

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Інститут електроенергетики

(інститут)

Факультет інформаційних технологій

(факультет)

Кафедра Програмного забезпечення комп'ютерних систем

(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня
магістра

(назва освітньо-кваліфікаційного рівня)

студента *Ігнашева Василя Олександровича*

(ПІБ)

академічної групи *121М-19-1*

(шифр)

спеціальності *121 Інженерія програмного забезпечення*

(код і назва спеціальності)

на тему: *Розробка програмного забезпечення*

для дослідження ефективності обробки запитів

пошукових систем та їх функції у корпоративній розвідці

В.О. Ігнашев

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтин говою	інституці йною	
розділів кваліфікаційної роботи				
спеціальний	проф. Алексєєв М.О.			
економічний	Доц. Касьяненко Л.В.			

Рецензент

Нормоконтролер

Доц. Сироткіна О.І.

Дніпро
2020

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

ЗАТВЕРДЖЕНО:

Завідувач кафедри

Програмного забезпечення комп'ютерних систем
 (повна назва)

І.М. Удовик

(підпис)

(прізвище, ініціали)

« »

20 20 Року

ЗАВДАННЯ

на виконання кваліфікаційної роботи магістра

спеціальності 121 Інженерія програмного забезпечення
 (код і назва спеціальності)

студенту 121М-19-1 Ігнашеву Василю Олександровичу
 (група) (прізвище та ініціали)

Тема кваліфікаційної роботи Розробка програмного забезпечення
для дослідження ефективності

пошукових систем та їх функції у корпоративній розвідці

1 ПІДСТАВИ ДЛЯ ПРОВЕДЕННЯ РОБОТИ

Наказ ректора НТУ «Дніпровська політехніка» від 22.10.2020 р. № 888-с

2 МЕТА ТА ВИХІДНІ ДАНІ ДЛЯ ПРОВЕДЕННЯ РОБІТ

Об'єкт досліджень – процес обробки запитів у пошукових системах Google, Yandex, Bing, TinEye.

Предмет досліджень – класифікація пошукових запитів за точністю відповіді системи.

Мета роботи – аналіз точності та швидкості результатів запитів у пошукових системах в залежності від типу вхідних даних, геолокації та форми запиту, аналіз пошукових систем як метода корпоративної розвідки та захисту даних.

3 ОЧІКУВАНІ НАУКОВІ РЕЗУЛЬТАТИ

Наукова новизна отриманих результатів кваліфікаційної роботи визначається тим, що удосконалено метод визначення точності та швидкості оброблення запитів пошуковими системами.

Практична цінність результатів полягає в тому, що запропонована робота дозволяє забезпечити максимальну продуктивність при пошуку інформації у відкритих джерелах.

- проведено аналіз відповідей пошукових систем;
- за отриманими результатами складено набір даних для продуктивного та точного використання пошукових систем.

4 ЕТАПИ ВИКОНАННЯ РОБІТ

Найменування етапів робіт	Строки виконання робіт (початок – кінець)
Збір інформації для дослідження предметної області	16.10.2020 – 02.11.2020
Дослідження методів для вирішення поставленого завдання	06.11.2020 – 01.12.2020
Експериментальні дослідження	01.12.2020 – 07.12.2020
Економічна частина	07.12.2020 – 10.12.2020

Завдання видав

(підпис)

Алексєєв М.О.

(прізвище, ініціали)

Завдання прийняв до виконання

(підпис)

Ігнашев В.О.

(прізвище, ініціали)

Дата видачі завдання: 12.09.2020

Термін подання кваліфікаційної роботи до ЕК 15.12.2020

РЕФЕРАТ

Пояснювальна записка: 41 стор., 26 рис., 2 додатки, 45 джерел.

Об'єкт досліджень: процес обробки запитів у пошукових системах Google, Yandex, Bing, TinEye.

Предмет досліджень: класифікація пошукових запитів за точністю відповіді системи.

Мета магістерської роботи: аналіз точності та швидкості результатів запитів у пошукових системах в залежності від типу вхідних даних для підвищення ефективності користувацького досвіду, геолокації та форми запиту, аналіз пошукових систем як метода корпоративної розвідки та захисту даних.

Методи дослідження: при вирішенні поставлених завдань виконано аналіз і наукове узагальнення літературних джерел по вихідним посилам досліджень.

Наукова новизна: удосконалено метод визначення ефективності оброблення запитів пошуковими системами та їх застосування у корпоративній розвідці.

Практична цінність результатів полягає в тому, що запропонована в роботі методика дозволяє забезпечити максимальну продуктивність при пошуку інформації у відкритих джерелах.

У розділі «Економіка» проведені розрахунки трудомісткості розробки програмного забезпечення, витрат на створення ПЗ і тривалості його розробки, а також проведені маркетингові дослідження ринку збуту створеного програмного продукту.

Список ключових слів: пошукова система, корпоративна розвідка, зображення, оператори розширеного пошуку.

ABSTRACT

Explanatory note: 41 pages, 26 figures, 2 appendices, 45 sources.

Object of research: search engines Google, Yandex, Bing, TinEye.

Subject of research: Development of software for studying the efficiency of search engine query processing and their functions in corporate intelligence.

The purpose of the master's work: to determine the effectiveness of search engines depending on the type of input data, geolocation and request form, to determine the role of search engines in corporate intelligence and data protection.

Research methods: in solving the tasks performed analysis and scientific generalization of literature sources on the original research.

Scientific novelty: a method of determining the effectiveness of search engine query processing and their application in corporate intelligence was upgraded.

The practical value of the results is that the proposed method allows to ensure maximum productivity when searching for information in open sources.

In the section "Economics" calculations of the complexity of software development, the cost of creating software and the duration of its development, as well as marketing research of the market for the created software product.

