіка комп'ютерних мереж методами машинного навчання. Вісник Хмельницького національного університету. Технічні науки. 2021. № 5. С. 22-26.

9. Джулій, В.М. Метод класифікації додатків трафіка комп'ютерних мереж на основі машинного навчання в умовах невизначеності / В.М. Джулій, О.В. Мірошніченко, Л.В. Солодєєва // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2022. – Вип. №74. – С. 73-82.

10. Лавров, Є. А. Математичні методи дослідження операцій: підручник / Є. А. Лавров, Л. П. Перхун, В. В. Шендрик – Суми : Сумський державний університет, 2017. – 212 с.

References

1. Lenkov, S.V. (2020), Model bezpeky poshyrennia zaboronenoi informatsii v informatsiino-telekomunikatsiinykh merezhakh / S.V. Lenkov, V.M. Dzhulii, V.S. ORLENKO, O.V. Sieliukov, A.V. Atamaniuk // Zbirnyk naukovykh prats Viiskovoho instytutu Kyivskoho natsionalnoho universytetu imeni Tarasa Shevchenka. – K.: VIKNU. – №68. – pp. 53-64.

2. Dzhulii V.M. Informatsiino-oznakova model dzherela shkidlyvoi informatsii v sotsialnykh merezhakh/ V.M. Dzhulii I.V. Muliar O.O Zatsepina V. M. Pichura – Vymi-riuvalna ta obchysliuvalna tekhnika v tekhnolohichnykh protsesakh № 3 (2022)- 73 – 78

3. Lenkov, S.V. (2008), Metodyy sredstva zashchytы ynformatsyy. V 2-kh tomakh / S.V. Lenkov, D.A. Perehudov, V.A. Khoroshko –K: Aryi–464s.

4. Ostapov, S. E. (2016) Tekhnolohii zakhystu informatsii: navchalnyi posibnyk / S.E. Ostapov, S.P. Yevseiev, O.H. Korol– Kharkiv : Vyd-vo KhNEU. – 476 s.

5. Lenkov, S.V. (2017), AnalIz Isnuyuchih metodiv ta algoritmiv viyavlennya atak v bezdrotovih merezhah peredachI danih / S.V. Lenkov, V.M. Dzhuliy, N.M. Bernaz, S.O. Bozhuk // Zbirnik naukovih prats Viyskovogo Institutu Kiyivskogo natsionalnogo universitetu imeni Tarasa Shevchenka. – K.: VIKNU. – Vip. No 56. – p.124-132

6. Dovhyi, S.O. (2012), Suchasni telekomunikatsii: merezhi, tekhnolohii, ekonomika, upravlinnia, rehuliuvannia /S.O. Dovhyi, O.I. Savchenko, P.P. Vorobiienko – K.: Ukrainskyi Vydatnychii Tsentr. – 520p.

7. Dzhulii, V.M. (2019), Model otsinky ymovirnisno-chasovykh kharakterystyk informatsiinoi vzaiemodii v merezhi internet rechei / V.M. Dzhulii, I.V. Muliar, O.V. Sieliukov, B.M. Kiziun // Zbirnyk naukovykh prats Viiskovoho instytutu Kyivskoho natsionalnoho universytetu imeni Tarasa Shevchenka. – K.: VIKNU. – Vyp. № 63. – p.96-106

8. Dzhulii V.M., Klots Yu.P., Muliar I.V., Zhylevych M.L., Dzhulii A.V. (2021), Kontrol dodatkiv internet-trafika kompiuternykh merezh metodamy mashynnoho navchannia. Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky. – Khmelnytskyi. – №5. – pp. 22–26.

9. Dzhulii, V.M. (2022), Metod klasyfikatsii dodatkiv trafika kompiuternykh merezh na osnovi mashynnoho navchannia v umovakh nevyznachenosti / V.M. Dzhulii, O.V. Miroshnichenko, L.V. Solodieieva // Zbirnyk naukovykh prats Viiskovoho instytutu Kyivskoho natsionalnoho universytetu imeni Tarasa Shevchenka. – K.: VIKNU. – Vyp. №74. – pp. 73-82.

10. Lavrov, Ye. A. (2017.), Matematychni metody doslidzhennia operatsii : pidruchnyk / Ye. A. Lavrov, L. P. Perkhun, V. V. Shendryk – Sumy : Sumskyi derzhavnyi universytet, – 212 p.

ДОДАТОК В
Презентація

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ПАХАР Олександр Валерійович

Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів

Науковий керівник
д.т.н., доцент Касянчук М.М.

кафедра кібербезпеки

Тема Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів

Мета магістерської роботи - підвищення ефективності методів та засобів виявлення вразливостей та загроз безпеки інформації на основі запропонованих алгоритмів, методів та інформаційно-аналітичної системи аналізу потоку повідомлень інтернет-ресурсів.

Наукова задача – розробка моделей, методу прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів, заснований на нечіткому логічному виводу, семантичному та статистичному аналізу даних.

Об'єкт дослідження: Потоки повідомлень тематичних інтернет-ресурсів, що містять відомості про вразливості та загрози безпеки інформації.

Предмет дослідження: Алгоритми, методи прогнозування вразливостей, загроз безпеки інформації, підходів до їх реалізації з використанням автоматизованого аналізу даних інтернет-ресурсів..

Задачі досліджень у роботі формулюються наступним чином:

1. Аналіз сучасних вразливостей та загроз інформаційної безпеки та засобів захисту інформації.
2. Розробка моделі бази даних інтернет-ресурсів, потоку даних тематичних ресурсів з метою виявлення вразливостей, загроз безпеки інформації.
3. Розробка алгоритму прогнозування вразливостей, загроз інформаційної безпеки, на основі отриманих результатів обробки інформації тематичних ресурсів та алгоритму фільтрації потоку текстових повідомлень інтернет-ресурсів.
4. Розробка інформаційно-аналітичної системи автоматизації проведення аналізу потоку даних тематичних інтернет-ресурсів та прогнозування появи нових вразливостей, загроз безпеки інформації з використанням нечіткого логічного виводу.

Наукова новизна роботи визначає:

1. Модель потоку повідомлень та бази даних, тематичного ресурсу, призначена для прогнозування вразливостей та загроз безпеки інформації, відрізняється можливістю опрацьовувати різнотипні дані, що застосовуються для організації дискусійних інформаційних тематичних ресурсів, а також можливістю статистичного аналізу та семантичної фільтрації повідомлень, дозволяє прогнозувати вразливості та загрози, враховуючи, при цьому, їхню тематичну приналежність.

2. Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів, заснований на нечіткому логічному виводу, семантичному та статистичному аналізі, відрізняється можливістю виявлення вразливостей та загроз до їх реалізації, дозволяє описувати закономірності інформаційного процесу наповнення тематичних ресурсів новими текстовими повідомленнями, що відображається на якості прогнозування.

