

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет кібербезпеки, комп'ютерної та програмної інженерії (ЗФН)

Кафедра комп'ютерних інформаційних технологій

ДОПУСТИТИ ДО ЗАХИСТУ

Завідувач випускової кафедри

Аліна САВЧЕНКО

«_____» _____ 2021 р.

ДИПЛОМНА РОБОТА
(ПОЯСНЮВАЛЬНА ЗАПИСКА)

ВИПУСКНИКА ОСВІТНЬОГО СТУПЕНЯ
«МАГІСТРА»

ЗА ОСВІТНЬО-ПРОФЕСІЙНОЮ ПРОГРАМОЮ «ІНФОРМАЦІЙНІ
УПРАВЛЯЮЧІ СИСТЕМИ ТА ТЕХНОЛОГІЇ»

Тема: «Програмний модуль інтелектуального аналізу текстових файлів для пошуку інформації з обмеженим доступом»

Виконавець: студент групи УС-201Мз Ревтюх Богдан Олегович

Керівник: к.т.н., доцент Климова Асія Сабирівна

Нормоконтролер: _____ Ігор РАЙЧЕВ

Київ-2021

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет кібербезпеки, комп'ютерної та програмної інженерії (ЗФН)

Кафедра комп'ютерних інформаційних технологій

Галузь знань, спеціальність, освітньо-професійна програма:: 12 «Інформаційні технології», 122 «Комп'ютерні науки», «Інформаційні управляючі системи та технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри

Аліна САВЧЕНКО

«___»_____2021р.

ЗАВДАННЯ

на виконання дипломної роботи студента

Ревтюха Богдана Олеговича

1. **Тема роботи:** «Програмний модуль інтелектуального аналізу текстових файлів для пошуку інформації з обмеженим доступом» затверджена наказом ректора від «12» жовтня 2021 р. № 2229/ст.
2. **Термін виконання роботи:** з 12 жовтня 2021 р. по 31 грудня 2021 р.
3. **Вихідні дані до роботи:** моделі життєвого циклу ІС, технічне завдання до системи, методології моделювання бізнес – процесів UML, IDEF, DFD.
4. **Зміст пояснювальної записки:** вступ, аналіз документації згідно до завдання роботи, розробка програмного забезпечення, тестування системи, висновки.
5. **Перелік обов'язкового графічного матеріалу:** таблиці, рисунки, діаграми.

6. Календарний план-графік

№ п/п	Завдання	Термін виконання	Підпис керівника
1.	Отримання завдання на дипломну роботу.	12.10.21	
2.	Аналіз літератури та джерел за темою дипломної роботи.	13.10.21- 20.10.21	
3.	Розробка розділу 1: Інформаційні ресурси як основна складова порушень безпеки інформації	21.10.21- 25.10.21	
4.	Розробка розділу 2: Методи інтелектуального аналізу текстових документів.	26.10.21- 31.10.21	
5.	Розробка розділу 3: Розробка програмного модуля інтелектуального аналізу текстових файлів для пошуку інформації з обмеженим доступом	01.11.21- 10.11.21	
6.	Розробка розділу 4: Програмна реалізація системи формування заліково – екзаменаційних відомостей.	11.11.21- 15.11.21	
7.	Технічне оформлення пояснювальної записки та графічних матеріалів.	16.11.21- 27.11.21	
8.	Підпис усіх необхідних документів у встановленому порядку.	28.11.21- 15.12.21	
9.	Підготовка до захисту дипломної роботи.	15.12.21- 20.12.21	

7. Дата видачі завдання «__» _____ 2021р.

Керівник дипломної роботи _____ Асія КЛИМОВА
(підпис керівника)

Завдання прийняв до виконання _____ Богдан РЕВТЮХ
(підпис випускника)

РЕФЕРАТ

Пояснювальна записка до дипломної роботи «Програмний модуль інтелектуального аналізу текстових файлів для пошуку інформації з обмеженим доступом» складається з чотирьох розділів, містить 61 сторінок, 9 рисунків, 1 таблицю, 1 додаток, 27 літературних джерел.

Ключові слова: ІЗoД, ПЕoМ, ПСІ, Програмний модуль, текстовий документ.

Об'єктом дослідження: оптимізація шляхів перевірки ПЕoМ.

Предмет дослідження: методи перевірки ПЕoМ за допомогою різноманітного програмного забезпечення.

Метою роботи: є розробка програмного додатку для процесу перевірки ПЕoМ на наявність інформації з обмеженим доступом.

В рамках дипломної роботи було описано та проаналізовано методи інтелектуального аналізу текстових документів, розроблено програмне забезпечення.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	7
ВСТУП.....	8
РОЗДІЛ 1. ІНФОРМАЦІЙНІ РЕСУРСИ ЯК ОСНОВНА СКЛАДОВА ПОРУШЕНЬ БЕЗПЕКИ ІНФОРМАЦІЇ.....	10
1.1. Інформаційні ресурси.....	11
1.2. Законодавство України у сфері захисту інформації.....	12
1.3. Відповідальності за порушення режиму секретності.....	14
РОЗДІЛ 2. МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ДОКУМЕНТІВ.....	17
2.1. Процес інтелектуального аналізу тексту.....	18
2.2. Виділення інформації з тексту.....	19
2.3 Інші методи інтелектуального аналізу тексту.....	20
2.4. Існуючі програмні додатки для перевірки ПЕОМ на наявність ІзОД, їх переваги та недоліки.....	21
2.5. Вибір підходу для виконання завдання	24
2.5.1. Метод виявлення опосередкованих порушень правил поведження з ІзоД	25
2.5.2. Метод виявлення прямих порушень правил поведження з ІзоД	31
РОЗДІЛ 3. РОЗРОБКА ПРОГРАМНОГО МОДУЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ФАЙЛІВ ДЛЯ ПОШУКУ ІНФОРМАЦІЇ З ОБМЕЖЕНИМ ДОСТУПОМ.....	33
3.1. Створення словника ключових слів ПСІ.....	34
3.2. Створення програмного модулю, для пошуку інформації з обмеженим доступом.....	42

РОЗДІЛ 4. ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНОГО ПІДХОДУ ДО АВТОМАТИЗАЦІЇ ПРОЦЕСУ ПЕРЕВІРКИ ПЕОМ НА НАЯВНІСТЬ ІзОД.....	45
4.1. Інструкція користувачу.....	45
4.2. Міжнародний стандарт оцінки якості ПЗ ISO 9126.....	46
4.3. Оцінка запропонованого підходу.....	50
ВИСНОВКИ.....	52
СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ.....	53
ДОДАТОК 1.....	56

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТА ТЕРМІНІВ

PC - Personal Computer

ВВНЗ – вищий військовий навчальний заклад

ІзОД – інформація з обмеженим доступом

ІТ – інформаційні технології

КСЗІ – комплексна система захисту інформації

МОУ – Міністерство Оборони України

НГШ ЗСУ – начальник генерального штабу збройних сил України

ОВТ – озброєння та військова техніка

ПЕОМ – персональна електронно-обчислювальна машина

ПЗ – програмне забезпечення

ПСІ – перелік службової інформації

СЗІ – система захисту інформації

ТЗІ – технічний захист інформації

АС – автоматизована система;

РП – робоча програма;

ІС - інформаційна система;

ПЗ - програмне забезпечення;

ПЗ- програмні засоби;

ТЗ – технічне завдання;

МКР – модульна контрольна робота;

КІТ - кафедра комп'ютерних інформаційних технологій;

ВНЗ – вищий навчальний заклад;

СОД – система обробки даних;

ІПС – інформаційно – пошукові системи;

ВСТУП

В умовах ведення бойових дій, Україна розпочала свій рух шляхом стрімкого розвитку науково-технічного прогресу у військовій сфері. Галузь розробки воєнних технологій трансформувалась у важливий чинник воєнно-технічної політики і воєнної безпеки провідних країн світу, потужну рушійну силу для розвитку озброєння та військової техніки (ОВТ). Під тиском нових відкриттів та винаходів сучасні війни все більше перетворюються на технологічні, тоді як самі воєнні технології набувають на світовому ринку озброєнь статусу окремого товару. Все вказане є ознакою переходу воєнної промисловості світу на новий етап розвитку, що полягає у безпосередньому використанні нових наукових знань (сучасних інформаційних технологій). Проте, багато шахраїв мають бажання перехопити дані технології та інформацію про них, що недопустимо відповідно до законів України “Про інформацію” та “Про доступ до публічної інформації”. З метою забезпечення безпеки інформації працюють багато служб, що займаються виявленням джерел витоків інформації та приводять дані про службову інформацію у відповідність з вимогами законодавства України.

Також існує інформація доступ до якої є обмеженим для осіб що мають певний рівень допуску - інформація з обмеженим доступом (ІЗОД). Перелік даної інформації описаний у ПСІ, затверджений начальником генерального штабу ЗСУ. Часто посадовим особам доводиться працювати з документами, які містять інформацію з обмеженим доступом. Але не всі особи, що працюють з цією інформацією дотримуються правил роботи з ІЗОД чим порушують законодавство та не усвідомлюючи цього стають порушниками режиму секретності, за що мають понести відповідальність по закону України.

Мета роботи: розробка програмного додатку для процесу перевірки ПЕОМ на наявність інформації з обмеженим доступом.

Виходячи з мети роботи, виникають наступні завдання:

- розглянути типи інформаційних ресурсів, як основну складову;
- дослідити шляхи та методи оптимізації системи перевірки;

- розглянути способи спрощення процесу перевірки;
- розробити ефективне програмне забезпечення для перевірки ПЕОМ на наявність інформації з обмеженим доступом;

Об'єкт досліджень: оптимізація шляхів перевірки ПЕОМ;

Предмет досліджень: методи перевірки ПЕОМ за допомогою різноманітного програмного забезпечення.

РОЗДІЛ 1

ІНФОРМАЦІЙНІ РЕСУРСИ ЯК ОСНОВНА СКЛАДОВА ПОРУШЕНЬ БЕЗПЕКИ ІНФОРМАЦІЇ

Під інформаційним ресурсом слід розуміти системну єдність впорядкованої у межах певного носія (сховища, місця розташування тощо) інформації (знань) з середовищем її формування, накопичення та розвитку, а також засобами та технологіями її зберігання, використання та передачі. До системи інформаційного ресурсу також входять організаційні структури з його обслуговування (персонал) та механізми захисту інформації від некоректного доступу до неї. Інформаційний ресурс як правило належить конкретному суб'єкту, має свою цінність для користувачів (споживчу вартість) та не є вичерпаним за своїм обсягом. Динаміка розвитку інформаційного ресурсу залежить від його здатності задовольняти мінливі потреби користувачів.

Основними об'єктами державно-управлінського впливу на інформаційний ресурс є: режим (порядок) формування, використання та накопичення інформації у межах певного ресурсу; норми та правила формування інформаційного ресурсу та надання доступу до нього; інформаційно-телекомунікаційна інфраструктура (інформаційні мережі локального, регіонального, державного та глобального рівнів; джерела виникнення інформації та канали її розповсюдження; засоби зберігання та розповсюдження інформації; організаційні структури з обслуговування відповідної інфраструктури та їх персонал тощо); технології формування, зберігання, адміністрування та захисту інформації; фізичне місце розміщення інформаційного ресурсу (бібліотеки, сховища, депозитарії тощо), а також безпосередньо самі носії інформації; середовище формування, зберігання та поширення інформації тощо

Кафедра КІТ				НАУ 21 12 98 000 ПЗ			
Виконав	Ревтюх Б.О.			ІНФОРМАЦІЙНІ РЕСУРСИ ЯК ОСНОВНА СКЛАДОВА ПОРУШЕНЬ БЕЗПЕКИ ІНФОРМАЦІЇ	Літера	аркуш	аркушів
Керівник	Климова А.С.					10	7
Консульт.							10
Н. контроль	Райчев І.Е.				УС-201Мз		122

1.1. Інформаційні ресурси

Інформація являє собою сукупність відомостей про факти, об'єкти, події та ідеї, які в даному контексті мають цілком певне значення [1].

У загальному розумінні інформація – це незвичайний ресурс, використання якого не зменшує його кількості та якості (рис1.1).

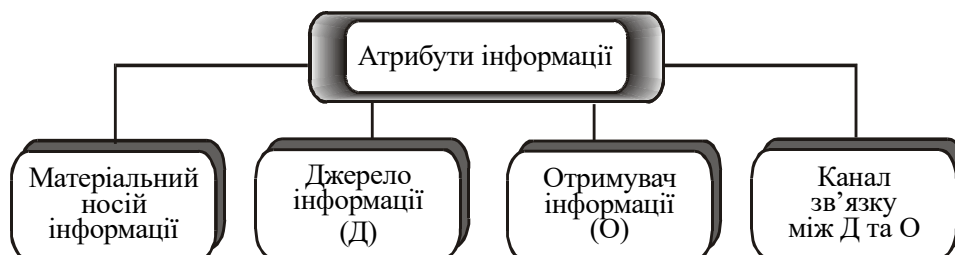


Рис.1.1. Атрибути інформації

Дані – це інформація, подана у формалізованому вигляді, прийнятному для обробки автоматичними засобами за можливої участі людини (рис1.2)[2].



Рис.1.2. Співвідношення понять «інформація» і «дані»

Інформаційні ресурси (*Information resources*) – документи і масиви документів в інформаційних системах (бібліотеках, архівах, фондах, банках даних, депозитаріях, музейних сховищах). Розрізняють інформаційні ресурси державні та недержавні.