Keyword list: search engine, corporate intelligence, images, advanced search operators.

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1. ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАВДАННЯ.....	10
1.1 Поняття пошукової системи	10
1.1.1 Класифікація пошукових систем.....	11
1.2 Корпоративна розвідка.....	12
1.2.1 Основні методи корпоративної розвідки.....	12
1.2.2 Принцип роботи.....	13
1.2.3 Переваги та застосування.....	14
1.3. Висновки до першого розділу.....	15
РОЗДІЛ 2. АНАЛІЗ ЕФЕКТИВНОСТІ ПОШУКОВИХ СИСТЕМ.....	15
2.1. Поняття пошукової системи.....	17
2.2. Класифікація пошукових систем.....	18
2.3. Поняття ефективності.....	18
2.4. Оператори розширеного пошуку.....	18
2.5. Комбінування операторів розширеного пошуку.....	19
2.6. Практичне застосування операторів розширеного пошуку у задачах корпоративної розвідки.....	20
2.7. Пошук зображень.....	21
2.8. Google зображення.....	22
2.9. Яндекс зображення.....	23
2.10. Bing зображення.....	24
2.11. TinEye.....	25
2.12. Висновки до другого розділу.....	25
РОЗДІЛ 3. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ.....	26
3.1 Проведення досліджень.....	26
3.2 Висновки до третього розділу.....	41
РОЗДІЛ 4. ЕКОНОМІКА.....	42

4.1	Маркетингові дослідження.....	42
4.2	Економічна ефективність.....	43
4.3	Висновки.....	43
	ВИСНОВКИ.....	44
	ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	45
	Додаток А. ВІДГУК КЕРІВНИКА ЕКОНОМІЧНОГО РОЗДІЛУ.....	48
	Додаток Б. ПЕРЕЛІК ДОКУМЕНТІВ НА ОПТИЧНОМУ НОСІЇ.....	49

ВСТУП

Актуальність роботи. В рамках роботи було розглянуто пошукові системи, так як вони мають ряд значних переваг:

- простота у використанні;
- широкий функціонал - у якості параметра пошуку можна використовувати текст, число, зображення та геодані;
- велика кількість індексованих пошуковими системами сторінок;
- автоматичний пошук;
- точний результат пошуку.
- зберігання попередніх версій сайту у кеші пошукових систем.

В останні десятиліття в світі бурхливо розвивається напрямок SEO - оптимізації та просування сайтів, що створює тісний зв'язок матеріалів та сторінок інтернет порталів із пошуковими системами, що, у свою чергу, робить пошукові системи незамінним інструментом корпоративної розвідки.

Актуальність досліджень в цьому напрямку підтверджується кількістю різних застосувань пошукових систем. Це є інструментом пошуку нової пісні лише за мотивом та ритмом, пошуку давніх друзів та знайомих, актуальних новин, нового будинку або нової машини, ресурсів для навчання, розважального контенту та безліч інших напрямків.

Широке коло завдань, що вирішується пошуковими системами, не дозволяє в даний час створювати універсальні та водночас точні запити, змушуючи розробляти текст запиту в залежності від цілі пошуку та очікуваних результатів.

Мета магістерської роботи: аналіз точності та швидкості результатів запитів у пошукових системах в залежності від типу вхідних даних для підвищення ефективності користувацького досвіду, геолокації та форми запиту, аналіз пошукових систем як метода корпоративної розвідки та захисту даних.

Об'єкт дослідження: процес обробки запитів у пошукових системах Google, Yandex, Bing, TinEye.

Предмет дослідження: класифікація пошукових запитів за точністю відповіді системи.

Методи дослідження: при вирішенні поставлених завдань виконано аналіз і наукове узагальнення літературних джерел по вихідним посилам досліджень.

Наукова новизна: удосконалено метод визначення ефективності оброблення запитів пошуковими системами та їх застосування у корпоративній розвідці.

Практичне значення: запропонований в роботі метод дозволяє забезпечити максимальну продуктивність при пошуку інформації у відкритих джерелах.

Особистий внесок автора полягає в розробці теоретичної частини магістерської роботи, в дослідженні і систематизації знань про існуючі методики, розробці методів досліджень і технологій реалізації, в оцінці отриманих результатів.

Структура та обсяг роботи. Робота складається з вступу, трьох розділів і висновків. Містить 49 сторінок друкованого тексту, в тому числі 43 сторінок тексту основної частини з 26 рисунками, списку використаних джерел з 50 найменуваннями на 4 сторінках, 2 додатків на 2 сторінках.

РОЗДІЛ 1

ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАВДАННЯ

1.1. Поняття пошукової системи

Пошукова система - це онлайн-служба (апаратно-програмний комплекс з веб інтерфейсом), що надає можливість пошуку інформації в Інтернеті. У просторіччі під пошуковою системою розуміють веб сайт, на котрому розміщено інтерфейс системи. Програмною частиною пошукової системи є пошукова машина (пошуковий рушій) — комплекс програм, що забезпечує функціональність пошукової системи.

Більшість пошукових систем шукають інформацію на сайтах Інтернету, але існують також системи, здатні шукати файли на ftp-серверах, товари в інтернет-магазинах, а також інформацію в групах новин Usenet.