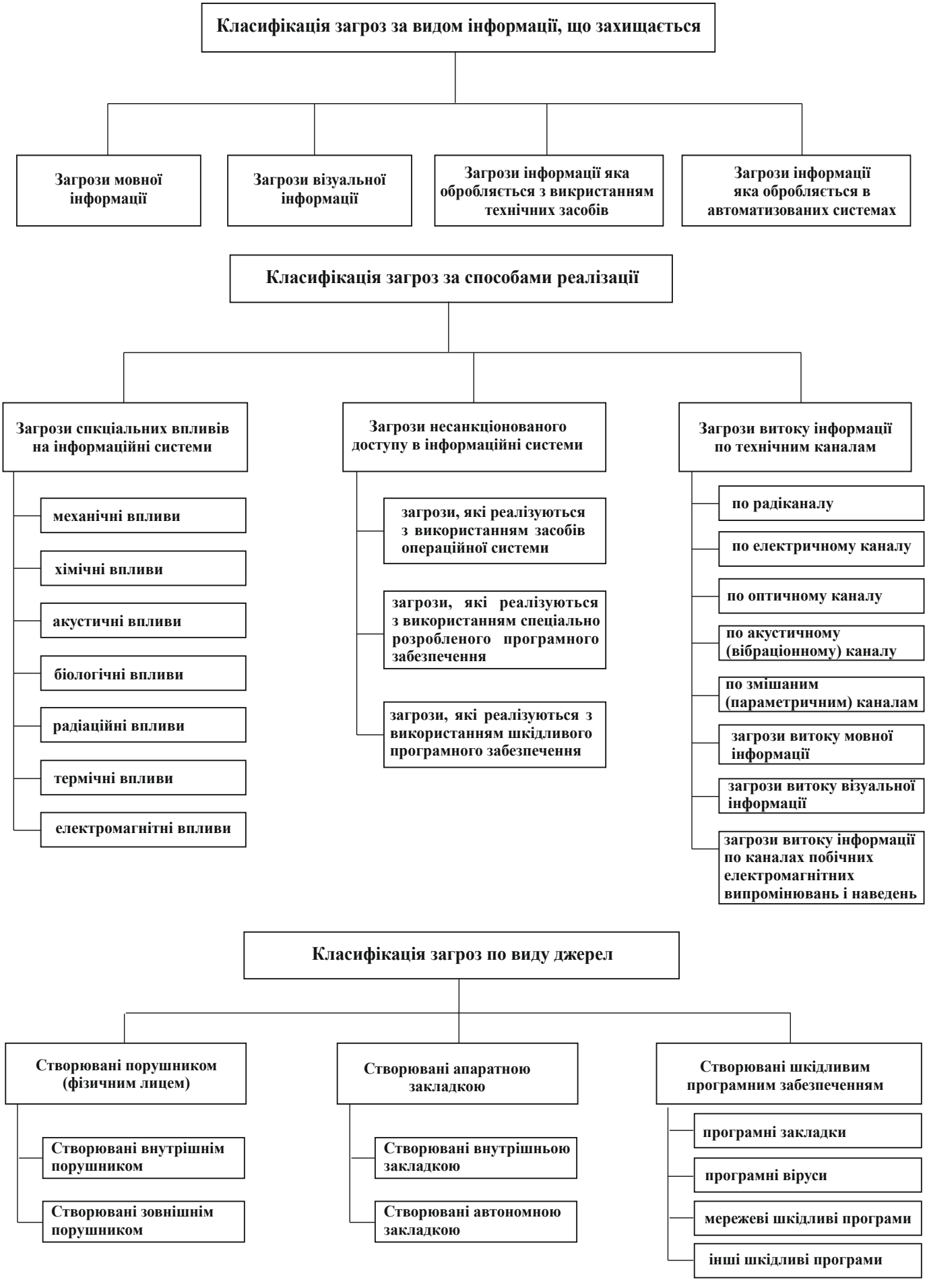
Методи дослідження. Для вирішення задач у магістерській роботі застосовувалися методи: логічного виводу, системного аналізу, пізнання та пошуку, теорії нечітких множин, інформаційного та функціонального моделювання, об'єктно-орієнтованого програмування, математичної статистики та логіки, семантичного аналізу тексту.

Практична цінність Реалізований в інформаційно-аналітичній системі алгоритм прогнозування вразливостей та загроз безпеки інформації на основі аналізу потоку даних тематичних ресурсів дозволяє автоматизувати інформаційний процес виявлення нових вразливостей, загроз, надає фахівцям інформаційної безпеки можливість оцінити своєчасно ступінь захищеності ресурсів та при необхідності вжити відповідних заходів щодо нейтралізації можливих загроз та вразливостей, тим самим підвищити інформаційну безпеку обчислювальних комп'ютерних систем від реалізації нових мережевих комп'ютерних атак.

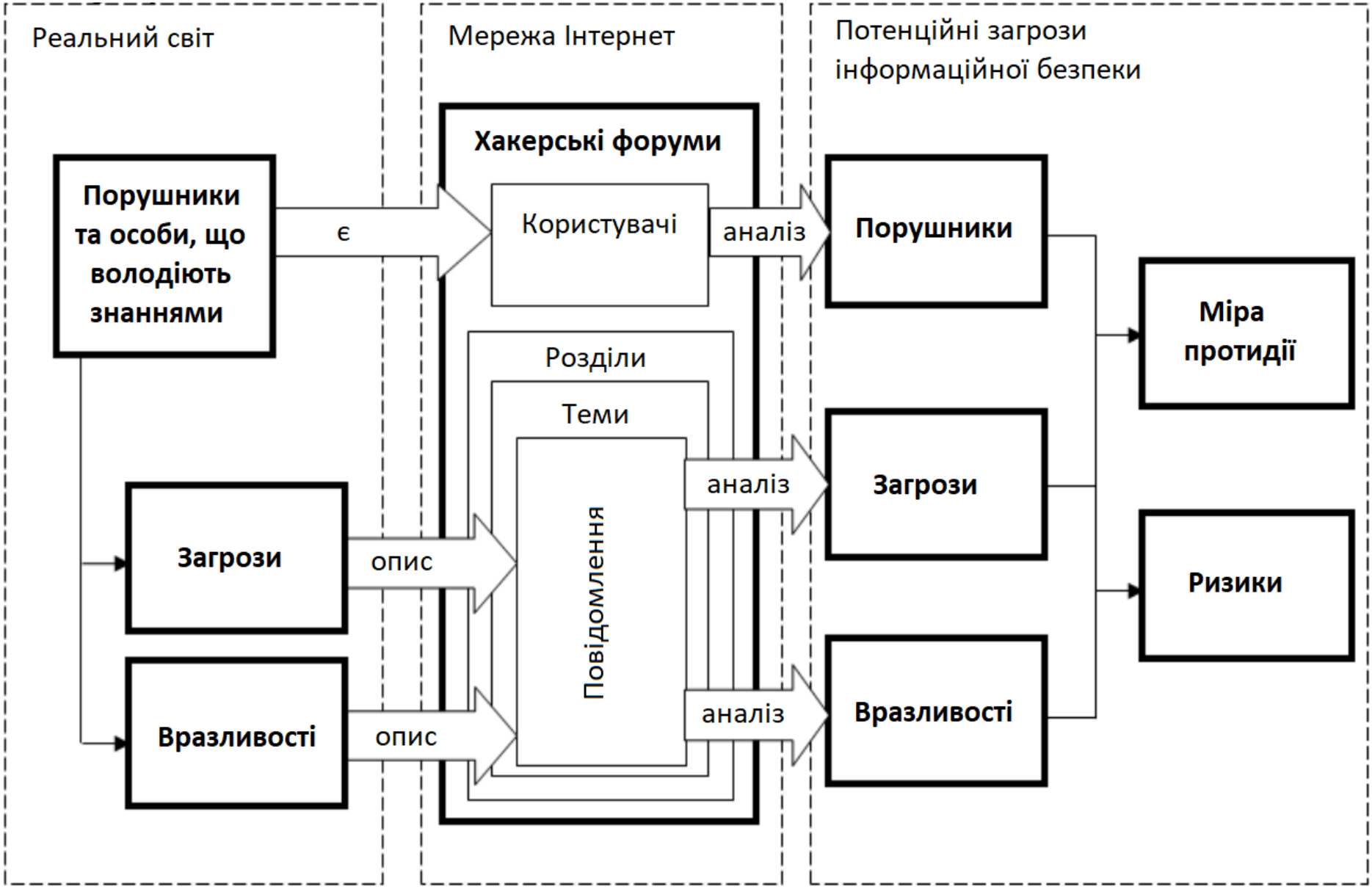
Апробація роботи. Наукові результати і основні положення магістерської роботи доповідались і обговорювались на всеукраїнських та міжнародних науково-технічних конференціях,