Захищена інформація, режим доступу до якої визначається її власниками і дозволяється за спеціальною угодою із споживачем. До цього виду ресурсів належать комерційні архіви (наприклад, комерційні версії програм в архівах *ftp.microsoft.com* або *ftp.bsdi.com*, закриті національні та міжнародні некомерційні

ресурси (наприклад, роботи за міжнародними проектами *CES* або *IAEA*, приватна некомерційна інформація з спеціальними режимами доступу (приватні благодійні фонди, наприклад).

Інформаційні ресурси обмеженого використання, до яких відносяться, наприклад, програми класу *shareware* (*Trumpet Winsock, Atis Mail, Netscape*, і т.п.). В даний клас можуть входити ресурси обмеженого часу використання (поточна версія *Netscape* перестане працювати в червні якщо тільки хтось не зламає захист) або обмеженого часу дії, тобто користувач може використовувати поточну версію на свій страх і ризик, але ніхто не буде надавати йому підтримку.

Вільно розповсюджені інформаційні ресурси або *freeware*, якщо мова йде про програмне забезпечення. До цих ресурсів відноситься все, що можна вільно отримати по мережі без спеціальної реєстрації. Це може бути документація, програми або що-небудь ще. Найвідомішими вільно розповсюджуваними програмами є програми проекту *GNU Free Software Foundation*. Слід зазначити, що вільно розповсюджується програмне забезпечення не має сертифіката якості, але як правило, його розробники відкриті для обміну досвідом[3]. Основними напрямками інформаційної діяльності є: політичний, економічний, соціальний, духовний, екологічний, науково-технічний, міжнародний тощо.

Держава зобов'язана постійно дбати про своєчасне створення, належне функціонування і розвиток інформаційних систем, мереж, банків і баз даних у всіх напрямках інформаційної діяльності. держава гарантує свободу інформаційної діяльності в цих напрямках всім громадянам та юридичним особам в межах їх прав і свобод, функцій і повноважень.

1.2. Законодавство України у сфері захисту інформації

Інформаційна безпека— це стан захищеності систем обробки і зберігання даних, при якому забезпечено конфіденційність, доступність і цілісність інформації, або комплекс заходів, спрямованих на забезпечення захищеності інформації від несанкціонованого доступу, використання, оприлюднення, руйнування, внесення

змін, ознайомлення, перевірки, запису чи знищення (у цьому значенні частіше використовують термін «захист інформації»)[4].

Інформаційна безпека держави характеризується ступенем захищеності і, отже, стійкістю основних сфер життєдіяльності (економіки, науки, техносфери, сфери управління, військової справи, суспільної свідомості і т.д.) стосовно небезпечних (дестабілізаційних, деструктивних, суперечних інтересам країни тощо), інформаційним впливам, причому як до впровадження, так і до вилучення інформації.

Поняття інформаційної безпеки не обмежується безпекою технічних інформаційних систем чи безпекою інформації у чисельному чи електронному вигляді, а стосується усіх аспектів захисту даних чи інформації незалежно від форми, у якій вони перебувають.

У нашій країні існує ряд законів про інформацію, її використання, зберігання, та її захист.

Відповідно до статті 5 закону України “Про інформацію ” від 02.10.92 та його офіційного тлумачення представлене у рішенні конституційного суду України, Кожен має право на інформацію, що передбачає можливість вільного одержання, використання, поширення, зберігання та захисту інформації, необхідної для реалізації своїх прав, свобод і законних інтересів [5].

Реалізація права на інформацію не повинна порушувати громадські, політичні, економічні, соціальні, духовні, екологічні та інші права, свободи і законні інтереси інших громадян, права та інтереси юридичних осіб [6].

А також, відповідно до статті 6 ч.2 право на інформацію може бути обмежене законом в інтересах національної безпеки, територіальної цілісності або громадського порядку, з метою запобігання заворушенням чи злочинам, для охорони здоров'я населення, для захисту репутації або прав інших людей, для запобігання розголошенню інформації, одержаної конфіденційно, або для підтримання авторитету і неупередженості правосуддя [7].

Захист інформації також гарантується законодавством, а за його порушення слідує відповідальність. Відповідно до Закону України “Про державну таємницю”

Звід відомостей, що становлять державну таємницю (далі –ЗВДТ), є єдиною формою реєстрації цих відомостей в Україні. З моменту опублікування ЗВДТ держава забезпечує захист і правову охорону відомостей, які зареєстровані в ньому[8,9].

1.3. Відповідальності за порушення режиму секретності.

Страх притягнення до адміністративної та кримінальної відповідальності за зняття грифів «ДСК». Українське законодавство передбачає суттєво жорсткіше покарання за незаконне оприлюднення інформації з обмеженим доступом, аніж за незаконну відмову у наданні інформації, що має бути відкритою. Порушення законодавства про державну таємницю регулюються Кодексом України про адміністративні порушення та Кримінальним кодексом України. За забезпечення охорони інформації із обмеженим доступом відповідає СБУ, і офіційна статистика підтверджує використання нею каральних норм обох кодексів для притягнення до відповідальності осіб звинувачених у недотриманні обмеженого режиму доступу до інформації [10,11].

Посадові особи та громадяни, винні в:

- розголошенні державної таємниці;
- втрати документів та інших матеріальних носіїв секретної інформації;
- недодержанні встановленого законодавством порядку передачі державної таємниці іншій державі чи міжнародній організації;
- навмисному не віднесенні до державної таємниці інформації, розголошення якої може завдати шкоди інтересам національної безпеки України, а також необґрунтованому зниженні ступеня секретності або необґрунтованому розсекречуванні секретної інформації;
- безпідставному засекречуванні інформації;
- наданні грифа секретності матеріальним носіям конфіденційної або іншої таємної інформації, яка не становить державної таємниці, або ненаданні грифа секретності матеріальним носіям інформації, що становить державну таємницю, а

також безпідставному скасуванні чи зниженні грифа секретності матеріальних носіїв секретної інформації;

- порушенні порядку надання допуску та доступу до державної таємниці;
- порушенні режиму секретності та невиконанні обов'язків щодо збереження державної таємниці;
- невжитті заходів щодо забезпечення охорони державної таємниці та незабезпеченні контролю за охороною державної таємниці;
- впровадженні діяльності, пов'язаної з державною таємницею, без одержання в установленому порядку спеціального дозволу на впровадження такої діяльності, а також розміщенні державних замовлень на виконання робіт, доведенні мобілізаційних завдань, пов'язаних з державною таємницею, в органах державної влади, органах місцевого самоврядування, на підприємствах, в установах, організаціях, яким не надано спеціального дозволу на провадження діяльності, пов'язаної з державною таємницею;
- недодержанні вимог законодавства щодо забезпечення охорони державної таємниці під час здійснення міжнародного співробітництва, прийому іноземних делегацій, груп, окремих іноземців та осіб без громадянства і проведення роботи з ними;
- невиконанні норм і вимог технічного захисту секретної інформації, внаслідок чого виникає реальна загроза порушення цілісності цієї інформації або просочування її технічними каналами, несуть дисциплінарну, адміністративну та кримінальну відповідальність, згідно із законодавством України [12].

Кодування являє собою процес створення кодів (набору цифр, букв та цифр і букв) та присвоєння їх підмножинам об'єктів, отриманих у процесі класифікації. Розрізняють два види методів кодування: а) реєстраційний; б) класифікаційний.

До реєстраційних належать порядковий і серійно-порядковий методи, а до класифікаційних – послідовний і паралельний.

Послідовний метод кодування – це створення коду класифікаційного групування і (чи) об'єкта класифікації з використанням кодів послідовно розміщених підпорядкованих групувань, що були отримані при ієрархічному методі класифікації, та його присвоєння.

Паралельний метод кодування – це створення коду класифікаційного групування і (чи) об'єкта класифікації з використанням кодів незалежних групувань, що були отримані при фасетному методі класифікації, та його присвоєння[13].

Висновки до розділу 1:

Отже, безпека інформації надзвичайно важливою у будь якому разі. Адже захищеність інформації забезпечує її від втрати, неліцензійної заміни, а також від витоку її до осіб, які можуть використати її у злочинних цілях. Особи по вині яких було здійснено втрату інформації мають понести відповідальність згідно чинного законодавства України. Ніяка терміновість та важливість подій не може стати причиною для розповсюдження, розголошення інформації з обмеженим доступом.

РОЗДІЛ 2

МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ДОКУМЕНТІВ

Реалізація програмного забезпечення – це процес перекладу системної специфікації в працездатну систему. Етап реалізації завжди включає процеси проектування і програмування, але якщо для розробки ПЗ застосовується еволюційний підхід, етап реалізації також може включати процес внесення змін у системну специфікацію.

Проектування систем завжди починається з визначення мети проекту. Основне завдання будь-якого успішного проекту, щоб у момент створення системи і протягом всієї її експлуатації можна було забезпечити:

- необхідну функціональність системи та рівень адаптації постійно змінюваних умов її функціонування;
- необхідну пропускну спроможність системи;
- необхідний час реакції системи на запит;
- безвідмовну роботу системи в необхідному режимі, інакше кажучи - готовність і доступність системи в обробці запитів користувачів;
- простоту експлуатації й підтримки системи;
- необхідну безпеку.

Продуктивність є, визначальним ефективність системи. Хороше проектне рішення є підставою високопродуктивної систем.

Згідно сучасної методології, процес створення АС являє собою процес побудови та послідовного перетворення ряду узгоджених моделей на всіх етапах життєвого циклу (ЖЦ) АС [6].

Кафедра КІТ				НАУ 21 12 98 000 ПЗ			
<i>Виконав</i>	<i>Ревтюх Б.О.</i>			МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ДОКУМЕНТІВ	<i>Літера</i>	<i>аркуш</i>	<i>аркушів</i>
<i>Керівник</i>	<i>Климова А.С.</i>					17	16
<i>Консульт.</i>							17
<i>Н. контроль</i>	<i>Райчев І.Е.</i>				УС-201Мз		122

2.1. Процес інтелектуального аналізу тексту

Переважна більшість інформації у електронному вигляді (до 80%) – звичайний неструктурований текст. Отримання значущої інформації – це і є задача інтелектуального аналізу тексту, який є процесом отримання якісної інформації з напів- та неструктурованих даних [2].

Інформація може бути отримана спеціальним програмним забезпеченням шляхом виявлення певних шаблонів чи трендів на основі статистичної обробки тексту (рис. 2.1).

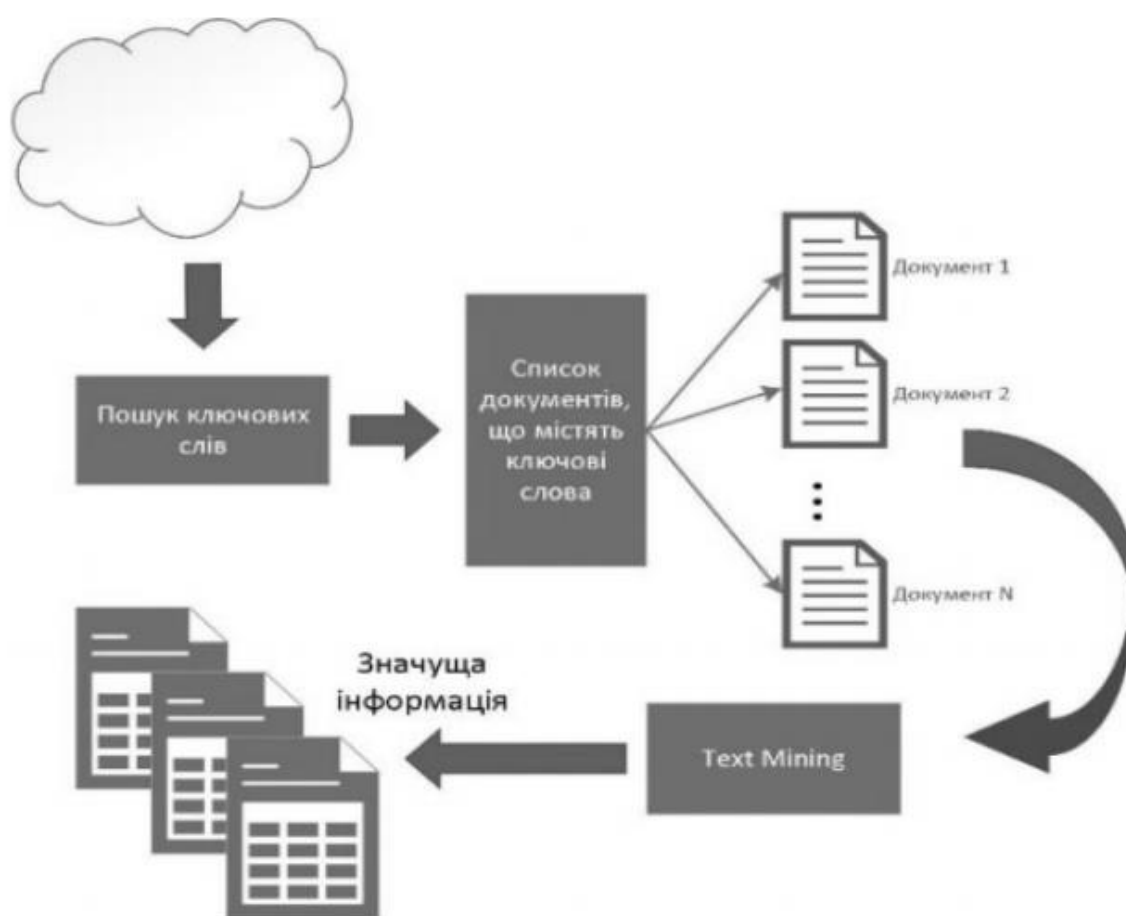


Рис. 2.1. Різниця між пошуком ключових слів та інтелектуальним аналізом тексту

Процес аналізу відбувається відповідно до наступної послідовності кроків.