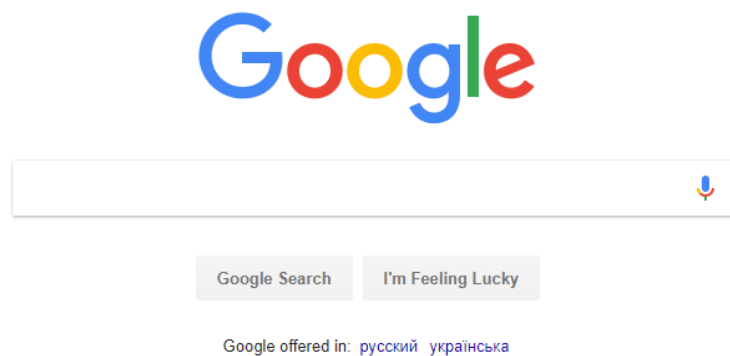


Рис.1.1. Приклад інтерфейсу пошукової системи

1.1.1 Класифікація пошукових систем

Класифікація пошукових систем відбувається за метою пошуку, типу даних, за якими здійснюється пошук, масиві даних, у якому здійснюється пошук та геолокації. За такими параметрами пошукові системи можна поділити на такі категорії:

- загальні;

- локальні;
- спеціалізовані.

Загальні пошукові системи

До загальних пошукових систем відносяться ті системи, що виконують пошук за індексованими у мережі Інтернет без прив'язки до певних локальних кордонів (таких як країна, частина світу, континент) та мови запиту, можуть індексувати різні типи файлів - від тексту до файлів з певним розширенням (.doc, .pdf, .rar та ін.). Загальними пошуковими системи є найпопулярніші у світі системи, такі як Google, Bing, Duckduckgo, Yahoo.

Локальні пошукові системи

До локальних пошукових систем відносяться системи, що орієнтовані на запити на певній мові або у певній геолокації. Вони також можуть індексувати різні типи файлів - від тексту до файлів з певним розширенням (.doc, .pdf, .rar та ін.), але показують гірший від загальних пошукових систем результат при пошуку на іншій мові або пошуку певної інформації за межами заданої геолокації. До локальних пошукових систем можна віднести німецьку систему Wseeker, українську UKR.NET, французьку Voilà та багато інших.

Спеціалізовані пошукові системи

До спеціалізованих пошукових систем можна віднести системи, що були створені для певного типу задач та найкращий результат показують при пошуку лише певних запитів, таких як пошук медіафайлів, пошук наукової інформації, пошук зображень та пошук даних у неіндексованих сторінках мережі Інтернет. Представниками цієї групи пошукових систем є система для пошуку файлів GlobalFileSearch, система для пошуку зображень TinEye, системи для пошуку наукової інформації ScienceDirect та Google Scholar, а також системи для пошуку серед неіндексованих сторінок у браузері Tor - Torch та NotEvil.

1.2. Корпоративна розвідка

Корпоративна розвідка на основі відкритих даних, або OSINT - дисципліна розвідки, що складається з пошуку та аналізу інформації, здобутої у відкритих джерелах, таких як: публікації, статті, соціальні мережі, інтерв'ю, резюме, публічні реєстри, звіти, декларації та інші джерела, що є загальнодоступними та відкритими.

Історія виникнення цього напрямку розвідки починається с 1941 року, коли у Сполучених Штатах Америки було сформовано Службу моніторингу зарубіжних трансляцій, співробітники котрої записували передачі на коротких хвилях, перекладали та відправляли до військових відомств у вигляді щотижневих докладів.

1.2.1. Основні методи корпоративної розвідки

Методами корпоративної розвідки є:

- формулювання запитів у пошукових системах;
- аналіз інформації у соціальних мережах;
- моніторинг інтернет порталів;
- зіставлення фактів;
- пошук інформації у відкритих реєстрах.

Формулювання коректних та доцільних запитів у пошукових системах є основною складовою у корпоративній розвідці. Завдяки великій популярності просування сайтів та Інтернет-порталів у пошукових системах, правильно складений запит може вирішити поставлене завдання ще на етапі формулювання.

Завдяки моніторингу і аналізу соціальних мереж, можна дізнатися про особисті переваги, смаки та хобі людини, дізнатися її коло спілкування, близьких людей та родичів, роботу та геопозицію. Якщо ж мова йде про юридичну особу,

наприклад, компанію, проаналізувавши акаунти можна дізнатись штат, поточні проекти, зміни у топ-менеджменті та навіть фінансові показники.

Моніторинг інтернет-ресурсів спрямований на спостереження та аналіз змін на сторінках порталу, що може визначити зміну власника порталу, отримати контактні дані або компрометуючий матеріал. Особливо корисним є аналіз сторінки у минулому. Для цього використовується або кеш пошукових систем, або портал InternetArchive.

Зіставлення фактів - основна частина роботи у корпоративній розвідці. Завдяки цьому формуються певні висновки щодо поставленої задачі.

Реєстри є невід'ємною складовою пошуку інформації, адже вони містять точну та достовірну інформацію. У більшості випадків використовується інформація з судових реєстрів, аби дізнатися про судові справи, адміністративні та кримінальні правопорушення об'єкта та реєстри з інформацією по юридичним особам, котрі можуть містити важливу інформацію про власників, бенефіціарів, засновників та пов'язаних з об'єктом осіб.

1.2.2. Принцип роботи

Корпоративна розвідка є сферою, у котрій неможливо скласти однаково успішний алгоритм дій для досягнення результатів у будь-яких завданнях, на це впливає значна варіативність завдань та доступної у конкретному випадку інформації, але можна виділити загальний план, на який можна опиратися при роботі, а саме: формулювання технічного завдання, пошук необхідної інформації, аналіз отриманих даних, проекція результату на технічне завдання, висновок. Важливою особливістю корпоративною розвідки є увага до деталей, бо саме зіставлення непомітних, на перший погляд, деталей є ключовим фактором до результативної роботи.

Також, завдяки гнучкості сфери, існує безліч методик пошуку та аналізу інформації. У якості базових методик, що є універсальними, є прогресивний джипег, або ітеративний метод, та метод дерев. У основі ітеративного методу є

збір максимальної кількості інформації за короткий проміжок часу, після чого увесь знайдений матеріал аналізується та сортується, переглядається, та вибірково дошукується у логічних та смислових розбіжностях. Метод дерев передбачає пошук безперервний пошук інформації по заціпкам - від однієї до іншої, таким чином, пошук інформації нагадує дерево з гілками, на кожній гілці якого ще по декілька гілок.