Публікації. За темою дипломної роботи ОКР «Магістр» опубліковано 1 теза доповідей, 1 фахова стаття.

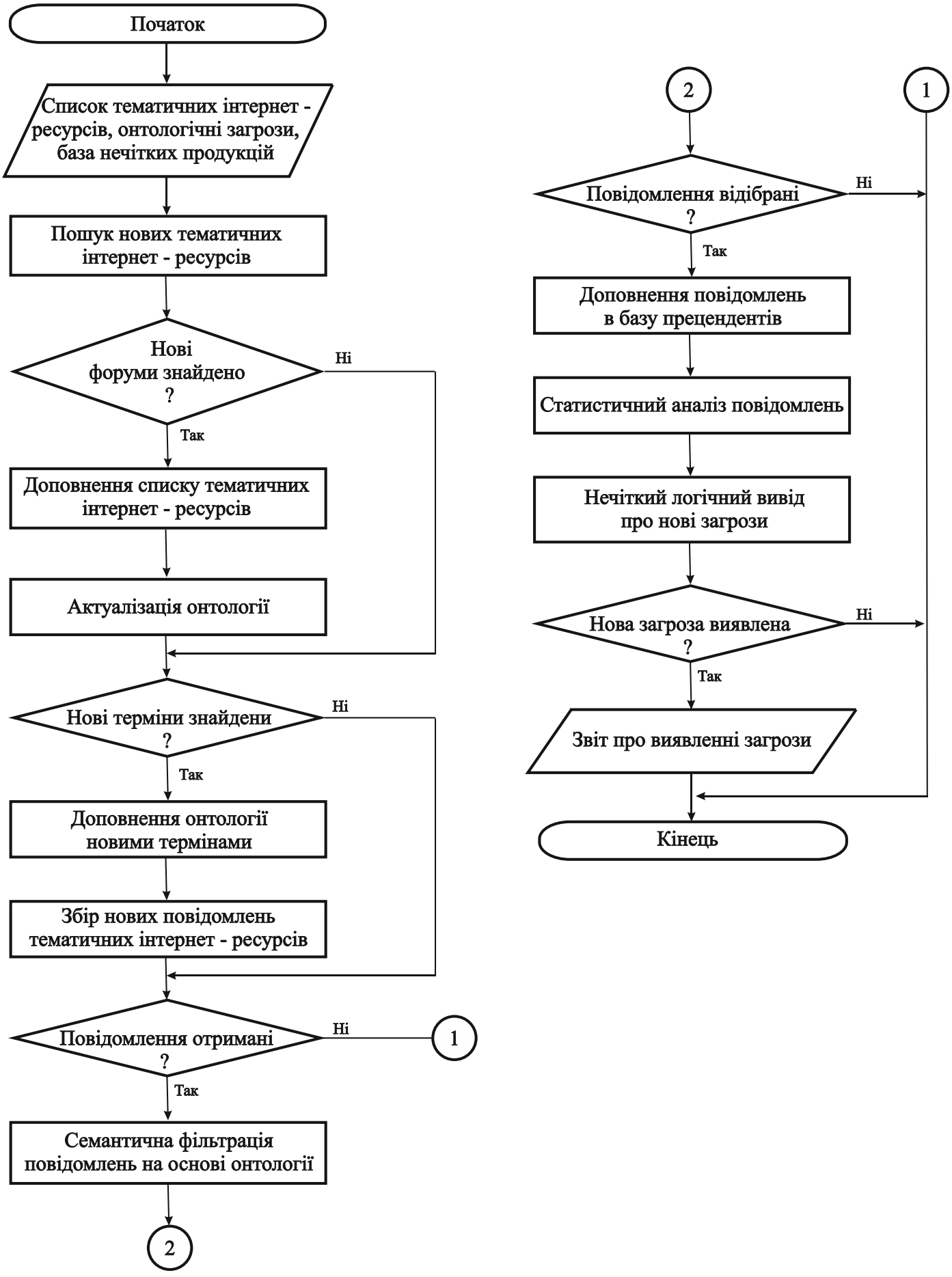
Класифікація вразливостей та загроз інформаційної безпеки, характерних для інтернет-ресурсів