- Збір (gathering) текстових документів з різних джерел з наступним виконанням попередньої обробки (preprocessing) (усунення реклами; токенизація (tokenization), тобто розділення його на складові частини; усунення незначущих слів

(артиклів, сполучників тощо); вирівнювання (stemming), тобто зведення всіх однокореневих слів до основної форми і, зменшення загальної кількості слів; перетворення тексту до вигляду, придатного для подальшого аналізу, наприклад, у вигляді гістограми частоти появи слів).

- Вибірка потрібних та усунення непотрібних властивостей тексту.
- Застосування одного чи декількох методів інтелектуального аналізу тексту з метою отримання патернів (шаблонів).
- Оцінювання патернів відповідно до критеріїв пошуку. На основі аналізу літературних джерел можна виділити наступні технології інтелектуального аналізу тексту.

2.2. Виділення інформації з тексту

Виділення інформації означає отримання структурованих даних з неструктурованого тексту для отримання набору сутностей (імена, локації, взаємозв'язки тощо). Використовується для відслідковування новин; дописів користувачів соціальних мереж з метою отримання їх думки стосовно певної події; отримання інформації з повідомлень електронної пошти; отримання особистих даних з профілів користувачів; отримання інформації в цифрових онлайн бібліотеках; виділення таблиць з тексту. Виділення в тексті застосовують для того, щоб підкреслити головні положення та висновки; зробити логічний наголос на окремих словах, групах слів або реченнях; виділити правила, приклади, визначення, цитати, ремарки, тощо.

Засоби виділення: шрифтові (курсив, півжирний, великі літери, шрифти іншої гарнітури або кегля тощо), не шрифтові (розрядка, втяжки, лінійки, інша фарба тощо) і комбіновані — шрифтові й нешрифтові. (Розрядка— збільшення інтервалів між літерами. Втяжка— відступ)

Загальний принцип підбору виділень залежить від читацького призначення і виду літератури. У книжках для дітей, особливо дошкільного та молодшого

шкільного віку застосовують півжирний шрифт, друк іншою фарбою, систему виділень за допомогою втяжок і лінійок, значно рідше — розрядку. У літературі для дітей середнього шкільного віку, переважно у підручниках, пів жирного менше, натомість широко вживають розрядку та курсив.

У підручниках для старших класів середньої школи, технікумів, училищ переважають курсив і розрядка, значно рідше трапляється півжирний, ще рідше — виділення за допомогою лінійок і втяжок. Для виділень у літературі для вищої школи, а також наукових і виробничо-технічних виданнях користуються передусім розрядкою та курсивом, а також півжирним. У всьому виданні треба дотримуватись єдиної системи

Занадто велика кількість виділень у тексті не тільки не сприяє зосередженню уваги читача, але, навпаки, розпорошує її внаслідок строкатості набору.

У тексті окремі слова, терміни, визначення найчастіше виділяють світлим курсивом, нові поняття— розрядкою. Прізвища в основному тексті дають шрифтом без виділень; у бібліографічних списках— розрядку, рідше курсив або пів жирний. Для кожного виду спеціальної літератури існує своя система виділень. Так для кожної групи виділень (логічних, термінологічних тощо). Шрифтові виділення застосовують ширше, бо вони простіші для сприймання. У наукових виданнях з біології, палеонтології латинські назви родини, роду та виду подають курсивом, прізвище автора— в розрядку, скорочені слова— шрифтом основного тексту.

2.3. Інші методи інтелектуального аналізу тексту

Узагальнення текстових даних є процесом створення стислого представлення великої кількості даних. Під час узагальнення здійснюється пошук найбільш релевантних до теми речень і, як наслідок, зменшення об'єму текстових даних без втрати інформації. Такий метод може використовуватись для: обробки інформації з агрегаторів новин; генерування звітів; сумісної роботи з поштовими клієнтами; виділення інформації про події.

Класифікацією називається процес знаходження певних ознак класу (поміток) серед частково структурованих даних. Нехай, до прикладу, менеджер хоче

встановити, чи купуватиме замовник комп'ютер. Тоді класами для покупок будуть значення "Так" і "Ні". Під час класифікації створюється класифікатор на відповідних тестових даних з наступним означенням міток класів. Класифікація тексту є процесом присвоєння категорії новому текстовому документові. Таким чином можна розподілити текстові документи відповідно до належності певним класам. Області використання такої обробки тексту: бізнес; медицина; юриспруденція; суспільні науки.

Кластеризація – це процес групування елементів даних споріднених типів у окремі кластери . Таким чином, кластеризація зменшує часові затрати при пошуку інформації. Інформація, що не належить до жодного кластеру, вважається нерелевантною. Можливі області застосування такої методики: розпізнавання образів; аналіз зображень; Таксонометрія– виділення тематики з тексту. Таким чином, у даному огляді описано і класифіковано наявні на сьогодні методики інтелектуального аналізу текстових даних. Серед них – отримання (виділення) інформації, узагальнення, кластеризація. Всі вони можуть бути застосовані відповідно до задач, які ставляться при інтелектуальному аналізу текстових даних.

2.4. Існуючі програмні додатки для перевірки ПЕОМ на наявність ІЗОД, їх переваги та недоліки

Програмних додатків призначених спеціально для пошуку інформації з обмеженим доступом, які зараз використовуються немає, відповідно моніторинг ПЕОМ здійснюється вручну. Однак, існують додатки, які дозволяють за тими, чи іншими параметрами, як правило це назви файлів та їх вміст, здійснити пошук файлів, що містять інформацію з обмеженим доступом на персональному комп'ютері. Яскравими прикладами таких програмних додатків є: *FileSearchy*, *Folder Find Text* та навіть звичайний пошук у *Windows* (рис. 2.2, рис. 2.3, рис 2.4).

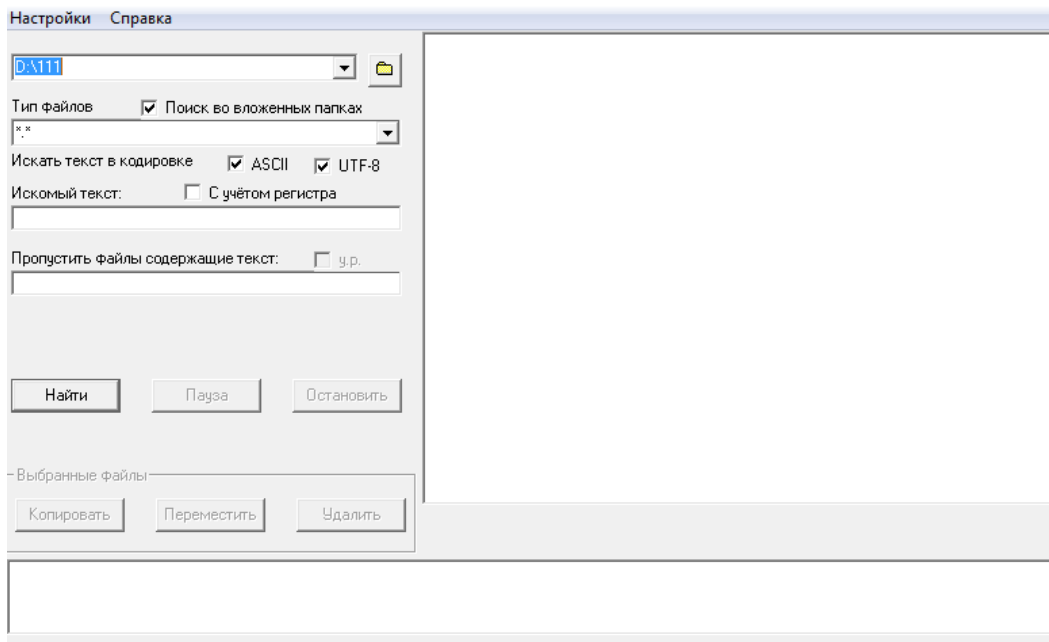


Рис.2.2. Программный додаток для пошуку файлів *Folder Find Text*

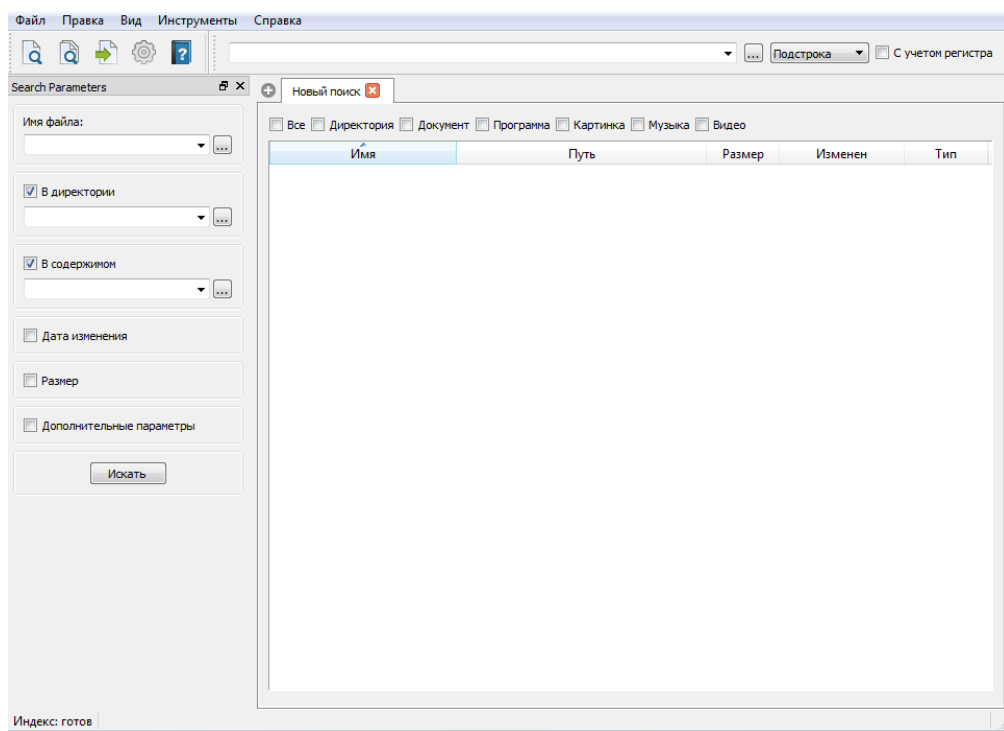


Рис.2.3. Программний додаток для пошуку файлів *FileSearchy*

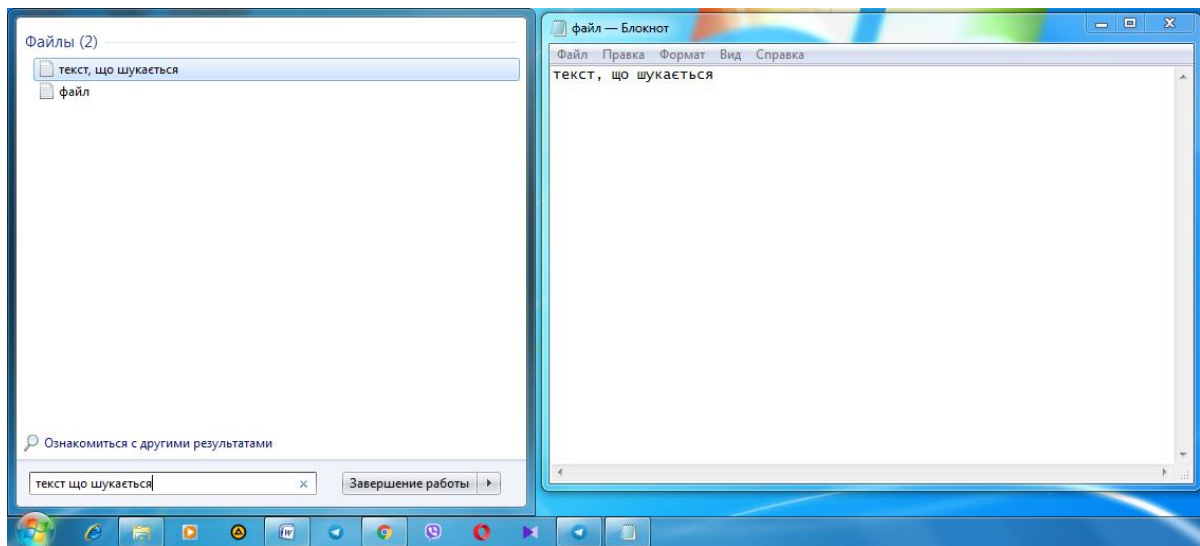


Рис.2.4. Пошук файлів у *Windows*

Представлені, програмні додатки не повністю відповідають вимогам щодо перевірки, оскільки призначені лише для пошуку файлів. Наприклад не здійснюють пошук файлів з форматом відмінним від *.txt* або *.doc*.