1.2.3. Переваги та застосування

Конкурентна розвідка має ряд значних переваг: перевірка контрагентів, перевірка потенційних партнерів та інвесторів, перевірка бекграунду осіб, відсутність потреби у спеціальному обладнанні для роботи, легальність методів пошуку інформації.

Всі перераховані переваги, роблять конкурентну розвідку незамінним помічником при пошуку партнерів у бізнесі, при пошуку найманих співробітників та аналізу пропозицій конкурентів

1.3. Висновки до першого розділу

У цьому розділі ми прослідкували основні аспекти та складові корпоративної розвідки та пошукових систем. В цілому, роботу корпоративної розвідки можна описати наступним чином: поставити задачу, яку необхідно розв'язати, зібрати інформацію з відкритих джерел, таких як публікації, статті, соціальні мережі, інтерв'ю, резюме, публічні реєстри, звіти, декларації та інші джерела, що є загальнодоступними та відкритими, проаналізувати отриману інформацію, зіставити факти та докази. Якщо у відповідь на власне запитання “Чи вирішене завдання?” ми відповідаємо “так”, задача виконана, якщо ні - повторюємо алгоритм. Кожен з етапів надзвичайно важливий, адже неякісно виконаний етап може призвести до неправдивих результатів, що може мати дуже неприємні результати.

РОЗДІЛ 2

АНАЛІЗ ЕФЕКТИВНОСТІ ПОШУКОВИХ СИСТЕМ

2.1. Поняття пошукової системи

Пошукова система - це онлайн-служба (апаратно-програмний комплекс з веб інтерфейсом), що надає можливість пошуку інформації в Інтернеті. У просторіччі під пошуковою системою розуміють веб сайт, на котрому розміщено інтерфейс системи. Програмною частиною пошукової системи є пошукова машина (пошуковий рушій) — комплекс програм, що забезпечує функціональність пошукової системи.

Більшість пошукових систем шукають інформацію на сайтах Інтернету, але існують також системи, здатні шукати файли на ftp-серверах, товари в інтернет-магазинах, а також інформацію в групах новин Usenet.

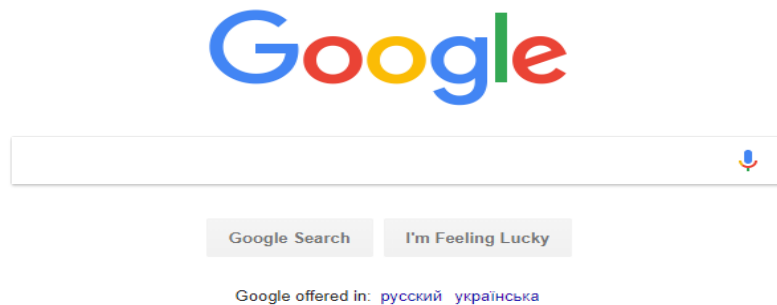


Рис.2.1. Приклад інтерфейсу пошукової системи

2.2. Класифікація пошукових систем

Класифікація пошукових систем відбувається за метою пошуку, типу даних, за якими здійснюється пошук, масиві даних, у якому здійснюється пошук

та геолокації. За такими параметрами пошукові системи можна поділити на такі категорії:

- загальні;
- локальні;
- спеціалізовані.

Загальні пошукові системи

До загальних пошукових систем відносяться ті системи, що виконують пошук за індексованими у мережі Інтернет без прив'язки до певних локальних кордонів (таких як країна, частина світу, континент) та мови запиту, можуть індексувати різні типи файлів - від тексту до файлів з певним розширенням (.doc, .pdf, .rar та ін.). Загальними пошуковими системи є найпопулярніші у світі системи, такі як Google, Bing, Duckduckgo, Yahoo.

Локальні пошукові системи

До локальних пошукових систем відносяться системи, що орієнтовані на запиту на певній мові або у певній геолокації. Вони також можуть індексувати різні типи файлів - від тексту до файлів з певним розширенням (.doc, .pdf, .rar та ін.), але показують гірший від загальних пошукових систем результат при пошуку на іншій мові або пошуку певної інформації за межами заданої геолокації. До локальних пошукових систем можна віднести німецьку систему Wseeker, українську UKR.NET, французьку Voilà та багато інших.

Спеціалізовані пошукові системи

До спеціалізованих пошукових систем можна віднести системи, що були створені для певного типу задач та найкращий результат показують при пошуку лише певних запитів, таких як пошук медіафайлів, пошук наукової інформації, пошук зображень та пошук даних у неіндексованих сторінках мережі Інтернет. Представниками цієї групи пошукових систем є система для пошуку файлів GlobalFileSearch, система для пошуку зображень TinEye, системи для пошуку наукової інформації ScienceDirect та Google Scholar, а також системи для пошуку серед неіндексованих сторінок у браузері Tor - Torch та NotEvil.

2.3. Поняття ефективності

Під ефективністю використання пошукової системи мається на увазі точність та кількість результатів відповіді пошукової системи на певні запити. Після проведення досліджень буде складено таблицю з пошуковими системами та типами вхідних даних, у котрій будуть виставлені оцінки відповідно до результатів дослідження. Для дослідження було обрано такий тип даних як зображення.