Інформаційне наповнення тематичних інтернет-ресурсів

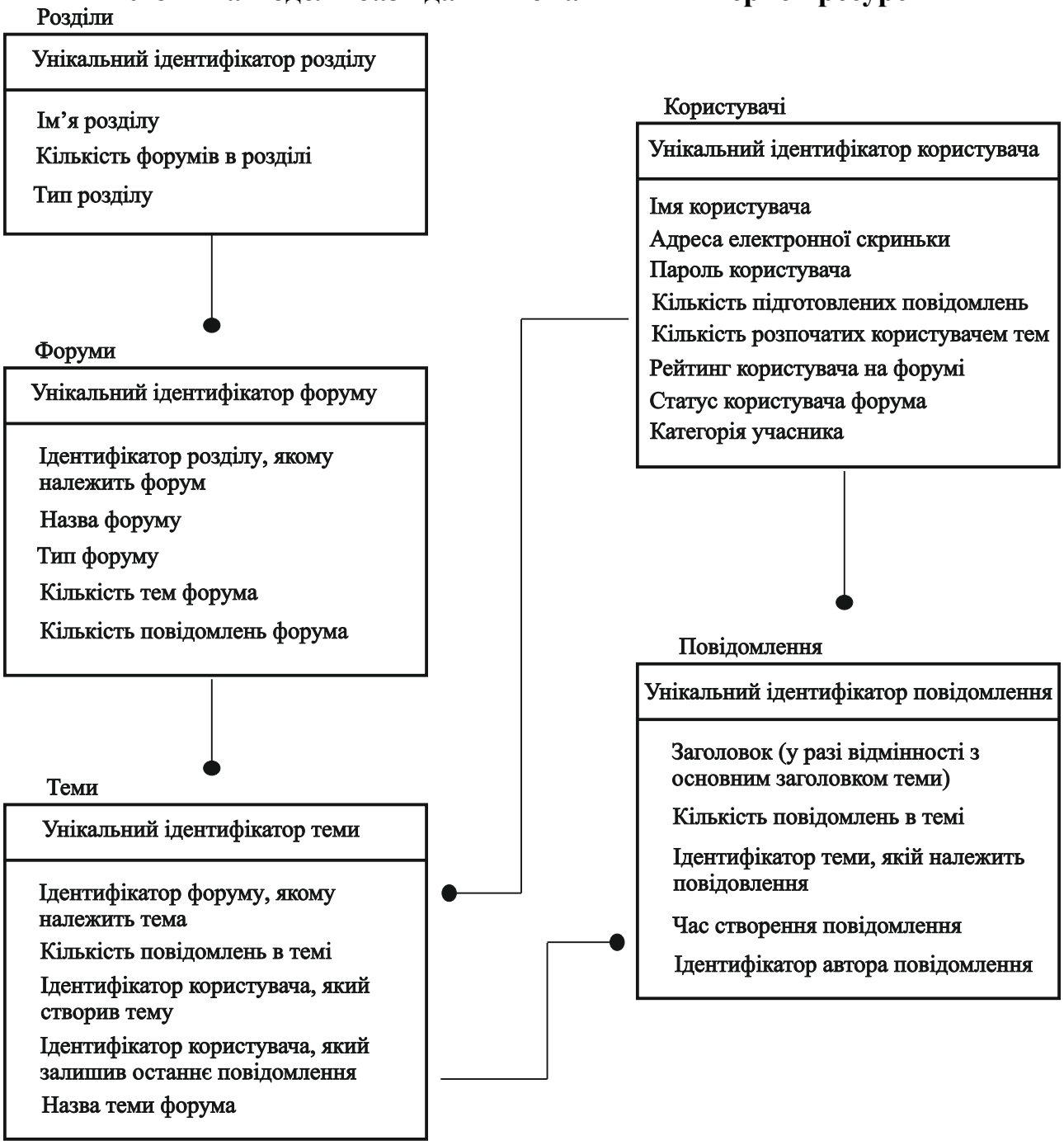


Алгоритм прогнозування вразливостей та загроз інформаційної безпеки



Модель потоку текстових повідомлень та бази даних інтернет-форуму

Логічна модель бази даних тематичних інтернет-ресурсів



Структура повідомлення тематичного інтернет-форуму

Модель потоку текстових повідомлень тематичних форумів

Модель потоку текстових повідомлень, представляється трійкою (1):

$$S_\tau = (M, O, T), \tag{1}$$

де S_τ – потік текстових повідомлень інтернет мережі у поточний час τ; M - множина інтернет повідомлень у потоку даних; O - онтологія даної предметної області; $T = \{1,...,\tau\}$ - множина часових інтервалів, в які велися спостереження за потоком текстових повідомлень (годин, днів, місяців).

$$O = (E, R, F), \tag{2}$$

де E - множина термінів заданої предметної області; R - множина взаємозв'язків між термінами даної предметної області; F - множина заданих на відношеннях та термінах онтології функцій аксіоматизації (інтерпретації).

Кожне текстове повідомлення $d \in M$ може бути представлене наступним чином:

$$d \in (s, t, F_d, A), \tag{3}$$

де s - текст інтернет повідомлення, $t = \{1,...,\tau\}$ - момент часу створення текстового повідомлення; $F_d = \{w_1,...,w_k\}$ – вектор, що представляє текстове повідомлення заданої предметної області, заданою відповідною онтологією O, k – кількість в онтології O термінів, координати $w_i\,(i = 1,...,k)$ – ваги термінів у текстовому повідомленні, A – рейтинг автора текстового повідомлення.

При розрахунку ваги термінів використовується модель TF–IDF (Term frequency – Inverse document frequency), відповідно до якої вага терміна текстового повідомлення прямо пропорційна частоті входження терміна в інтернет повідомлення і обернено пропорційна кількості текстових повідомлень, у яких зустрічається термін (4):

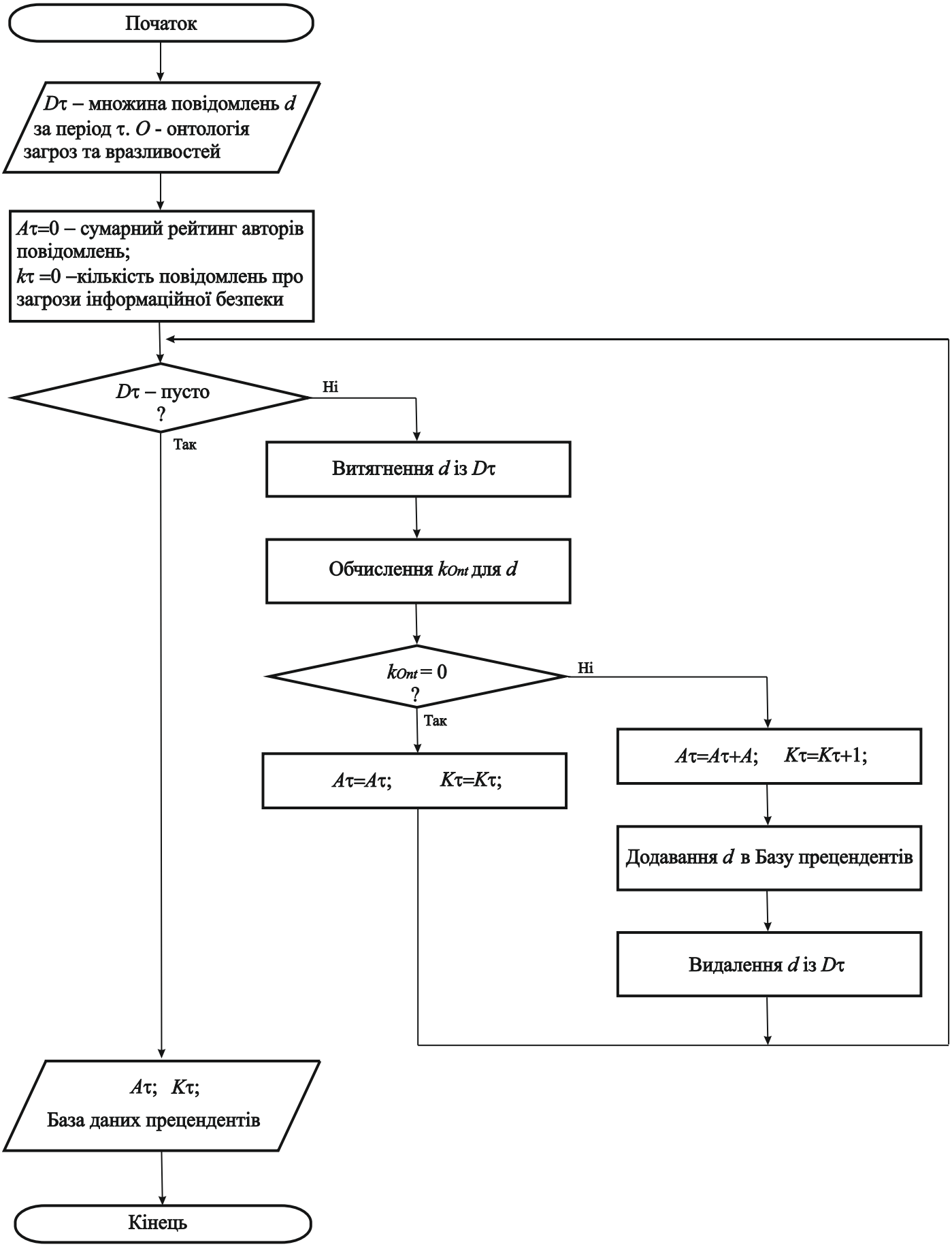
$$w_i = F_i \cdot \log\left(\frac{D}{DF_i}\right), \tag{4}$$