Тобто, щоб здійснити повноцінний пошук потрібних файлів необхідно використати декілька різних програм, що є недоцільним через велику затрату часу на перевірку. Багато програмних додатків дозволяють здійснювати багатоцільовий пошук інформації, але для оптимальної перевірки ПЕОМ на наявність ІзОД ніякий з розглянутих додатків не може використовуватись через недостатній функціонал та відсутність відповідного словника, за яким здійснюється пошук інформації на ПЕОМ.

Кожен з розглянутих програмних додатків має і свої переваги. Наприклад,

Folder Find Text - здійснює пошук файлів з різними типами кодування. Але не зберігає заданий словник та не відкриває файли, що було знайдено а також не здійснює пошук у файлах після їх архівації, що дає можливість приховати від пошуку інформацію. *FileSearchy* - здійснює пошук лише у вузьких директоріях, що суттєво ускладнює процес сканування, а також не працює на *Windows XP*, *Windows Vista* і інших старіших версіях *Windows*, які часто встановлені на ПЕОМ у ЗСУ.

2.5. Вибір підходу до виконання завдання

Порушення правил поведження з ІЗoД можна логічно розділити на два типи, а саме: неумисні – коли користувач неумисно порушує встановлені правила безпеки через впливи людського фактору та неуважність, та умисні – коли користувач ПЕoМ усвідомлено намагається порушити правила безпеки з метою доступу до інформації до якої він не має правомірного доступу, або ж передачі даних третім особам які не мають відповідного доступу.

Логічно допустити, що при неумисному порушенні правил безпеки поведження з ІЗoД, користувач не буде ніяким чином шифрувати, переписувати, кодувати чи будь-яким іншим чином видозмінювати інформацію у зазначених файлах та документах. А користувач, який виконує це умисно, може докласти певних як незначних так і значних зусиль для певного видозмінення інформації (змінити кодування, формат, назву, передати зміст іншими словами тощо), що може значно ускладнити запобігання та виявлення таких порушень.

Відповідно до видів порушень правил поведження з ІЗoД можна також виділити методи запобігання, які можна умовно окреслити як Прямий та Опосередкований методи аналізу текстових даних на наявність інформації з обмеженим доступом.

Історично відомо багато способів шифрування текстової інформації шляхом кодування текстових даних, які лягли в основу ідей сучасних методів шифрування. Також в історичних даних згадуються способи різноманітного маскуванню інформації під допустиму. Наприклад формули, коли для читання правильної інформації потрібно зчитувати першу букву кожного слова, або букву послідовну за номером кожного слова для складання правильних слів, що ускладнює аналіз такої інформації так як читання такої інформації, яка зовнішньо може бути схожа на дитячу казку, статтю з газети чи листування старих друзів може підкорятися певним формулам, які відомі лише обмеженому колу осіб, які користуються таким підходом. Кількість різноманітних подібних формул не піддається підрахунку через велику кількість різноманітних факторів.

Якщо порушення в плані умисної зміни кодування файлів доволі легко виявити звичайними програмними засобами проаналізувавши кодування файлу на відповідність певним стандартам, то опосередкований спосіб маскуванню буду дуже складно проаналізувати програмними засобами не знаючи формули, а метод можливого підбору формули по невизначеній множині різноманіття може зайняти проміжок часу який максимально наближений до нескінченності.

2.5.1. Метод виявлення опосередкованих порушень правил поводження з Ізод

Людина або спеціаліст при виявленні такого методу має більше шансів розгадати істинний зміст замаскованого тексту, спираючись на свій життєвий досвід та досвід роботи з подібними файлами, просто на інтуїцію та можливу уважність у пошуці закономірностей та просто певну удачу.

Саме тому для вирішення подібної задачі доцільно було б використати підхід штучного інтелекту та штучних нейронних мереж.

Штучні нейронні мережі – це комп'ютерні програми та системи, в основу реалізації яких було вкладено приклад біологічних нейронних мереж, що складають мозок тварин. Такі програмні системи здатні навчатися певним притаманним задачам на певних прикладах або ж без них і з часом та затратами на навчання покращувати свої результати у виконанні цих завдань. Основним недоліком та одночасно перевагою даних алгоритмів та систем є те що вони самі в процесі навчання виділяють певний набір характеристик вхідних даних, який впливає на результат роботи. Через це з одного боку вони можуть виявити певні закономірності, якими пов'яжуть два схожих об'єкта (наприклад як у розрізі задачі розпізнавання образів) та видати правильний результат, а з іншого – дані виявлення можуть бути помилковими, тому судження системи можуть бути невірними з точки зору людини.

З одного боку дослідження у сфері нейронних мереж йдуть протягом довгого часу, але досі залишається багато невідомого у сфері використання.

Як було вказано вище, штучні нейронні мережі працюють за принципами роботи схожими з роботою біологічних нейронних мереж тварин. Так як біологічні

нейрони працюють з електричними сигналами штучний нейрон працює з інформацією.

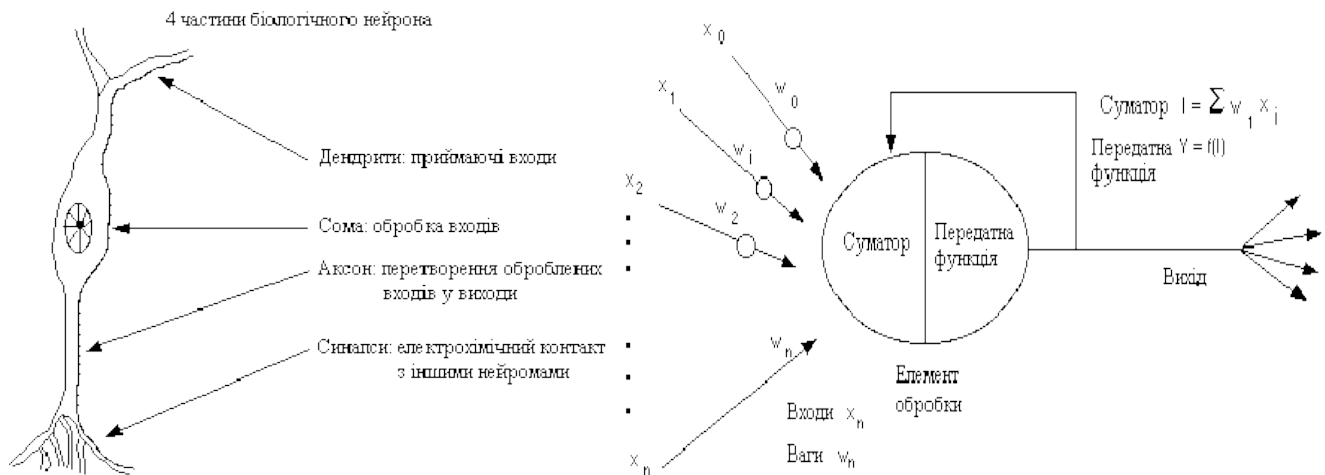


Рис. 2.5. Схема біологічного та штучного нейронів

Біологічний нейрон має дендрити – входи які, приймають імпульси та передають їх до тіла нейрони де вони обробляються, а аксон перетворює оброблені входи у виходи та передає інформацію через синапси до наступних нейронів. Синапси у свою чергу створюють певну хімічну речовину нейротрансмітери, яка збуджує або гальмує здатність наступного нейрона сприймати інформацію. Саме синапси навчаються у процесі мозкової активності що може впливати на час передачі інформації. Тобто інформація може розповсюдитися від одного нейрона до інших не в один проміжок часу а в різний залежно від тренуваності синапсів.

Штучний нейрон в свою чергу має входи, тіло обробки та виходи на інші нейрони. Структурно функції синапсів у штучних нейронах виконують вказані на схемі ваги. Але так як синапси впливають на час, то ваги є певним коефіцієнтом на який множиться вихід попереднього нейрона у схемі. Подібно до синапсів біологічних нейронних мереж ваги штучних нейронних мереж і є основним об'єктом тренувань. Мережа таких штучних нейронів і називається штучною нейронною мережею. За обробку даних відповідає передатна функція, або ж функція активації.

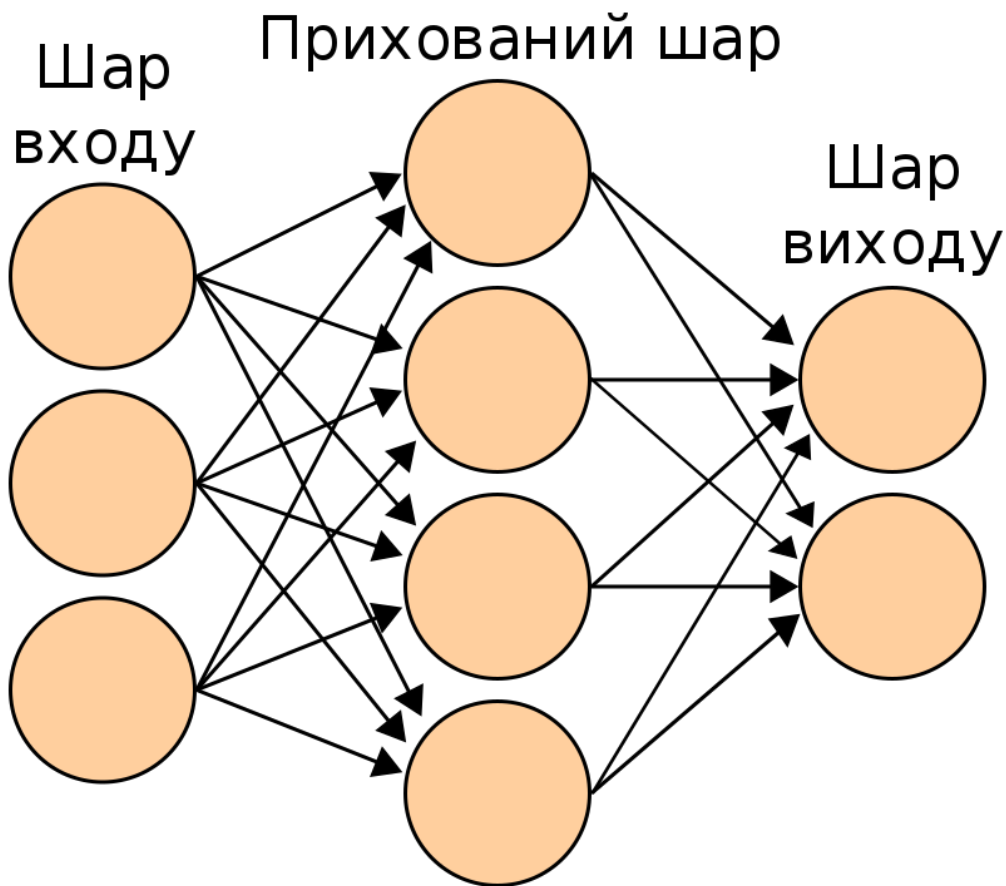


Рис. 2.6. Загальний вигляд штучої нейронної мережі

У загальному вигляді штучна мережа має: шар входу – даний шар призначений для вхідної інформації, яку буде обробляти мережа; шар виходу – для виводу рішення нейронної мережі; прихований шар – пов’язаний з шарами входу та виходу, у якому безпосередньо виконується обробка даних. Слід також зауважити, що даний шар може мати різну кількість шарів нейронів, які можуть відрізнятися один від одного за типом функції активації, кількістю нейронів та зв’язків.

Від кількості та типів нейронів у прихованому шарі зростає розмір нейронної мережі, що значно впливає на час її роботи та навчання, та впливає на точність результатів роботи мережі.

Основний алгоритм вирішення задачі складається з таких етапів:

- Збір даних для навчання;
- Підготовка і нормалізація даних;

- Вибір топології мережі;
- Експериментальний підбір характеристик мережі;
- Експериментальний підбір параметрів навчання;
- Власне навчання;
- Перевірка адекватності навчання;
- Коректування параметрів, остаточне навчання;
- Вербалізація мережі з метою подальшого використання.

Збір даних та підготовка даних до навчання є одним із найбільш складних та довготривалих етапів при роботі зі штучними нейронними мережами. Більша кількість даних якісно впливає на роботу мережі, але збільшує час необхідний на її навчання. Але великої кількості даних недостатньо – дані повинні бути також репрезентативними, тобто такими які відповідають дійсному положенню речей у предметній області, тобто такими які взяті з умов наближених до реальних, та несуперечливими.

Тобто для вирішення задач інтелектуального аналізу тексту на наявність ІЗоД потрібна велика кількість тексту, яка містить у собі ІЗоД. Для запобігання саме опосередкованим порушенням – потрібна велика кількість тексту та документів, які саме містять у собі умисне приховування ІЗоД. Так як видів опосередкованих умисних порушень є незліченна кількість, то вибірка повинна покривати прикладами найбільший із можливих видів різноманітних типів порушень, які підкоряються різним принципам, та мати подібних прикладів велику кількість.

Підготовка і нормалізація даних являє собою перетворення інформації до вигляду, який можна подати безпосередньо на входи нейронної мережі. Кожен такий запис за яким нейромережа повинна видати результат називається навчальним вектором. Навчальний вектор містить по одному значенню на кожен вхід мережі, і в залежності від типу навчання із учителем або без може мати по одному значенню на вихід нейронної мережі.

В даному випадку доцільно було б перетворити слова у тексті на відповідні їм числові значення від одного до кількості унікальних слів у навчальному датасеті.