2.4. Оператори пошуку

Для ефективної роботи з пошуковими системами використовуються оператори пошуку - символи та команди, що певним способом взаємодіють з пошуковою системою та повертатимуть лише певний результат. Операторів пошуку багато і всі мають свій функціонал, нижче наведені лише ті, що є найпопулярнішими:

- оператор “” - оператор примусового пошуку точного збігу, що шукатиме лише фразу або слово, яке знаходиться в лапках. Використовується для уточнення результатів видачі пошукової системи та виключення синонімів.
- оператор “-” - оператор винятку. Він виключає з результату пошуку певний термін або фразу;
- оператор “filetype:” - оператор обмеження типу файлів. Виключає з результату пошуку усі типи файлів, окрім заданого;
- оператор “cache:” - оператор кешу. Повертає кешовану копію заданої сторінки, за умови, якщо вона існує;
- оператор “site:” - оператор обмеження сайту. Виключає з результату пошуку всі сторінки, крім сторінок із заданого сайту;
- оператор “inurl:” - оператор обмеження url. Виключає з результату пошуку усі посилання, що не містять певної фрази у своєму URL;

2.5. Комбінування операторів розширеного пошуку

Увесь потенціал використання операторів розширеного пошуку у пошукових системах розкривається при їх комбінуванні. Саме влучне поєднання декількох операторів розширеного пошуку є інструментом досягнення бажаного результату та економії часу. Обмежень на кількість операторів у комбінації немає, можна комбінувати як 2 оператори розширеного пошуку, так і 5 або 10. Також не існує суворих правил щодо комбінування - будь-який оператор можна поєднати з цим же оператором, або з будь-яким іншим, в залежності від результату, якого планується досягти пошуковим запитом.

2.6. Практичне застосування пошукових операторів для задач корпоративної розвідки

Рішення поставленої задачі застосуванням операторів розширеного пошуку та їх комбінаціями є одним із найважливіших методів роботи на етапі пошуку інформації або її перевірки. Наприклад, метою завдання є визначення співробітників компанії ТОВ “N”. Для виконання цього завдання найкраще підійде соціальна мережа LinkedIn, але вона має деякі особливості, що можуть ускладнити виконання завдання. Ця мережа працює за моделлю кол: перше коло - ваші друзі, друге коло - друзі друзів. Люди, що знаходяться у цих колах, будуть відображатися у LinkedIn з повною інформацією про них, але ті люди, що поза цими колами, будуть показані як “учасник LinkedIn” та без можливості перейти до їх особистого профілю. Такий випадок є яскравим прикладом ефективності комбінування операторів пошуку, адже ці сторінки проіндексовані пошуковими системами, і виконавши запит “ТОВ “N” site:linkedin.com” ми отримаємо усі проіндексовані сторінки, що містять “ТОВ “N” та лише з сайту linkedin.com. Якщо необхідно визначити конкретного співробітника, наприклад, бухгалтера, до запиту необхідно додати назву посади - “ТОВ “N” “бухгалтер” site:linkedin.com.

2.7. Пошук зображень

У корпоративній розвідці багато роботи з різноманітними типами даних: звіти, код програми, виписки з реєстрів, і т.д. Одним з типів даних, на який рідко звертають увагу, є зображення, такі як зображення профіля у соціальних мережах та форумах, фотографії у небезпечному програмному забезпеченні, логотипи компаній та інші. Виконуючи пошук за зображенням, з'являється можливість отримати багато інформації про об'єкт досліджень. Для деяких випадків може бути достатньо пошуку зображень за допомогою звичайної пошукової системи, такої як Google або Яндекс. Наприклад, якщо у роботі трапився незнайомий логотип, виконавши простий пошук зображення можна отримати інформацію про походження логотипу, інформацію про бренд або геолокацію. Іноді, при спробі виконати пошук зображення може бути корисним змінити початкове зображення для досягнення найкращих результатів. Наприклад, іноді зображення може бути опубліковано та заявлено, що воно є оригіналом, коли в дійсності це відзеркалена або перевернута версія вже існуючого фото. Повертаючи або відзеркалюючи фотографію та виконуючи пошук зміненого зображення, є ймовірність знайти додаткові результати, які, можливо, не були отримані при пошуку початкової версії. Також, кропітливе обрізання фотографії може дати кращі результати, так як інші об'єкти на зображенні можуть заважати пошуковій системі сфокусуватися на основному об'єкті. Хоча зображення та фотографії не являються найпоширенішим джерелом інформації, цей метод може бути корисним в деяких випадках, коли необхідна ідентифікація людини або логотипу.

2.8. Google зображення

Google виконує все можливе для визначення що саме є об'єктом зображення, а не хто. Результати діляться на три частини: декілька результатів пошуку по об'єкту, що, за думкою алгоритму, зображено на фотографії,

візуально схожі (але не ідентичні) результати та сторінки, що містять дане зображення. До мінусів можна віднести те, що система працює добре лише з відомими людьми та логотипами, тобто якщо об'єкт пошуку не дуже популярний, велика ймовірність неповних або неякісних результатів. Також, пошук обмежується лише одним предметом, наприклад, задавши у параметр пошуку фотографію будинку поруч з озером та горами, у результаті будуть відображені не фотографії будинків поруч з горами та озером, а просто візуально схожі будинки, не враховуючи гори та озеро.