де w_i – вага i - терміна у текстовому повідомленні; F_i – частота i - терміна у інтернет повідомленні; D – загальна кількість текстових повідомлень; DF_i – кількість текстових повідомлень, у яких зустрічається i - термін.

Розглянута модель не враховує того, що текстові повідомлення тематичних форумів можуть мати різний розмір, у зв'язку з чим, вага терміна і відповідно частота будуть зменшуватися зі зростанням розміру текстових повідомлень. Враховуючи дану ситуацію необхідно проводити нормування ваг термінів у інтернет повідомленні, діленням їх на довжину вектора-повідомлення (еклідову норму) (5):

$$w_i^* = \frac{w_i}{\overline{d}} = \frac{w_i}{\sqrt{\sum_{i=1}^{k} w_i}} \tag{5}$$

Алгоритм фільтрації потоку текстових повідомлень та статистичного аналізу інформаційної безпеки



Метод прогнозування вразливостей та загроз інформаційної безпеки

1. Формування правил логічних нечітких продукцій у вигляді :

$$ЯКЩО\left(u_1 \epsilon A_1\right) I \ldots I\left(u_n \epsilon A_n\right) TO\left(y \epsilon Q_j\right)$$

2. Фазифікації вхідних параметрів: $\mu_{A_i}(x)=\begin{cases}1, x=u_i\\ 0, інакше\end{cases}$

3. Обчислення коефіцієнтів ступенів приналежності підумов відповідно до правил нечітких логічних продукцій: $A_i\left(u_i\right)=A_i\left(u_i\right) \wedge A_i\left(u_i\right),$

4. Агрегування умов, відповідно до правил нечітких логічних продукцій. Визначення ступенів приналежності продукційних правил. При перетині множин використовується метод Т-норма, часним випадком є операція мінімуму:

$$a_j=A_1\left(u_1\right) \wedge A_2\left(u_2\right) \wedge \ldots \wedge A_n\left(u_n\right),$$

де a_j - ступінь приналежності передумови для j - го правила; $A_1\left(u_1\right) \wedge A_2\left(u_2\right) \wedge \ldots \wedge A_n\left(u_n\right)$ - нечіткі логічні множини для n підумов j-го правила. Для подальших розрахунків використовуються продукційні правила, для яких значення ступенів приналежності не дорівнює нулю.

5. Активізації нечітких виводів у правилах продукцій. Операція, здійснюється із застосуванням методу мінімуму. Для вихідних змінних визначаються «усічені» функції приналежності, розглядаються лише активні правила продукцій.

$$\overline{Q}_i(y)=a_j \wedge Q_i(y),$$

де a_j - значення коефіцієнта ступеня приналежності передумови j - го правила продукції, $Q_i(y)$- нечітка множина виводів j - го продукційного правила, $\overline{Q}_i(y)$ - «усічена» нечітка множина виводів j - го продукційного правила.

6. Акумуляції виводів правил продукцій. Здійснюється об'єднанням «усічених» функцій приналежності та отриманням для вихідної змінної підсумкової множини. Для об'єднання множин застосовується метод S-норма:

$$\overline{Q}(y)=\overline{Q}_1(y) \vee \overline{Q}_2(y) \vee \ldots \vee \overline{Q}_j,$$

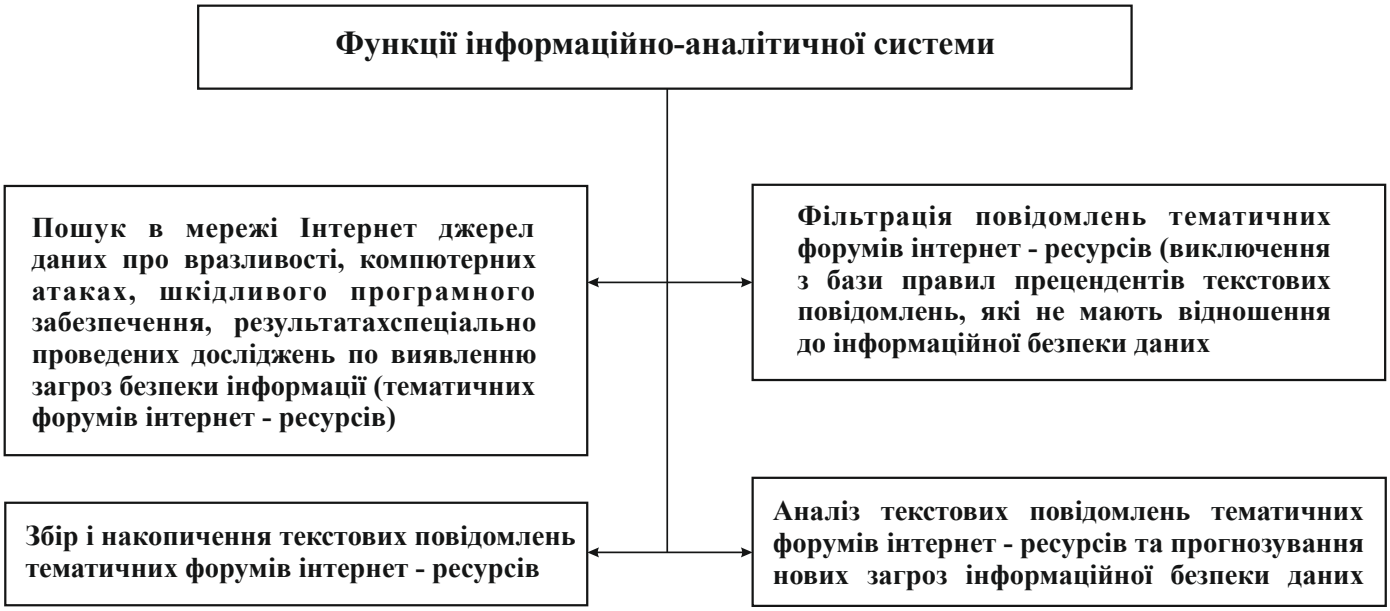
де $\overline{Q}(y)$ – логічна нечітка множина, що відповідає результату роботи логічної нечіткої системи; $\overline{Q}_1(y) \vee \overline{Q}_2(y) \vee \ldots \vee \overline{Q}_j$ – «усічені» нечіткі логічні множини, що відповідають виводам продукційним активним правилам.