Проблему виникнення нових невідомих слів у подальшій роботі можливо вирішити доданням новознайдених слів у сформований словник та присвоєння їм нового унікального числового значення.

Також можливо перетворити кожен символ у числове значення для входу на мережу, але такий спосіб нормалізації даних для подальшої обробки призведе до збільшення вхідного та прихованого шарів, що значно вплине на час обробки та навчання.

У випадку зі словами нейормережа скоріш за все навчиться працювати зі словами як логічними та змістовими структурами а не як з набором букв. Проте спосіб із символами можливо було б застосувати для читання бінарних файлів побайтово перетворивши байт в символ.

Вибір топології мережі залежить від поставленої задачі та даних для навчання. Для навчання з учителем потрібна наявність для кожного елемента вибірки «експертної» оцінки. Іноді одержання такої оцінки для великого масиву даних просто неможливо. У цих випадках природним вибором є мережа, що навчається без учителя. Також слід встановити відповідність вхідного шару нормалізованим даним для навчання, та вихідного шару для виводу параметрів оцінки.

В розрізі задачі в залежності від підходу до нормалізації та підготовки навчального датасета можливо було б обрати кількісний показник вхідного шару за максимальною кількістю слів у реченні, яке було знайдено у навчальній вибірці. Та подавати на вхід ціле речення та приймати рішення. Також можна було б збільшити даний показник в декілька разів для можливої обробки кількох речень або ж цілої сторінки.

В даному випадку вихідний шар повинен приймати рішення «Чи міститься в даному реченні ІЗод?» але також можна було б додати відповідь на питання «Де саме?». Вихідний шар залежить саме від того як саме і яку кількість параметрів повинне мати рішення нейронної мережі, щодо наданого на обробку об'єкта. Дане питання при навчанні з учителем повинне відповідати на кількість параметрів відповіді у навчальному векторі.

Експериментальний підбір параметрів мережі полягає у підборі кількості шарів у прихованому шарі, кількості нейронів у прихованих шарах, функції активації прихованих шарів. З одного боку більша кількість зв'язків покращує здатність мережі до узагальнення, а з іншого боку вона обмежена кількістю навчальних записів та даних. Після вибору топології потрібно підібрати параметри навчання мережі, такі як функція помилки та швидкість навчання що впливають на час сходження результатів до необхідного рівня якості. Дані параметри підбираються лише експериментальним шляхом керуючись при цьому критерієм завершення навчання.

Власне навчання мережі полягає у перегляді мережею навчальної вибірки у певному порядку. Мережі, які навчаються без учителя, переглядають вибірку тільки один раз. При навчанні з учителем мережа переглядає вибірку множини разів, при цьому один повний прохід по вибірці називається епохою навчання.

Зазвичай вибірку поділяють на два типи – тренувальна та валідаційна. Тренувальна вибірка подається для власне тренування. А тестова вибірка подається лише для перевірки роботи нейромережі і не впливає на її навчання. Якщо на валідаційних даних похибка зменшується то навчання виконується правильно. Якщо ж на тестових даних похибка зменшується а на валідаційних даних збільшується то нейромережа перестала навчатися на узагальнення і почала «запам'ятовувати» навчальні дані.

Так як кількість даних може бути дуже великою то і сам процес навчання зазвичай може тривати довгий час. У сучасному світі найчастіше для навчання нейронних мереж використовують графічні процесори відеокарт. Дані повністю завантажуються в оперативну пам'ять графічного процесора та проходять через мережу. При великих масивах даних завантажити всі може бути неможливо через розмір буфера пам'яті та прохід однієї епохи навчання може бути завеликим. Для вирішення цієї проблеми було винайдено підхід батч-нормалізації. Батч-нормалізація полягає у розбитті датасету на деякі відрізки, які батчами подаються для навчання нейронній мережі. І вже керуючись похибкою на батчі нейромережа робить крок змін ваг нейронів і саме навчається. Це дозволяє за одну повну епоху

декілька разів вирахувати похибку та провести зміни вагів нейронної мережі. Але так як дін можуть бути неоднорідними топомилка деяких батчів може відрізнятись від помилки інших батчів, що буде схожим на перетягування канату в різні боки. Тому навчання за повною епохою вираховує всю помилку для всього датасету і робить крок у більш точному напрямку але займає багато часу, в той час як батч-нормалізація робить більшу кількість кроків, але вектор розвитку може бути не завжди точним.

Для роботи зі штучними мережами та їх навчанням зазвичай застосовуються такі фреймворки як TensorFlow (Python, JavaScript), PyTorch, Keras, Darknet (C++).

2.5.2. Метод виявлення прямих порушень правил поведження з Ізод

Даний метод базується на пошуці певних слів та словосполучень у назвах файлів та їх текстовому наповненню. Даний підхід зможе вказати які саме слова та словосполучення і в яких саме файлах було знайдено та доповісти інформацію оператору.

Для реалізації даного підходу пошуку підходить практично будь-яка сучасна мова програмування високого рівня. Сучасні мови програмування високого рівня поділяються на компільовані та інтерпретовані.

Компільовані мови програмування – мови програмування в яких сирцевий код перед виконанням повністю перетворюється в машинний код і записується у файл зі спеціальним розширенням та назвою. Кожна така програма компілюється під певну платформу та архітектуру процесора на машині розробника, та при переносі на машину клієнта ніяких програмних функцій виконувати не потрібно. Це запобігає розповсюдженню та розкриттю сирцевого коду як такого і дозволяє виконувати файли навіть зі знімних носіїв. Для внесення будь-яких змін у програму потрібно повторити процес компіляції. Проте потрібно мати зкомпільовану програму для кожної платформи окремо. Прикладами таких мов є C, C++, C#, Java.

Інтерпретовані мови програмування – це мови програмування в яких сирцевий код не перетворюється в машинний код, а за допомогою програми інтерпретатора виконується покроково на ПЕОМ клієнта. Для внесення змін у такі програми не

потрібно виконувати компіляцію а достатньо змінити скрипт та перезапустити програму. Основним недоліком даного підходу є те що зазвичай сирцевий код знаходиться на машині клієнта і його можна доволі просто прочитати та змінити. До того ж інтерпретований код виконується повільше ніж компільований.

Для вирішення даної задачі слід обрати компільовану мову програмування, для більшої швидкодії та захисту сирцевого коду. Також повинна бути можливість написання зручного інтерфейсу користувача, який буде безпосередньо виконувати перевірку.

Даний підхід є більш простим, так як алгоритм аналізу не потребує складних зусиль, та затрат людських ресурсів однак може не вирішувати всіх завдань.

Висновок до розділу 2:

Не зважаючи на те що у перспективі застосування штучних нейронних мереж може вирішити більший спектр завдань він потребує завеликих затрат часу та роботи для формування навчальної та тестових вибірок, часу для тренування нейронних мереж та експериментів з параметрами навчання. Тому для вирішення даної задачі доцільніше буде використати прямий метод пошуку по словнику термінів та словосполучень, які натякають або ж містять у собі ІЗoД.

РОЗДІЛ 3

РОЗРОБКА ПРОГРАМНОГО МОДУЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ФАЙЛІВ ДЛЯ ПОШУКУ ІНФОРМАЦІЇ З ОБМЕЖЕНИМ ДОСТУПОМ

Даний розділ містить результати роботи над програмною реалізацією програмного модуля інтелектуального аналізу текстових файлів для пошуку інформації з обмеженим доступом, що створювалась.

Кафедра КІТ				НАУ 21 12 98 000 ПЗ			
<i>Виконав</i>	<i>Ревтюх Б.О.</i>			РОЗРОБКА ПРОГРАМНОГО МОДУЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ФАЙЛІВ ДЛЯ ПОШУКУ ІНФОРМАЦІЇ З ОБМЕЖЕНИМ ДОСТУПОМ	<i>Літера</i>	<i>аркуш</i>	<i>аркушів</i>
<i>Керівник</i>	<i>Климова А.С.</i>					33	12
<i>Консульт.</i>							33
<i>Н. контроль</i>	<i>Райчев І.Е.</i>				УС-201Мз		122

3.1. Створення словника ключових слів ПСІ

Для того, щоб створити повноцінний додаток призначений для пошуку інформації з обмеженим доступом на ПЕОМ потрібно створити словник - перелік слів, визначень, за якими буде проводитись пошук.

Інформація з обмеженим доступом = інформація, доступ до якої має лише обмежене коло осіб і оприлюднення якої заборонено розпорядником інформації відповідно до закону. Обмеження доступу до інформації здійснюється в інтересах національної безпеки або охорони законних прав фізичних та юридичних осіб. Обмежується доступ до інформації, а не до документу. Відповідно, якщо в одному документі міститься відкрита і закрита інформація, перша може бути надана на ознайомлення зацікавленій особі у вигляді окремого документу [14,15].

В ст. 29 Закону України «Про інформацію» вказано, що інформація з обмеженим доступом може бути поширена, якщо вона є суспільно необхідною, тобто є предметом суспільного інтересу, і право громадськості знати цю інформацію переважає потенційну шкоду від її поширення[16].

В свою чергу, предметом суспільного інтересу вважається інформація, яка свідчить про загрозу державному суверенітету, територіальній цілісності України; забезпечує реалізацію конституційних прав, свобод і обов'язків; свідчить про можливість порушення прав людини, введення громадськості в оману, шкідливі екологічні та інші негативні наслідки діяльності (бездіяльності) фізичних або юридичних осіб тощо. [17,18].

Таблиця 3.1 Словник ключових слів ПСІ

1	2
Слово, за яким здійснюється пошук	Пункт у ПСІ
Бойова готовність, завдання управління	Відомості, що розкривають зміст завдань органів військового управління, військ щодо бойової готовності, оперативної підготовки та застосування військ

Приведення у бойову готовність, бойова готовність, інструкція чергового частини	Відомості про зміст інструкцій чергових частин (структурних підрозділів Міністерства оборони України та Генерального штабу Збройних Сил України), що стосуються приведення у бойову готовність.
План командно-штабних тренувань, план тактико-стройових занять	Відомості про зміст планів проведення командно-штабних тренувань і тактико-стройових занять у військовій частині.
Результати навчань, результати тренувань, результати тренувань	Відомості про результати проведених навчань, тренувань.
Підготовка навчань Підготовка тренувань; Підготовка маневрів; Підготовка воєнних ігор; Проведення навчань; Проведення тренувань; Проведення маневрів;	Відомості про підготовку та проведення навчань (тренувань, маневрів, воєнних ігор).
Перевірка стану бойового чергування; План перевірки;	Відомості щодо планування та результати здійснених перевірок стану бойового чергування, оперативно-чергової (чергової) служби
Військову підготовку; Перевірку техніки; Перевірка техніки; Вивчення військовозобов'язаних;	Відомості про результати вивчення військовозобов'язаних та перевірку техніки національної економіки України.
Чисельний склад; Резервістів;	Відомості про чисельний склад резервістів
Укомплектованість, бойові можливості	Відомості про склад, укомплектованість, забезпеченість, бойові можливості, положення і дії (характер діяльності) військ (сил) до роти (батареї) та їм рівних включно.

Слово, за яким здійснюється пошук	Пункт у ПСІ
Наказ; Особовий склад;	Сукупність відомостей, що містяться в наказах по особовому складу щодо військовослужбовців.
Спеціальні набори; Дострокові випуски;	Відомості про спеціальні набори і дострокові випуски з військових навчальних закладів
Чисельність; Офіцер запасу;	Відомості про чисельність підготовлених офіцерів запасу і тих, які готуються.
Службова характеристика; Автобіографія; Подання; Нагородні листи; Атестації; Особова справа;	Зведені відомості персонального обліку військовослужбовця стосовно проходження ним військової служби, його службової характеристики, автобіографія, подання, нагородні листи, атестації, оцінні матеріали, особова справа військовослужбовця.
Прийому; Відбору; Робота комісії;	Відомості про результати роботи комісій з відбору, прийому і випуску слухачів (курсантів).
Результатів підготовки військ;	Відомості про роботу з офіцерськими кадрами, сержантами та старшинами, які містять у собі узагальнені дані результатів підготовки військ.
Укомплектованість;	Відомості мирного часу про укомплектованість особовим складом (за військову частину і вище).
Плани комплектування;	Відомості про плани комплектування військ особовим складом строкової служби (за військову організаційну).
Нагородження відзнакою; Нагороджений відзнакою; Нагороджена відзнакою; Номер зброї; Заводський номер;	Відомості про осіб, що зазначаються в наказах по особовому складу щодо нагородження відзнакою Міністерства оборони України “Вогнепальна зброя”, “Холодна зброя” (в тому числі їх персональні дані), назва (марка) та заводський номер зброї.