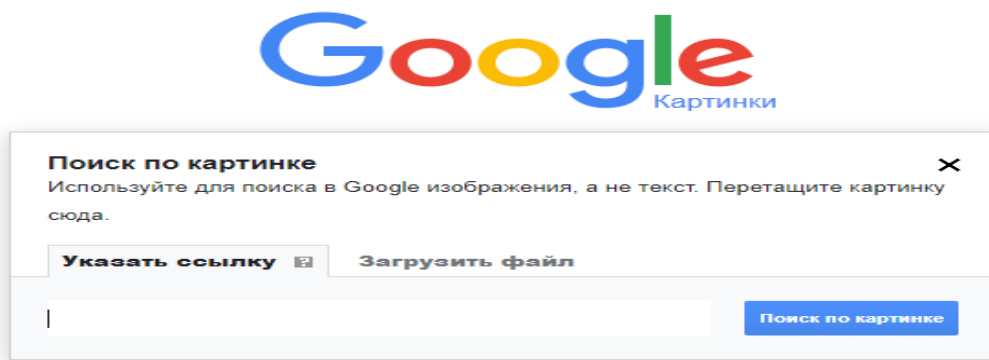


Рис.2.2. Інтерфейс Google зображень

2.9. Яндекс зображення

Яндекс, який є російськомовним аналогом Google, являється дуже ефективним ресурсом для пошуку зображень. Він надає додаткові розміри одного й того ж зображення, візуально схожі зображення, і багато результатів, коли схожі зображення наявні на сторінках. Яндекс має тенденцію бути найсильнішою пошуковою системою для зіставлення та ідентифікації облич та визначення місцезнаходження.

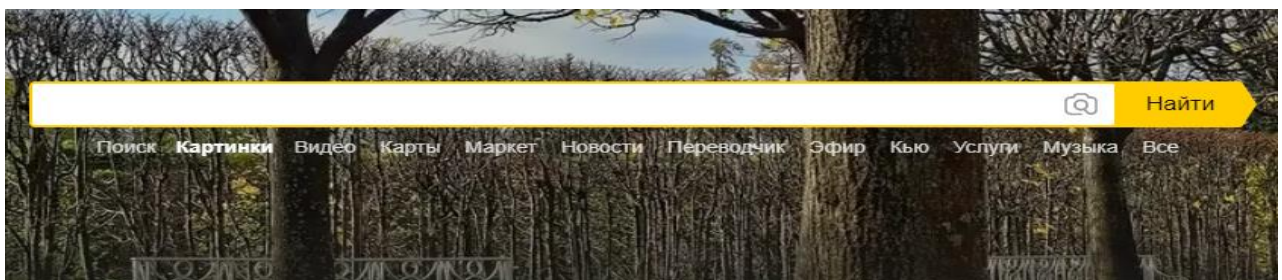


Рис.2.2. Інтерфейс Яндекс зображень

2.10. Bing зображення

Bing має унікальну функцію - можливість обрізати області зображення або фотографії та бачити результати у режимі реального часу. Це дуже ефективно для високоякісних зображень з великою кількістю елементів, що необхідно ідентифікувати. Крім того, порівнюючи Bing з Google, він намагається ідентифікувати елементи на фотографії та знайти зображення, що містять усі ці елементи. Наприклад, зображення старовинного автомобіля, що стоїть біля дерева, надасть у результатах зображення, що містять автомобіль та дерево, коли Google вибирає один елемент та шукає саме його. Також непогано виконує ідентифікацію людей за обличчям.

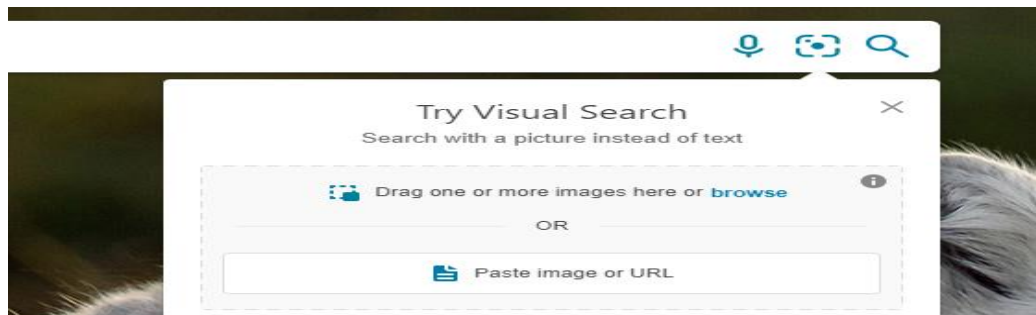


Рис.2.3. Інтерфейс Bing зображень

2.11. TinEye

Початкова та єдина мета TinEye полягала в тому, щоб знайти інші розміри одного й того ж зображення, і завдяки цій системі є можливість знайти задане зображення у інших зображеннях, що є корисним у деяких специфічних задачах, таких як пошук логотипів, ідентифікація користувачів форумів та інші.



Рис.2.4. Інтерфейс TinEye

2.12. Висновки до другого розділу

У цьому розділі ми ознайомились з такими поняттями як оператори розширеного пошуку та пошук по зображенню, розглянули можливість комбінування операторів розширеного пошуку та їх використання на прикладі пошуку співробітників певної компанії. Також ми розглянули інтерфейс пошукових систем та описали їх переваги та недоліки, що буде доведено у наступному розділі. Комбінування операторів розширеного пошуку є невід'ємною частиною під час виконання завдання у корпоративній розвідці, адже варіативність можливих комбінацій дійсно вражає, завдяки цьому можна підібрати запит практично до будь-якої задачі: від аналізу цін на кукурудзяне борошно у межі супермаркетів міста Дрогобич до пошуку людей, що вчилися в одному класі якоїсь спеціалізованої школи. Плюсом пошуку по зображенню є навпаки, дуже конкретна відповідь, що може або додати нових даних до аналізу, або підтвердити та ідентифікувати людину, логотип або місце. Це може бути дуже корисним, бо загалом люди не задумуються над вибором фотографій і зображень, і це може зіграти у нашу користь, коли мова йде, наприклад, про ідентифікацію особи під різними іменами або нікнеймами, або коли юридична особа незаконно використовує чужі фотоматеріали, мінімально або зовсім не редагувавши, та видаючи їх за свої.