7. Дефазифікації. Отриманий результат логічного виводу приводиться до чіткого представлення, із застосуванням методу центра ваги.

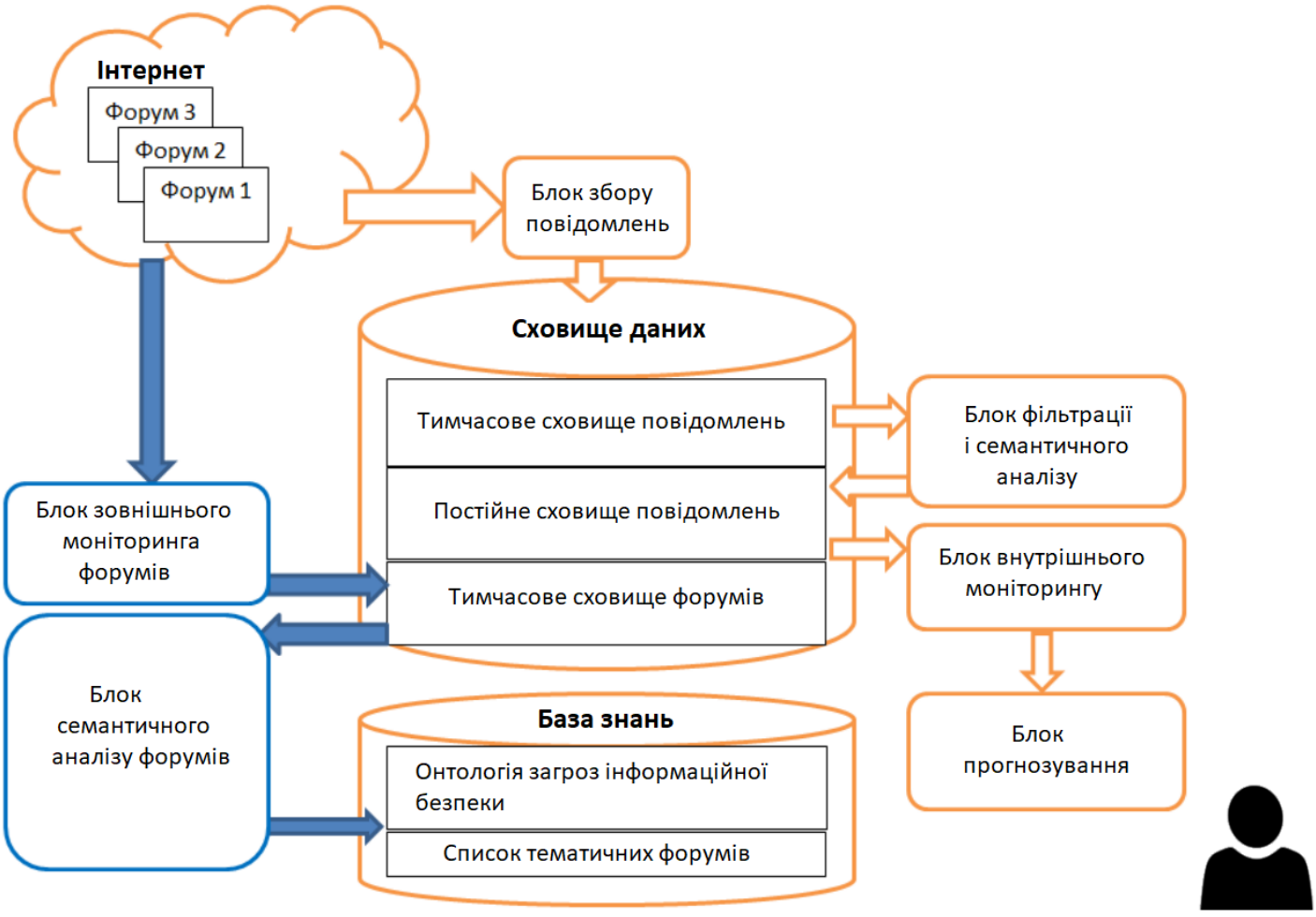
$$y=\frac{\sum_{j=1}^{R} b_j \int \mu_{\overline{Q}_j(y)} dy}{\sum_{j=1}^{R} \int \mu_{\overline{Q}_j(y)} dy},$$

де y - чітке значення результату виходу нечіткої логічної системи; b_j - центри функцій приналежності відповідних термів онтології вихідної нечіткої змінної y для j - го правила продукції; R – кількість правил логічних нечітких продукцій; $\int \mu_{\overline{Q}_j(y)} dy$ – величина площі під усіченою нечіткою множиною $\overline{Q}_j$ для j-го правила продукції.

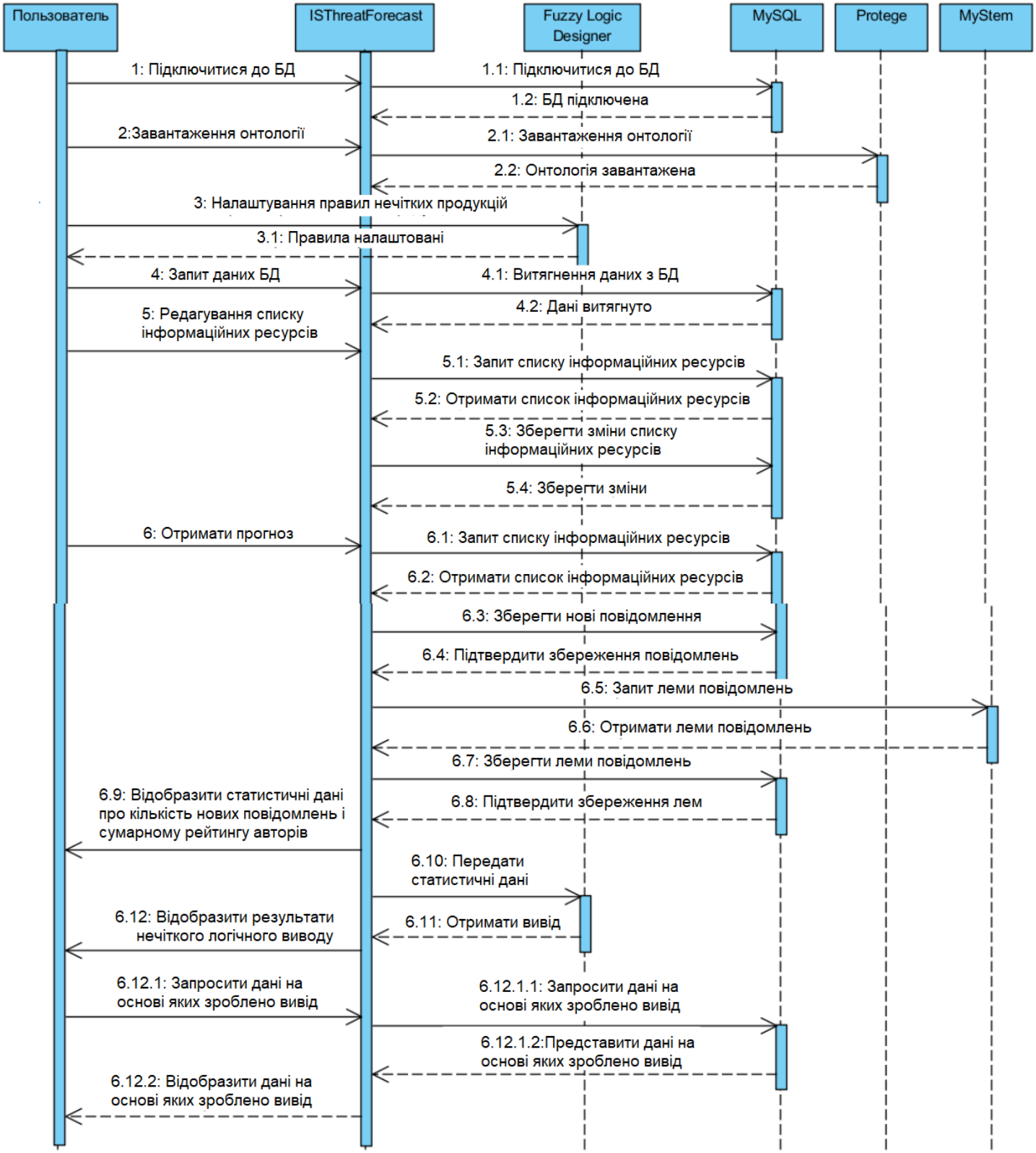
Структура та функції інформаційно-аналітичної системи