Слово, за яким здійснюється пошук	Пункт у ПСІ
Результати розвідувальної Діяльності; Розвіддані; Даними розвідки; Розвідка; План розвідки;	Відомості про зміст, плани, організацію, фінансування, матеріально-технічне забезпечення, засоби, форми, методи, оцінку і результати розвідувальної діяльності.
Причетність; Причетний; Причетна; Бере участь; Розвідувальна інформація; Місцезнаходження; Розташовується; Розташована; Розташувати;	Відомості про причетність військових частин до діяльності, пов'язаної із збиранням, обробкою, диспетчеризацією та використанням розвідувальної інформації, а також до діяльності, пов'язаної з підготовкою фахівців для органів розвідки, їх місцезнаходження.
Виїзд за кордон; Відбір кандидатів;	Відомості про вивчення особового складу під час прийому на службу до органів розвідки, виїзду за кордон, відбору кандидатів для вступу до ВВНЗ розвідки.
Военно-політична обстановка;	Відомості про воєнно-політичну обстановку, викладені у звітах, доповідях за тиждень (добу).
Первинна обробка; Розвід матеріал;	Відомості про результати первинної обробки розвідматеріалів та відомостей.
Втрата носіїв інформації; Для службового користування; ДСК; дск;	Відомості про втрату матеріальних носіїв інформації, які містять службову інформацію та яким присвоєно гриф обмеження доступу “Для службового користування”.
Бойовий склад; Штат; Забезпеченість ; Порядок комплектування; Озброєння; Кількість озброєння;	Відомості про організацію, бойовий склад, штат, чисельність особового складу, кількість озброєння, військової техніки, транспорту, засобів матеріально-технічного забезпечення, укомплектованість і забезпеченість, порядок комплектування.

Слово, за яким здійснюється пошук	Пункт у ПСІ
Бойовий склад; Штат; Забезпеченість ; Порядок комплектування; Озброєння; Кількість озброєння;	Відомості про організацію, бойовий склад, штат, чисельність особового складу, кількість озброєння, військової техніки, транспорту, засобів матеріально-технічного забезпечення, укомплектованість і забезпеченість, порядок комплектування.
Призначений; Контррозвідувальний;	Відомості, викладені в службовому листуванні щодо розвідувальної діяльності (запитах, звітах, доповідях, донесеннях, інформаційно-аналітичних довідках, інформаційних матеріалах) які не становлять державної таємниці, але розкривають призначення структурного підрозділу для здійснення розвідувальної, оперативно-розшукової діяльності або виконання окремих контррозвідувальних заходів.
Новітніх досягнень; Криптологія;	Відомості про можливості застосування в розвідувальній діяльності результатів аналізу новітніх досягнень у криптології.
Експлуатація військової техніки; Ремонт ОВТ; Ремонт озброєння;	Відомості, викладені у звітах, доповідях, донесеннях, планах тощо про експлуатацію (ремонт) озброєння та військової техніки, які входять до номенклатури розвідки.

Слово, за яким здійснюється пошук	Пункт у ПСІ
Бойовий склад; Штат; Забезпеченість ; Порядок комплектування; Озброєння; Кількість озброєння;	Відомості про організацію, бойовий склад, штат, чисельність особового складу, кількість озброєння, військової техніки, транспорту, засобів матеріально-технічного забезпечення, укомплектованість і забезпеченість, порядок комплектування.
Обсяг озброєння; Обсяг військової техніки;	Відомості про номенклатуру, обсяги постачання, потребу і обсяги озброєння, військової та спеціальної техніки, які входять до номенклатури розвідки, знаходяться на озброєнні та не мають грифу секретності.
Ремонтні підприємства; Ремонтне підприємство; Номенклатура розвідки;	Зведений перелік ремонтних підприємств, їх місцезнаходження та можливості щодо ремонту військової техніки та озброєння, що входять до номенклатури розвідки.
Заходи ТЗІ; Технічний захист інформації;	Відомості про плани вдосконалення заходів ТЗІ.
Створення комплексів; Випробування комплексів; Проектування комплексів; Обстеження комплексів; Дослідна експлуатація; СЗІ;	Відомості про зміст організаційних документів із питань проектування, категоріювання, обстеження, створення, випробування комплексів технічного захисту інформації та дослідної експлуатації системи захисту інформації на об'єктах інформаційної діяльності.

1	2
Слово, за яким здійснюється пошук	Пункт у ПСІ
Технічна характеристика документу; Криптографічний захист;	Відомості про порядок використання, поводження, технічні характеристики ключових документів до засобів криптографічного захисту службової інформації
Навчально-бойової діяльності; Морально-психологічне забезпечення;	Відомості, які за сукупністю показників розкривають зміст морально-психологічного забезпечення навчально-бойової діяльності конкретно визначених військових частин, з'єднань і вище.
стан підготовки військ; стан підготовки військ (сил);	Відомості про морально-психологічне забезпечення, які за сукупністю показників розкривають реальний стан підготовки військ (сил) до застосування
МПЗ;	Морально-психологічне забезпечення військ (сил).
Постачання; Часткове постачання; Вилучення;	Відомості про часткове постачання або вилучення продовольства, речового майна, ракетного палива, основних видів пального та мастильних матеріалів, технічних засобів, інших матеріальних засобів
матеріальне забезпечення передислокації;	Відомості про матеріальне забезпечення передислокації з'єднань, військових частин, закладів та установ Збройних Сил України на воєнний час.
Вартість; Кількість; Система охорони; Спортивна зброя;	Відомості про місця зберігання та систему охорони спортивної зброї у закладах фізичної культури і спорту Міністерства оборони України.

1	2
Слово, за яким здійснюється пошук	Пункт у ПСІ
Матеріальна база служби військ; Склад добового наряду;	Відомості про стан матеріальної бази служби військ, склад добового наряду.
Розподіл випускників;	Відомості про розподіл випускників вищих військових навчальних закладів та військових навчальних підрозділів вищих навчальних закладів

Слово, за яким здійснюється пошук	Пункт у ПСІ
Стрілецької зброї; Боєприпаси;	Відомості про наявність стрілецької зброї та боєприпасів на арсеналах, базах, складах.
Модернізовані зразки озброєння;	Відомості про технічне обслуговування і ремонт нових та модернізованих зразків озброєння та військової техніки.
Відвантаження продукції;	Відомості про одержання відвантаженої на замовлення Міністерства оборони України продукції.
Приймання літаків; Постачання літаків;	Зведені відомості про приймання та постачання літаків.
Обсяг продукції військового призначення;	Інформація про обсяги продукції військового призначення, яка контролюється в поточному році.
Конструкторська документація;	Зведені відомості про наявність на підприємствах (в установах, організаціях) конструкторської документації на озброєння та техніку, яка є державною власністю.
Рух засобів охорони; Переміщення техніки охорони;	Відомості про облік, наявність і рух технічних засобів охорони.

Окремо до словника необхідно додати наступний перелік слів:

- Таємно;
- ДСК;
- Цілком таємно;
- Для службового користування;
- Особливої важливості;

3.2. Створення програмного додатку, для пошуку інформації з обмеженим доступом.

При виборі мови програмування потрібно, щоб вона відповідала наступним вимогам:

- об'єктно-орієнтованість;
- багатопоточність;[19].

Враховуючи дані особливості, а також функціональні можливості, наявна кількість фреймворків та рішень, для здійснення швидкої та гнучкої розробки у якості мови програмування для реалізації обрано об'єктно-орієнтовану мову програмування високого рівня *C#*.

Серед наявних середовищ розробки мовою програмування *C#* для розробки програмного модулю обрано безкоштовне інтегроване середовище розробки *Visual Studio Community 2019*. Дане середовище має зручний користувацький інтерфейс, велику кількість допоміжних інструментів розробки, а також підтримку значну кількість фреймворків.

Будь-який додаток доводиться змінювати з часом – змінюються вимоги, додаються нові. Чим швидше і зручніше можна внести зміни в існуючий функціонал – тим гнучкішою буде система.

Повинна бути можливість додавати в систему нові сутності та функції, не порушуючи її основної структури. На початковому етапі в систему має сенс закладати лише основний і самий необхідний функціонал. Але при цьому архітектура повинна дозволяти легко нарощувати додатковий функціонал в міру

необхідності. Причому так, щоб внесення найбільш ймовірних змін вимагало найменших зусиль [20].

Гнучкість і можливість розширення забезпечується використанням принципу *SOLID – The Open Closed Principle (OCP)* – так званий принцип відкритості/закритості, який вимагає, щоб програма була закрита для змін і відкрита для розширення [21, 22].

Для реалізації цього принципу в рішенні використовуються інтерфейси та абстрактні класи. Це значно полегшує розробку, запобігає дублюванню коду та дозволяє легко та швидко додавати новий функціонал. Класи для обробки команд створені на основі таких патернів проектування як *Template Method* (шаблонний метод) та *Command* (команда). Також у них реалізовано породжуючий патерн *Prototype* (прототип), який дозволяє створювати екземплярами класів їх клонів. Цю їх властивість застосовано в спеціальному менеджері команд, який представляє собою *Fabric* (фабрику) для створення екземплярів класів для обробки команд [23].

При розробці програмного додатку враховувалось те, що даний додаток можливо використовувати не тільки для пошуку інформації з обмеженим доступом, але і інших файлів на ПЕОМ за рахунок створення вікна вводу слів у словник вручну. Чим забезпечується багатофункціональність додатку.

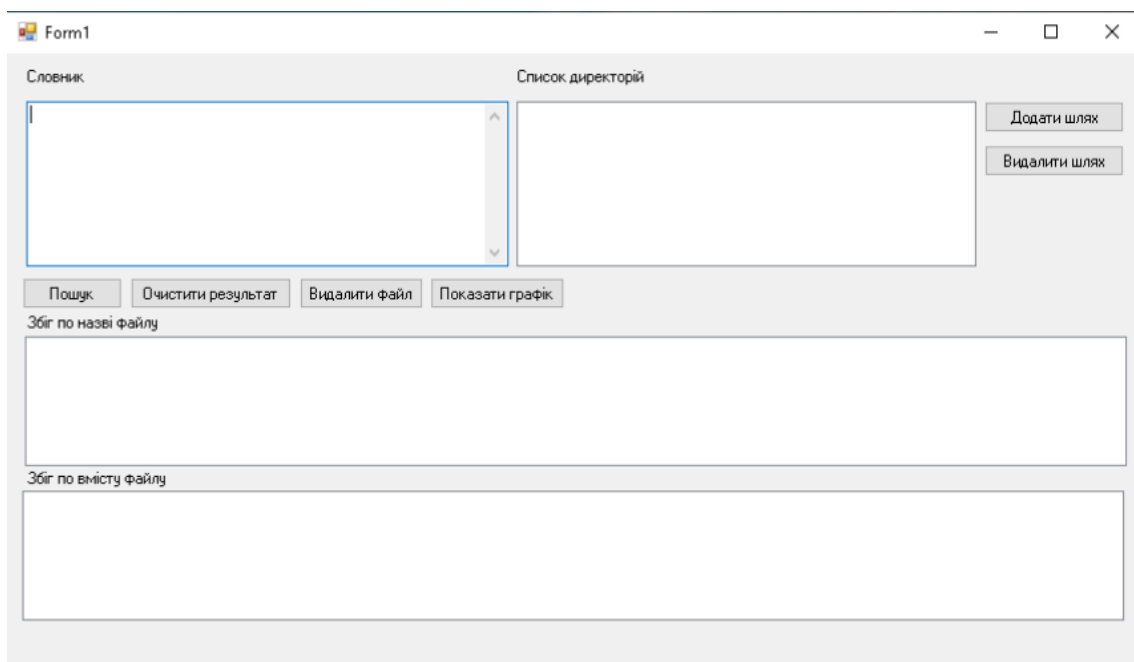


Рис.3.1. розроблений програмний додаток для пошуку файлів

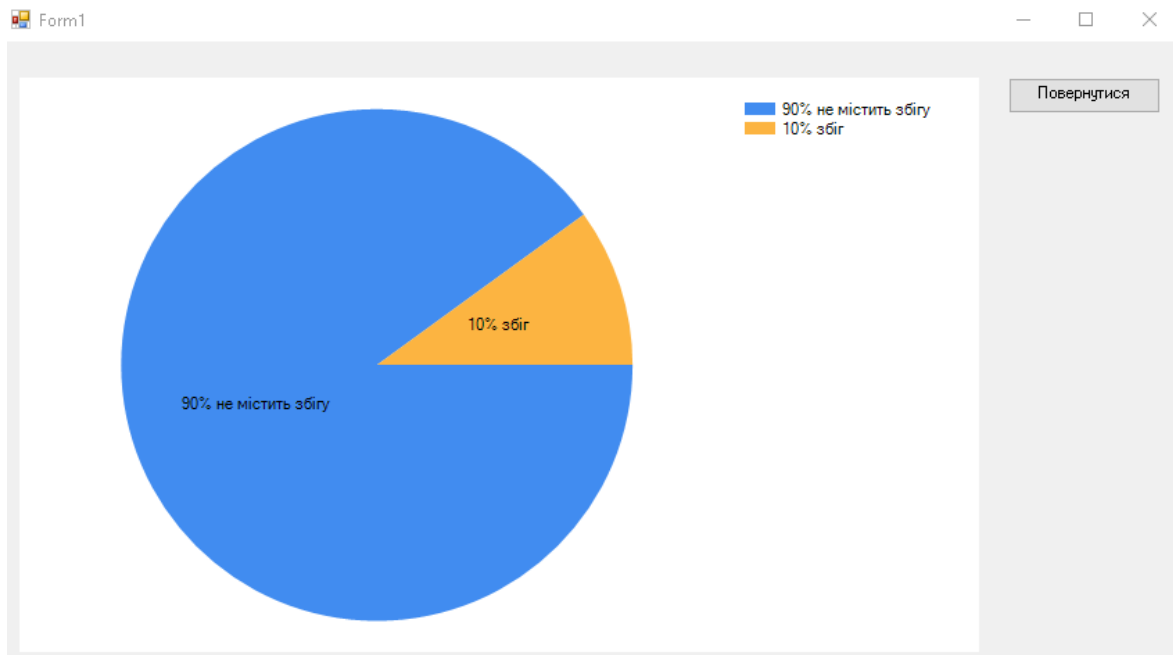


Рис.3.2. приклад виведення діаграми

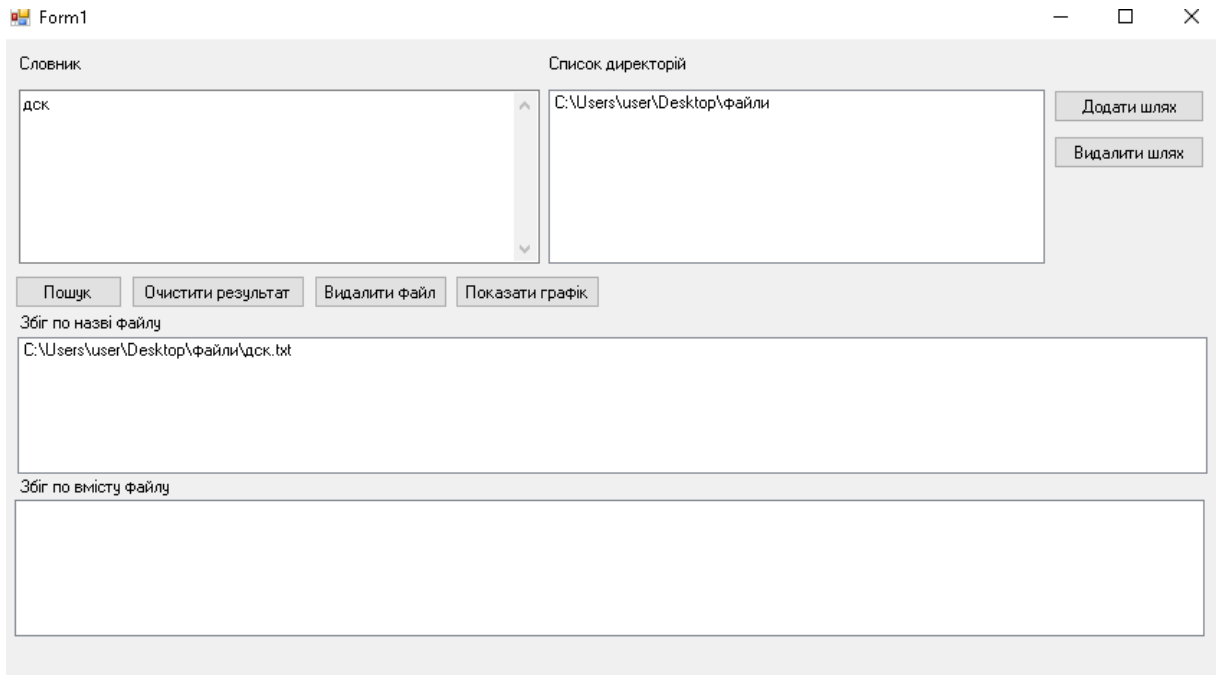


Рис.3.3. приклад роботи програмного додатку

Висновок до розділу 3:

Отже, розглянуті програмні програмні додатки для пошуку інформації на ПЕОМ задовольняють потреби користувача лише частково, прийнято рішення та розроблено програмний додаток з урахуванням недоліків розглянутих додатків та з використанням створеного словника ключових слів.

РОЗДІЛ 4

ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНОГО ПІДХОДУ ДО АВТОМАТИЗАЦІЇ ПРОЦЕСУ ПЕРЕВІРКИ ПЕОМ НА НАЯВНІСТЬ ІзОД

4.1. Інструкція користувачу

Інтерфейс програми є дуже простим та зрозумілим. При роботі з програмним додатком необхідно чітко усвідомлювати мету пошуку файлу. Знати принаймні одне слово з назви файлу або з тексту файлу. Слово за яким здійснюється пошук необхідно додати у словник. Після чого вибрати директорію пошуку файлу (можна обрати всю пам'ять ПЕОМ). Швидкість пошуку на пряму залежить від кількості файлів в обраній директорії та їх розміру. Поле вибору директорії створене для прискорення процесу пошуку за умови що користувач знає приблизне місцезнаходження файлу. По завершенню пошуку результат виводиться у окреме поле та зберігається у окрему папку для створення подальшого звіту про проведений пошук. Окремо виводяться файли у яких було виявлено збіг зі словником у назві та у змісті.

Кафедра КІТ				НАУ 21 12 98 000 ПЗ			
Виконав	Ревтюх Б.О.			ОЦІНКА ЕФЕКТИВНОСТІ ЗАПРОПОНОВАНОГО ПІДХОДУ ДО АВТОМАТИЗАЦІЇ ПРОЦЕСУ ПЕРЕВІРКИ ПЕОМ НА НАЯВНІСТЬ ІзОД	Літера	аркуш	аркушів
Керівник	Климова А.С.					45	7
Консульт.							45
Н. контроль	Райчев І.Е.				УС-201Мз		122

4.2. Міжнародний стандарт оцінки якості ПЗ ISO 9126

ISO 9126 – це міжнародний стандарт, що визначає оціночні характеристики якості програмного забезпечення. [24]

Стандарт поділяється на три частини, що описують наступні питання:

- модель якості;
- зовнішні метрики якості;
- внутрішні метрики якості.

Стандарт пропонує використовувати для опису внутрішнього та зовнішнього якості ПЗ багаторівневу модель. На верхньому рівні виділено 6 основних характеристик якості ПЗ. Кожна характеристика описується за допомогою кількох вхідних у неї атрибутів. Для кожного атрибута визначається набір метрик, що дозволяють його оцінити[25,26].

Визначення цих характеристик і атрибутів за стандартом *ISO 9126*:

- функціональність (*functionality*). Здатність ПЗ в певних умовах вирішувати задачі, потрібні користувачам. Визначає, що саме робить ПЗ, які задачі воно вирішує;
- функціональна придатність (*suitability*). Здатність вирішувати потрібний набір задач;
- точність (*accuracy*). Здатність видавати потрібні результати;
- здатність до взаємодії (*interoperability*). Здатність взаємодіяти з потрібним набором інших систем;
- захищеність (*security*). Здатність запобігати неавторизованому, тобто без вказівки особи, що намагається його здійснити, і недозволеному доступу до даних і програм;
- надійність (*reliability*). Здатність ПЗ підтримувати визначену працездатність у заданих умовах;
- стійкість до відмов (*fault tolerance*). Здатність підтримувати заданий рівень працездатності при відмовах і порушеннях правил взаємодії з середовищем;

- здатність до відновлення (*recoverability*). Здатність відновлювати визначений рівень працездатності й цілісність даних після відмови, необхідні для цього час і ресурси;

- зручність використання (*usability*). Здатність ПЗ бути зручним у навчанні та використанні, а також привабливим для користувачів;

- зрозумілість (*understandability*). Показник, зворотний до зусиль, які затрачаються користувачами на сприйняття основних понять ПЗ та усвідомлення їх застосовності для розв'язання своїх задач;

- зручність навчання (*learnability*). Показник, зворотний зусиллям, затрачуваним користувачами на навчання роботі з ПЗ

- зручність роботи (*operability*). Показник, зворотний зусиллям, що вживається користувачами для розв'язання своїх задач за допомогою ПЗ;

- привабливість (*attractiveness*). Здатність ПЗ бути привабливим для користувачів;

- продуктивність (*efficiency*) або ефективність. Здатність ПЗ при заданих умовах забезпечувати необхідну працездатність стосовно виділюваного для цього ресурсам. Можна визначити її і як відношення одержуваних за допомогою ПЗ результатів до затрачуваних на це ресурсів усіх типів;

- часова ефективність (*time behaviour*). Здатність ПЗ видавати очікувані результати, а також забезпечувати передачу необхідного об'єму даних за відведений час;

- ефективність використання ресурсів (*resource utilisation*). Здатність вирішувати потрібні задачі з використанням визначених об'ємів ресурсів визначених видів;

- зручність супроводу (*maintainability*). Зручність проведення всіх видів діяльності, пов'язаних із супроводом програм;

- аналізованість (*analyzability*) або зручність проведення аналізу. Зручність проведення аналізу помилок, дефектів і недоліків, а також зручність аналізу необхідності змін і їх можливих наслідків;

- зручність внесення змін (*changeability*). Показник, зворотний трудозатратам на виконання необхідних змін;
- стабільність (*stability*). Показник, зворотний ризику виникнення несподіваних ефектів при внесенні необхідних змін;
- зручність перевірки (*testability*). Показник, зворотний трудозатратам на проведення тестування і інших видів перевірки того, що внесені зміни привели до потрібних результатів.
- переносність (*portability*). Здатність ПЗ зберігати працездатність при перенесенні з одного оточення в інше, включаючи організаційні, апаратні й програмні аспекти оточення;
- адаптованість (*adaptability*). Здатність ПЗ пристосовуватися різним оточенням без проведення для цього дій, крім заздалегідь передбачених;
- зручність встановлення (*installability*). Здатність ПЗ бути встановленим або розгорнутим у визначеному оточенні;
- здатність до співіснування (*coexistence*). Здатність ПЗ співіснувати з іншими програмами у загальному оточенні, ділячи з ними ресурси.
- зручність заміни (*replaceability*) іншого ПЗ даним. Можливість застосування даного ПЗ замість інших програмних систем для вирішення тих же задач у певному оточенні [27].

Для оцінки програмного модулю згідно моделі представлений в стандарті *ISO 9126*, необхідно визначитися із метрикою для оцінки характеристик та атрибутів моделі.

Не існує заздалегідь визначених метрик для оцінки якості програмного забезпечення, так як кожен програмний компонент має свою прикладну сферу, метрику для оцінки його якості та ефективності потрібно підбирати.

Підберемо метрики для оцінки якості програмного модулю взаємодії бази даних з системою вступ з оптимізацією запитів керуючись стандартом *ISO 9126*.

Функціональна придатність – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – виконує понад 85% зазначених задач, середня 65% - 85%, низька – до 65%.

Здатність до взаємодії – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – взаємодіє з потрібним набором усіх інших модулів системи, середня – взаємодіє з потрібним набором деяких інших модулів системи, низька – не взаємодіє з іншими модулями системи.

Захищеність – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – гарантує неможливість компрометації даних, що передаються на час до коли вони стануть не актуальними, середня – гарантує неможливість компрометації даних, що передаються на час до коли вони стануть не актуальними але з допущенням невеликої ймовірності компрометації, низька – не гарантує неможливість компрометації даних.

Стійкість до відмов - оцінюється як Висока (В), Середня (С), Низька (Н). Висока – допускає 2% відмов на 1000 сесій, середня – 5% відмов на 1000 сесій, низька – понад 8% відмов на 1000 сесій.

Здатність до відновлення - оцінюється як Висока (В), Середня (С), Низька (Н). Висока – гарантовано відновлює визначений рівень працездатності й цілісність даних після відмови, середня – відновлює працездатність за визначених умов, низька – не відновлює працездатність після відмови.

Зручність використання – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – зручне у використанні та навчанні, привабливе для користувачів, середня – не зовсім зручне у використанні та навчанні, низька – не зовсім зручне у використанні та навчанні, не привабливе для користувачів.

Ефективність використання ресурсів – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – споживає менше 2% загального об'єму ресурсів на пристрої, де буде використовуватися, середня – менше 5%, низька – 10% і більше.

Зручність супроводу – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – ПЗ має високу доступність для аналізу, зручне для внесення змін, стабільне, зручне для тестування, середня – має високу доступність для аналізу, частково зручне для внесення змін, стабільне, не зручне для тестування, низька – має низьку доступність для аналізу, частково зручне для внесення змін, зручне для тестування.

Здатність до співіснування – оцінюється як Висока (В), Середня (С), Низька (Н). Висока – ПЗ без проблем співіснує з іншими програмами у загальному оточенні поділяючи з ними ресурси. Середня – ПЗ конфліктує з визначеним набором програм. Низька – ПЗ не передбачає роботу інших програм в одному загальному оточенні разом з ним.

Маємо набір метрик для оцінки програмного модулю організації взаємодії бази даних з системою вступу з оптимізацією запитів. Користуючись ними складемо таблицю оцінки якості програмного модулю.

4.3. Оцінка запропонованого підходу

Кожен розроблений програмний продукт підлягає ретельному тестуванню та оцінюванню його ефективності використання. Відповідно до цього виконується перевірка модулю на рахунок ефективності його використання. Оцінювання проводиться за певними критеріями які обираються відповідно до сфери застосування програмного продукту та його функціональності у визначеній предметній області.

Відповідно до розробленого програмного модулю, можливо винести такі критерії оцінки ефективності як:

- час обробки інформації;
- простота у використанні;
- актуальність у сфері використання.

Відповідно вищевказаних критеріїв можливо зробити адекватну оцінку ефективності модулю.

Час обробки інформації дорівнює приблизно декільком секундам, що забезпечує оперативність в роботі з додатком. Вся інформація про проведений пошук зберігається у додатковому файлі, що допомагає запобігти фальсифікації результату та уникненню від відповідальності.

Даний модуль має досить зручний інтерфейс користувача, який не є обтяженим для користувача і не потребує детального вивчення продукту для подальшої його

експлуатації. Графічний інтерфейс модулю відповідає всім вимогам користувача, щодо його простоти і ефективності.