Структура інформаційно-аналітичної системи



Діаграма діяльності інформаційно-аналітичної системи



Показники якості прогнозування інформаційно - аналітичної системи

№ п/п	Назва, формула, опис
1	*MAPE* – середня абсолютна процентна помилка системи прогнозування $$MAPE = \frac{1}{h}\sum_{i=1}^{h}\left|\frac{f_{T,i} - y_{T+i}}{y_{T+i}}\right| \cdot 100\%, \quad (1)$$ де h - довжина інтервалу, на якому проводиться прогнозування загроз; $f_{T,i}$ - прогнозне значення часового ряду, отримане в момент часу T на i кроків наперед; y_{T+i} - значення часового ряду в момент часу $T+i$
2	*MAE* - середня абсолютна помилка системи прогнозування: $$MAE = \frac{1}{h}\sum_{i=1}^{h}\left| f_{T,i} - y_{T+i}\right|, \quad (2)$$
3	*RMSE* - квадратний корінь із середньої квадратичної помилки системи прогнозування: $$RMSE = \sqrt{\frac{1}{h}\sum_{i=1}^{h}\left(f_{T,i} - y_{T+i}\right)^2}, \quad (4.3)$$
Для оцінки якості прогнозування загроз зручніше використання середньої абсолютної процентної помилка (*MAPE*), вимірюється у відсотках від значення прогнозованого показника. Показник може бути використаний для порівняння якості прогнозування загроз, систем побудованих із застосуванням різних моделей, також в якості прогнозування конкретних моделей, для яких визначено рівень помилки прогнозування критичний.	
4	Відносна кількість випадків прогнозування до загального числа випадків: $$\eta = \frac{p}{p+q}, \quad (4)$$ де p – число випадків прогнозування, які підтверджені фактичними даними; q – число випадків, які не знайшли фактичного підтвердження.
5	Критерій Дарбіна-Уотсона: $$d = \frac{\sum_{i=2}^{m}\left(e_i - e_{i-1}\right)^2}{\sum_{i=2}^{m} e_i^{\,2}}, \quad (5)$$ де m - довжина часового ряду, e_i - помилка прогнозування загроз: $e_i = x_i - \bar{x}_i$

ВИСНОВКИ

В результаті магістерського дослідження вирішена наукова задача, полягає в підвищенні ефективності засобів та методів виявлення вразливостей та загроз інформаційної безпеки конфіденційних даних на основі розробки інформаційно-аналітичної системи та алгоритмів для проведення аналізу потоку повідомлень тематичних форумів інтернет-ресурсів.

Для досягнення мети поставленої в магістерській роботі вирішено та сформульовано наступні задачі:

1. Розроблено модель бази знань тематичного форуму інтернет-ресурсу, призначена для прогнозування вразливостей та загроз безпеки інформації, відрізняється можливістю працювати з різнотипними потоками даних, різних програмних платформ, що застосовуються для реалізації дискусійних тематичних форумів інформаційних ресурсів, а також представлена модель потоку текстових повідомлень, що належать до заданої предметної області, описаної заданою онтологією, відрізняється від аналогів можливістю статистичного аналізу та семантичної фільтрації текстових повідомлень, дозволяє прогнозувати вразливості та загрози, враховуючи їхню приналежність до конкретного форуму, рейтингу автора, кількості повідомлень теми форуму, часу створення, темі форуму.

2. Розроблено алгоритм прогнозування вразливостей та загроз безпеки інформації, заснований на логічному нечіткому виводу, семантичному та статистичному аналізі, відрізняється від аналогів можливістю виявлення вразливостей та загроз до їх безпосередньої реалізації, а також гнучко дозволяє описувати закономірності процесу наповнення тематичних форумів інтернет-ресурсів новими текстовими повідомленнями, що в результаті сприяє покращенню якості прогнозування загроз.

3. Розроблено для вирішення задачі прогнозування вразливостей та загроз безпеки інформації алгоритм проведення аналізу потоку текстових повідомлень тематичних форумів інтернет-ресурсів, заснований на семантичному та статистичному аналізі, відрізняється від наявних можливістю обчислювати статистичні параметри, здійснювати семантичну фільтрацію текстових повідомлень, для прогнозування подій системи нечіткого логічного виводу.

4. Запропоновано інформаційно-аналітичну систему прогнозування вразливостей та загроз безпеки інформації шляхом проведення автоматизованого аналізу текстових повідомлень форумів тематичних інтернет-ресурсів, реалізує запропоновані алгоритми, дозволяє прогнозувати вразливості та загрози, вживати адекватних заходів щодо захисту інформації. Отримані результати свідчать про ефективність запропонованого методу прогнозування вразливостей та загроз, а також коректній роботі розробленої інформаційно-аналітичної системи та можливості застосування на практиці.

За темою роботи опубліковано 1 теза та 1 наукова стаття.

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ

КАФЕДРИ КІБЕРБЕЗПЕКИ

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджусмо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод прогнозування вразливостей інформаційної безпеки на основіданих інтериет-ресурсів

Автор: Пахар Олександр Валерійович
Спеціальність: 123 – Комп'ютерна інженерія
Освітня програма: Програмування та захист комп'ютерних систем і мереж
Науковий керівник: Касьянчук М.М., к.т.н, доц.
Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних). Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи.	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допушена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
5	Інше:	

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

1) запозичення розміщені в розділах аналізу існуючих аналогів та прототипів, які не описують безпосередньо авторське дослідження і не стосуються результатів роботи;
2) усі запозичення фрагментарні, або мають належним чином оформленні посилання;
3) всі зафіксовані системою ознаки модифікації тексту відносяться до комбінування латинських символів зі україномовними скороченнями індексів в формулах, що не є модифікацією тексту.

Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ ідентичності/схожості, складає 3,32% з урахуванням наведених обгрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи.

Керівник роботи _____ Михайло КАСЬЯНЧУК

Завідувач кафедри кібербезпеки _____ Юрій КЛЬОЦ

Дата:

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

РЕЦЕНЗІЯ НА ДИПЛОМНУ РОБОТУ

освітнього ступеня «магістр»

Магістр Пахар Олександр Валерійович

Тема Метод прогнозування вразливостей інформаційної безпекина основі даних інтернет-ресурсів

Спеціальність 125 «Кібербезпека»

спеціалізація «Кібербезпека»

Обсяг дипломної роботи освітнього ступеня «магістр»:

кількість листів креслень 13; кількість сторінок записки 80

1. Короткий зміст ДР та прийнятих рішень В рамках магістерської роботи вирішена задача підвищення ефективності методів та засобів виявлення вразливостей, загроз безпеки інформації на основі запропонованих алгоритмів, методів та інформаційно-аналітичної системи аналізу потоку повідомлень інтернет-ресурсів, та отримані основні результати: модель потоку повідомлень та бази даних, тематичного ресурсу, відрізняється можливістю опрацьовувати різнотипні дані, що застосовуються для організації дискусійних інформаційних тематичних ресурсів, а також можливістю статистичного аналізу та семантичної фільтрації повідомлень, дозволяє прогнозувати вразливості та загрози, враховуючи, при цьому, їхню тематичну приналежність; метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів, заснований на нечіткому логічному виводу, семантичному та статистичному аналізі, відрізняється можливістю виявлення вразливостей та загроз до їх реалізації, дозволяє описувати закономірності інформаційного процесу наповнення тематичних ресурсів новими текстовими повідомленнями, що відображається на якості прогнозування.

2. Висновок про відповідність ДР дипломному завданню Дипломна робота освітнього ступеня «магістр» у повній мірі відповідає поставленому завданню як в теоретичній, так і в практичній частині дипломної роботи

3. Характеристика виконання кожного розділу роботи, ступінь використання останніх досягнень науки і техніки і передових методів роботи: У вступі обгрунтовується актуальність теми роботи, дається аналіз досліджуваної проблеми і обгрунтовується застосовуваний підхід до її вирішення, формулюються цілі і завдання дослідження, описується наукова новизна і практична значимість отриманих результатів. У першому розділі якісно та в повній мірі проаналізовано сучасний стан вразливостей та загроз інформаційної безпеки та засобів захисту інформації. Наступні розділи присвячені розробці моделі та методу прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів. Розглянуто питання оцінки методу прогнозування вразливостей та загроз інформаційної безпеки.

4. Позитивні сторони проекту Дипломна робота містить ряд інноваційних рішень, зокрема, в розробці моделей і алгоритмів автоматизації проведення аналізу потоку даних тематичних інтернет-ресурсів та прогнозування появи нових вразливостей, загроз безпеки інформації з використанням нечіткого логічного виводу.

5. Негативні сторони проекту В роботі не наведена інформація про рівні доступу та їх роль при роботі з базою правил та інформаційно – аналітичною системою. Як даний підхід реалізований в магістерській роботі?

6. Оцінка графічного оформлення та пояснювальної записки роботи Графічне оформлення виконане відповідно до теми дипломної роботи з дотриманням стандартів. В загальному графічне оформлення виконане на достатньому рівні. Пояснювальна записка відповідає нормам для її оформлення.

7. Відгук про роботу в цілому В загальному дипломна робота заслуговує позитивної оцінки. Весь матеріал дипломної роботи структурований, чіткий та послідовний. Усі розділи роботи послідовні та логічні, що дозволяє чітко розуміти викладений матеріал в рамках тематики дипломної роботи. Графічний матеріал дозволяє наочно побачити доцільність та ефективність рішень, які були прийняті за основу для досягнення поставленої задачі.

8.Інші зауваження

9. Оцінка дипломної роботи Розглянувши позитивні та негативні сторони представленої дипломної роботи, можна зробити висновок, що вона заслуговує оцінку «відмінно».

РЕЦЕНЗЕНТ (прізвище, ім'я, по батькові, посада, місце роботи)

Лисенко Сергій Миколайович, д.т.н., професор, кафедра комп'ютерної інженерії та інформаційних систем, Хмельницького національного університету

« 2 » 12 2022р. _____ (підпис)

ХНУ. Зам. 54, тир. 6000, 2007.

Завідувачу кафедри кібербезпеки
к.т.н., доц. Кльоцу Ю.П.
Пахаря Олександра Валерійовича
ПІБ здобувача вищої освіти

студента ФІТ, 2 курсу, групи КІм-21-1

ЗАЯВА

З правилами чинного Положення «Про дотримання академічної доброчесності в Хмельницькому національному університеті» від 26.09.2020 (зі змінами від 26.11.2020), згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування заходів дисциплінарної та академічної відповідальності, ознайомлений (а). Про використання програмно-технічних засобів для перевірки кваліфікаційних робіт здобувачів вищої освіти на плагіат оповіщений (а) та надаю свою згоду на обробку та збереження університетом моєї роботи в інституційному репозитарії університету.

Також надаю університету право на передачу моєї роботи для обробки та збереження в базах даних програмно-технічних засобів (Unicheck та Anti-Plagiarism) та використання роботи для виявлення плагіату в інших роботах, які перевіряються програмно-технічними засобами та користувачами, що мають доступ до цих програмно-технічних засобів, виключно в обмежених цілях для виявлення плагіату в текстах робіт.

Робота для перевірки університетом надається в друкованому та електронному варіанті. Електронна версія моєї роботи збігається (ідентична) з друкованою.

5.12.2022
дата

підпис

Anti-Plagiarism v-15.257

Максимальное совпадение с одним документом 1.0%

Словари проверки: en_US, ru_RU, ua_UA. **Ошибок в документах: 8%**

	Документ		Суммарное совпадение по Базе Данных	
ID: 108801 Название: Метод прогнозування вразливостей інформаційної безпеки на основі даних інтернет-ресурсів Добавлено в БД: 2022-11-28 Авторы: Пахар О.В. Руководители: Касянчук М.М. Консультанты: Опоненты:	Символы	Лексемы	Символы	Лексемы
	109355	651	1580 (1%)	26 (4%)

Источник плагиата

ID	Описание	Наличие плагиата в документе	
		Символы	Лексемы

UNICHECK

Ім'я користувача:
Кафедра кібербезпеки

ID перевірки:
1013208399

Дата перевірки:
06.12.2022 12:01:33 EET

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
06.12.2022 14:17:03 EET

ID користувача:
100008300

Назва документа: **Пахар_О_В_МагістерськаNew2**

Кількість сторінок: **81** Кількість слів: **15135** Кількість символів: **123653** Розмір файлу: **20.13 MB** ID файлу: **1012824901**

3.32%
Схожість

Найбільша схожість: **2.47%** з джерелом з Бібліотеки (ID файлу: **1012824874**)

0.83% Джерела з Інтернету	8	Сторінка 83
2.64% Джерела з Бібліотеки	15	Сторінка 83

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи	7