На даний момент автоматизація процесу моніторингу ПЕОМ на наявність файлів що містять ДСК та/або ІзОД являється одним із актуальним питанням у сфері автоматизації процесу попередження витоку інформації у збройних сил. Тому даний модуль являється одним із можливих рішень питання контролю над обігом ДСК та ІзОД [26,27].

Висновок до розділу 4:

Отже, створивши словник для пошуку та програмний додаток, що здійснює пошук файлів було здійснено часткову автоматизацію процесу моніторингу ПЕОМ на наявність файлів що містять ДСК та/або ІзОД.

ВИСНОВКИ

Проаналізувавши існуючі методи інтелектуального аналізу текстових документів та програмні додатки, що здійснюють моніторинг наявності інформації з обмеженим доступом на незахищених ПЕОМ, можна побачити, що, не дивлячись на широту застосування, переваги та недоліки кожного з розглянутих програмних додатків існує ряд недоліків пов'язаних з їх застосуванням. Ці недоліки пов'язані з проблемами застосування техніки старого покоління та використанням програмного забезпечення з яким згадані вище додатки працюють лише частково виконуючи свої функції у неповному обсязі. До цих недоліків відносяться: несумісність додатків з певним типом кодування, відсутність оновлення. А отже, необхідно шукати способи часткового вирішення проблем на базі техніки та програмного забезпечення, що є в наявності. Проаналізувавши усі проблеми пропонується вирішення проблеми с автоматизацією обліку майна програмним шляхом.

Для представлення програмної реалізації запропоновано програмний додаток. Згідно цього, програмний додаток буде виконувати роль автоматизації робочого процесу, а функція управління покладається на оператора.

Розробка програмного додатку здійснювалась в послідовності, що визначена сучасними концепціями технологій комп'ютерного проектування.

Розроблений програмний додаток було протестовано. Створений програмний додаток показав стійку роботу при роботі з інформацією та перенесення на інший ПЕОМ.

СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ

1. О. І. Гарасимчук, В. Б. Дудикевич, В. А. Ромака, Комплексні системи санкціонованого доступу : навч. посібник / О. І. Гарасимчук, В. Б. Дудикевич, В. А. Ромака ; Львів. політех. - Л. : Вид-во Львів. політехніки, 2010. - 212 с
2. К.т.н., проф. А.Ю.Щеглов: публікація – / А. Ю. Щеглов, компьютерная безопасность.идентификация и аутентификация так ли все просто?–електронний ресурс, ЗАО НПП «Информационные технологии в бизнесе», режим доступу: www.npp-itb.spb.ru
3. Иванов, В.Г. Ідентифікація користувачів інформаційних систем: аналіз підходів / Иванов В.Г., Мазниченко Н.І. // Международная научная конференция MicroCAD : Секція №21 - Інформатика і моделювання - НТУ "ХПИ", 2012.
4. Комп'ютерні технології обробки інформації. / Під. ред. С.В. Назарова. - М.: Фінанси і статистика, 1995.,-166с.
5. Д. Веді Захист даних в комп'ютерних мережах / / Відкриті системи М, 1995, № 3, с. 12-18
6. Кручинин А. Специфікація *Aztec Code* (без *Small Aztec*) [Електронний ресурс]. – Режим доступа : <http://recog.ru/blog/standarts/6.html>.
7. Медведев А. Реальная Виртуальность. [Електронний ресурс]. – Режим доступа : <http://www.ixbt.com/peripheral/real-vrml.html>.
8. *Information technology – International symbology specification – Data Matrix (ISO/IEC 16022:2006) : ДСТУ – ISO/IEC 16022:2000 [Действителен с 2006г.]*. – *ISO copyright office, 2006.* – 133с. – (Міжнародний стандарт).
9. *Information technology – Automatic identification and data capture techniques – Aztec Code bar code symbology specification (ISO/IEC 24778:2008): ДСТУ – ISO/IEC 24778: 2008 [Дійсний з 2008.]*. – *ISO copyright office, 2008.* – 150с. – (Міжнародний стандарт).
10. Інтелектуальні системи автоматичного управління. / Под ред. І.М. Макарова, В.М. Лохіна - М.: Фізматліт, 2001.-124с.

11. Белов Г.В. Штриховое кодирование: технологии XXI столетия М.: Металлургия, 1998.-433с.
12. Безопасность компьютерных систем. Компьютерная преступность и ее предупреждение / М.С.Вертузаев, В.О.Голубев, О.И.Котляревский, О.М.Юрченко / Под ред. О.П.Снигерьева. - Запорожжя: ПВКФ "Павел", 1998.
13. Гуржий, А.М. Информатика та інформаційні технології: підручник / А. М. Гуржий, Н. І. Поворознюк, В. В. Самсонов. - Х. : Компанія СМІТ, 2003. - 352 с.
14. Д. Гроп. Методы идентификации систем. – М.:Мир,1979.-302с.
15. Л. Льюнг. Идентификация систем. – М.: Наука, 1991.-432 с.
16. Растринин Л.А., Маджаров Н.Е. Введение в идентификацию объектов управления. – М.: Энергия, 1977.-216 с.
17. Грайворонський М. В., Новіков О. М. Г14 Безпека інформаційно-комунікаційних систем. – К.: Видавнича група ВНУ, 2009. – 608 с
18. Дерех З.Д., Рейцен Є.О. Дослідження підвищення ефективності автоматизованих систем // Наук.-техн. вісник "Безпека комп'ютерних систем".-1999. – 1(2). – с. 79–85.
19. Герасименко В.А. Защита информации в автоматизированных системах обработки данных. - В 2-х кн. - М.: Энергоатомиздат, 1994.
20. Автоматизована система. Матеріали з Вікіпедії. [Електронний ресурс]. – Режим доступу: <http://ru.wikipedia.org/wiki/...>
21. Постановка задачи параметрической идентификации. [Електронний ресурс]. —Режим доступа: <http://www.ccas.ru/konkom/ident.htm>
22. Автоматизированные информационные технологии в экономике: Учебник/Под. ред. Г.А.Титаренко.-М.: Компьютер. ЮНИТИ, 1998.
23. *Information technology – Automatic identification and data capture techniques – QR Code (ISO/IEC 18004:2000):* ДСТУ – ISO/IEC 18004:2000 [Действителен с 2000г.]. – ISO copyright office, 2000. – 140с. – (Міжнародний стандарт).
24. *Lamb P. ARToolkit.* [Електронний ресурс]. – Режим доступа: <http://www.hitl.washington.edu/artoolkit/>

25. [Комп'](#)ютерні технології обробки інформації. / Під. ред. С.В. Назарова. - М.: [Фінанси і статистика](#), 1995.
26. *ISO 9126 Software Quality Characteristics* [Електронний ресурс] .– Режим доступу *URL: <http://www.sqa.net/iso9126.html>*
27. *ISO/IEC FDIS 9126-1* [Електронний ресурс] .– Режим доступу *URL: <http://www.cse.unsw.edu.au/~cs3710/PMmaterials/Resources/9126-1%20Standard.pdf>*

ДОДАТОК 1

Лістинг програмного коду

```
using System;

using System.Collections.Generic;

using System.Diagnostics;

using System.IO;

using System.Linq;

using System.Windows.Forms;

//using Word = Microsoft.Office.Interop.Word;

namespace Search
{
    public partial class Form1 : Form
    {
    {
        InitializeComponent();
    }

    private void button1_Click(object sender, EventArgs e)
    {
        //Додання шляху директорій до списку
        DialogResult result = folderBrowserDialog1.ShowDialog();
        if (result == DialogResult.OK)
        {
            listBox4.Items.Add(folderBrowserDialog1.SelectedPath);
        }
    }
}
}
```



```

private void button2_Click(object sender, EventArgs e)
{
    string writes;

    if (textBox2.Lines.Any() && listBox4.Items.Count != 0) //Перевірка
заповненості словника та списку директорій
    {
        string date = DateTime.Now.ToString()+".txt";

        //Створення назви файлу звіту з датою і часом проведення пошуку

        //Заміна некоректних символів

        date = date.Replace("/", "-");

        date = date.Replace(":", "-");

        string fileName = Environment.CurrentDirectory+"\" + date;
//Створення файлу звіту

        //Внесення в звіт допоміжної інформації

        writes = "Сканування проводилося " + DateTime.Now.ToString() + " "
+ Environment.NewLine+Environment.NewLine;

        System.IO.File.AppendAllText(fileName, writes);

        System.IO.File.AppendAllText(fileName, " Було вибрано наступні
параметри сканування:");

        //Створення масиву з слів згідно словника

        string[] dic = textBox2.Lines;

        //Запис до звіту слів з словника та допоміжної інформації

        File.AppendAllLines(fileName, dic);

        writes = Environment.NewLine + Environment.NewLine + "Було
знайдені такі файли:" + Environment.NewLine;

        System.IO.File.AppendAllText(fileName, writes);

        string[] doc = new string[0];

        //Пошук всіх файлів в вказаних директоріях

```

```

        foreach (string its in listBox4.Items)
        {
            doc = doc.Concat(Directory.GetFiles(its, "*.docx",
SearchOption.AllDirectories)).ToArray();

            doc = doc.Concat(Directory.GetFiles(its, "*.txt",
SearchOption.AllDirectories)).ToArray();

            doc = doc.Concat(Directory.GetFiles(its, "*.pptx",
SearchOption.AllDirectories)).ToArray();
        }

        var files = new List<string>();
        for (int i = 0; i < doc.Length; i++)
        {
            char slas = '\x005C';

            doc[i].ToLower();

            //Присвоєння всьому тексту нижнього регістру
            int slash = doc[i].LastIndexOf(slas);

            //Визначення індексу розміщення \\
            for (int j = 0; j < dic.Length; j++) //Перебір слів в словнику
            {
                if (slash < doc[i].LastIndexOf(dic[j]))
                //Умова пошуку тільки в назві файлу
                {
                    listBox1.Items.Add(doc[i]);

                    //Запис до списку результатів

                    //Запис результату до звіту з перенесенням на новий рядок
                    System.IO.File.AppendAllText(fileName, doc[i]);

                    System.IO.File.AppendAllText(fileName, Environment.NewLine);
                }
            }
        }
    }
}

```

```

        }
    }

    writes = Environment.NewLine + Environment.NewLine + "Було знайдено слова в
таких файлах:" + Environment.NewLine; // Формування допоміжної інформації для
звіту

        string texts = new
TikaOnDotNet.TextExtraction.TextExtractor().Extract(doc[i]).Text; //Зчитування тексту
що міститься в файлі

    texts.ToLower();

        //Присвоєння всьому тексту нижнього регістру

        for (int j = 0; j < dic.Length; j++)

        //Перебір слів в словнику

        {

            if (texts.IndexOf(dic[j]) != -1)

            //Пошук наявності в тексті слова

            {

                listBox2.Items.Add(doc[i]); //Запис до списку результатів

                files.Add(doc[i]); //Запис результату до списку який буде записаний в звіт

                j = dic.Length;

            }

        }

    }

        //Запис даних до звіту

        writes = Environment.NewLine + Environment.NewLine + "Було
знайдено слова в таких файлах:" + Environment.NewLine;

        File.AppendAllText(fileName, writes);

        File.AppendAllLines(fileName, files);

    }

```

```

else MessageBox.Show("Введіть дані до словника, або списку директорій"); //
Виведення помилки про те що не заповнені дані в словнику або списку директорій
    }

private void listBox3_SelectedIndexChanged(object sender, EventArgs e)
    {
    }

private void button5_Click(object sender, EventArgs e)
    {
        //Видалення об'єкту з списку
        listBox4.Items.Remove(listBox4.SelectedItem);
    }

private void listBox1_DoubleClick(object sender, EventArgs e)
    {
        //Відкриття файлу по двійному кліку
        Process.Start(listBox1.Items[listBox1.SelectedIndex].ToString());
    }

private void listBox2_DoubleClick(object sender, EventArgs e)
    {
        //Відкриття файлу по двійному кліку
        Process.Start(listBox2.Items[listBox2.SelectedIndex].ToString());
    }

private void listBox2_SelectedIndexChanged(object sender, EventArgs e)
    {
        //Зняття виділення з іншого поля вибору
        if (listBox1.SelectedIndex != -1 && listBox2.SelectedIndex != -1)
listBox1.SetSelected(listBox1.SelectedIndex, false);
    }

```

```

    }

    private void button6_Click(object sender, EventArgs e)
    {
        //Очищення полів результатів
        listBox1.Items.Clear();
        listBox2.Items.Clear();
    }

    private void button7_Click(object sender, EventArgs e)
    {
        //Перевірка того чи було вибрано файл на видалення
        if (listBox1.SelectedIndex != -1)
        {
            //Видалення файлу з комп'ютера та з списку
            File.Delete(listBox1.Items[listBox1.SelectedIndex].ToString());
            listBox1.Items.Remove(listBox1.SelectedItem);
        }
        else if (listBox2.SelectedIndex != -1)
        {
            File.Delete(listBox2.Items[listBox2.SelectedIndex].ToString());
            listBox2.Items.Remove(listBox2.SelectedItem);
        }
        else MessageBox.Show("Виберіть файл для видалення"); // Сповіщення
        про те що не було вибрано файл
    }

    private void listBox1_SelectedIndexChanged(object sender, EventArgs e)
    {

```

//Зняття виділення з іншого поля вибору

```
if (listBox2.SelectedIndex != -1 && listBox1.SelectedIndex != -1)
listBox2.SetSelected(listBox2.SelectedIndex, false);
}
```