РОЗДІЛ 3

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Для проведення дослідження ефективності пошуку зображень пошуковими системами було обрано елементи з різних категорій, а саме:

- логотип компанії DomainTools



Рис. 3.1. Логотип DomainTools

- логотип Культу Мертвої Корови

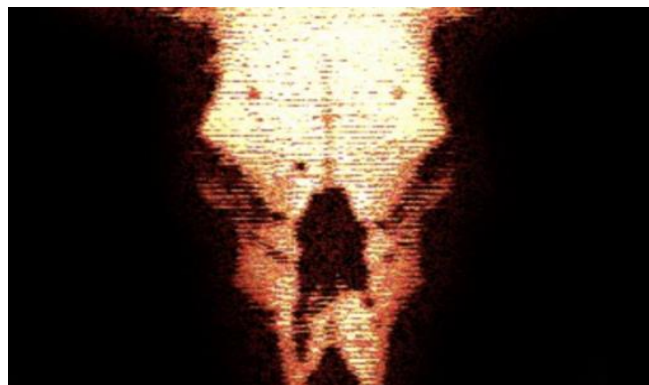


Рис. 3.2. Логотип Культу Мертвої Корови

- знімок екрану з зображенням Фішера Стівенса з фільму “Хакери”



Рис. 3.3. Знімок екрану з зображенням Фішера Стівенса

- фотографія Центрального залізничного вокзала Мілану



Рис. 3.4. Фотографія Центрального залізничного вокзала Мілану

Таким чином, використовуючи ці зображення, було досягнуто повного комплексу критеріїв для оцінювання:

- деталі;
- локація;
- обличчя;
- будівля;
- логотип;
- ресайз (зміна розміру).

У ході дослідження для всіх зображень, що були обрані для дослідження, було створено запит у пошукових системах Google, Yandex, Bing, TinEye.

Логотип DomainTools:

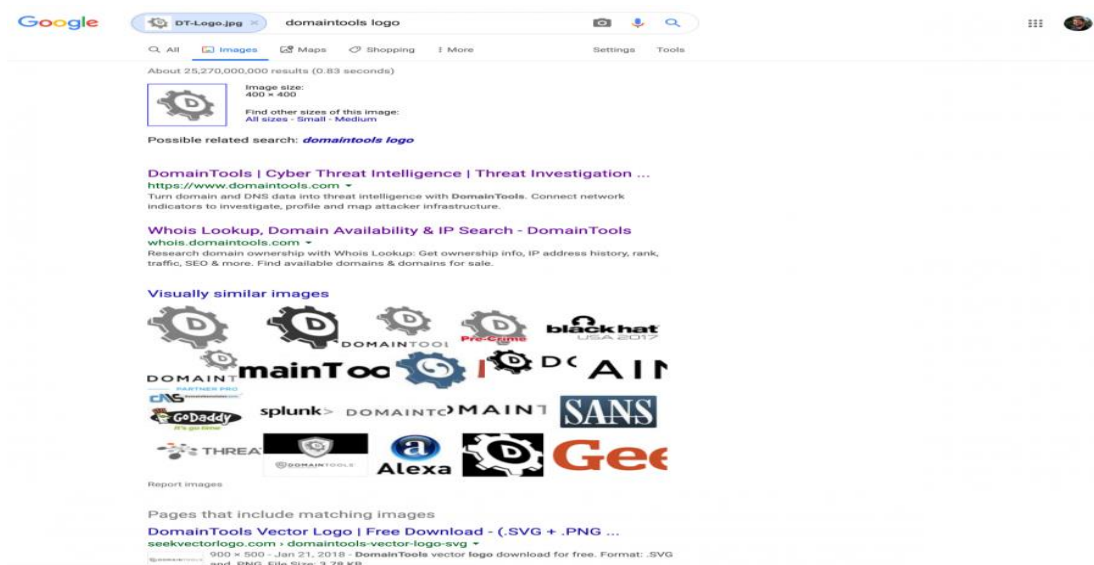


Рис. 3.5. Запит у пошуковій системі Google

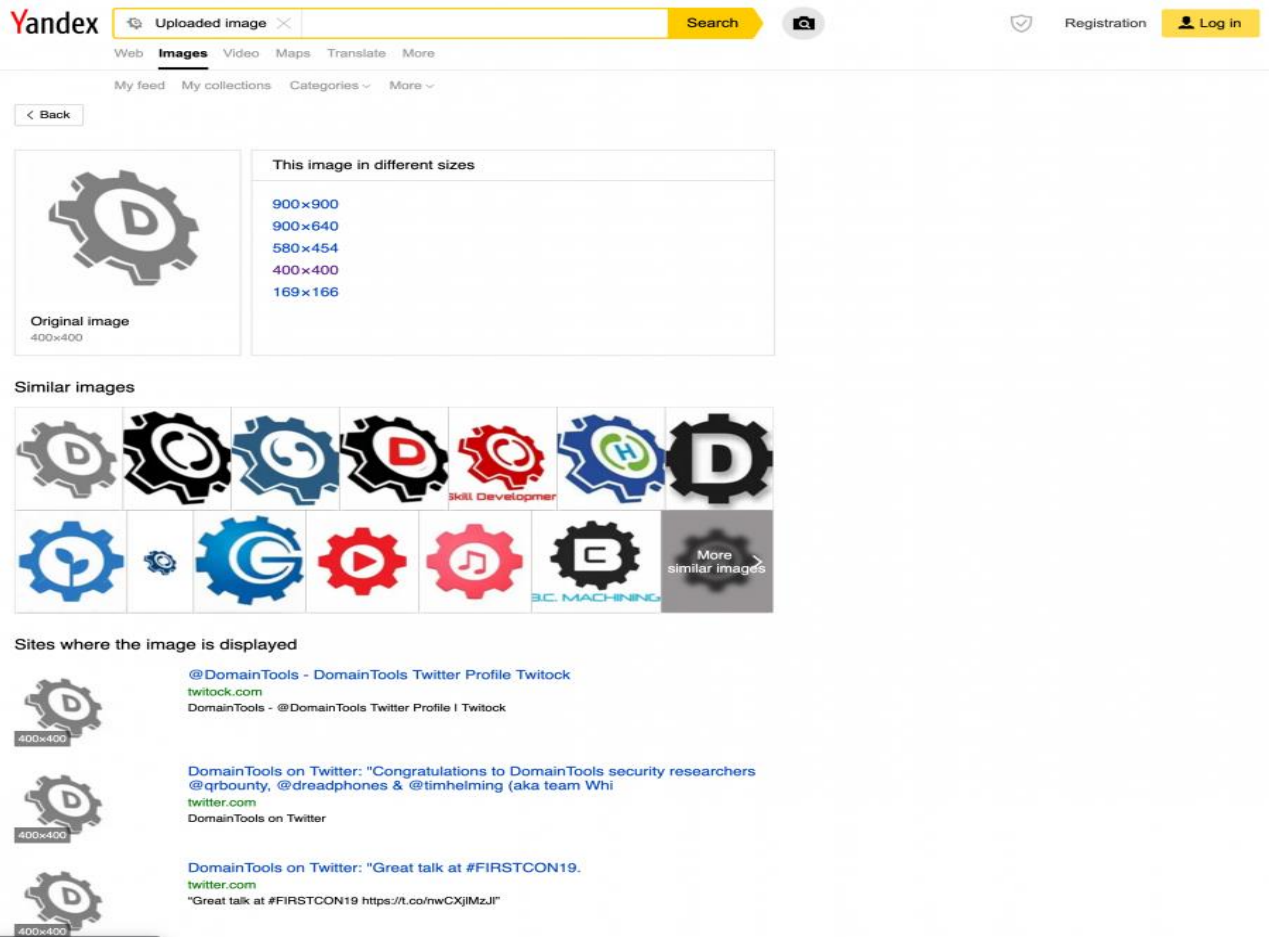


Рис. 3.6. Запит у пошуковій системі Яндекс

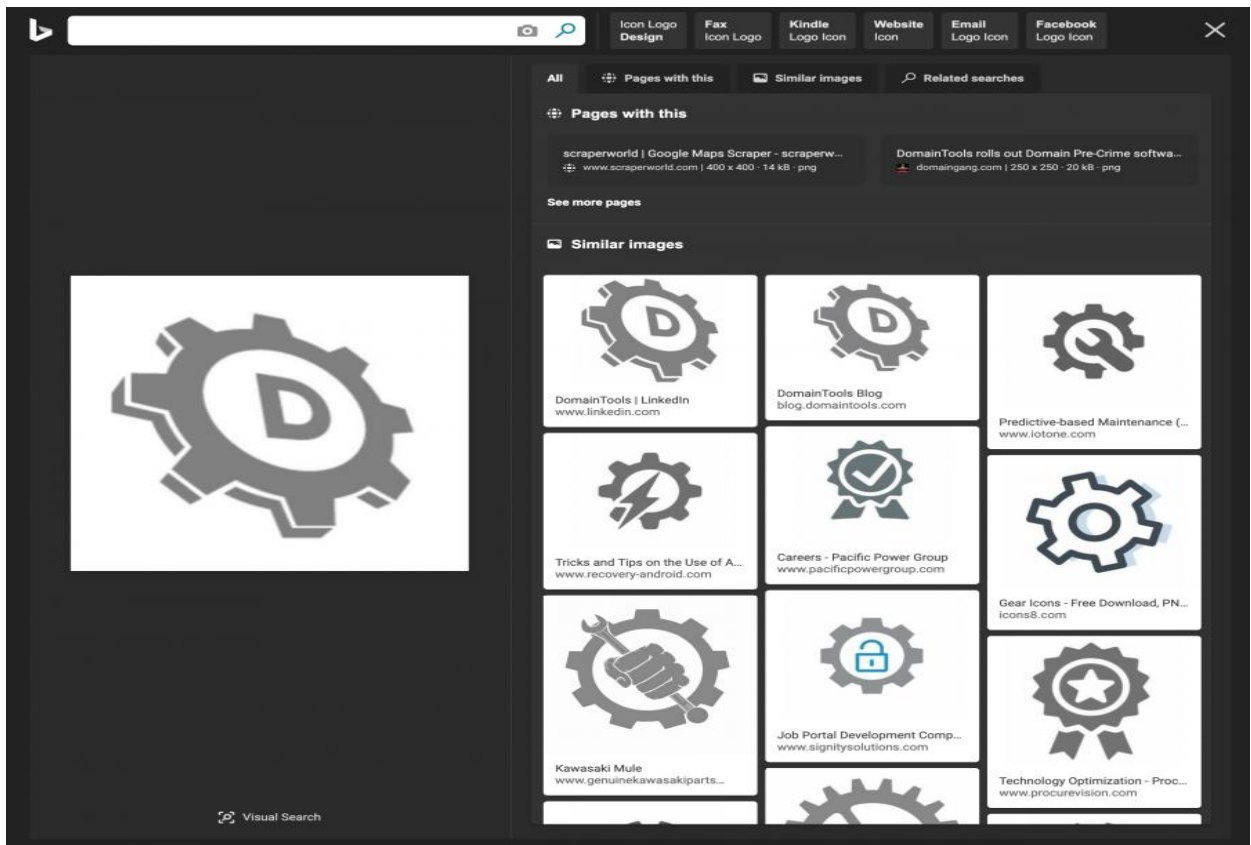


Рис. 3.7. Запит у пошуковій системі Bing

The screenshot shows the TinEye search interface. At the top, there is a search bar with the text "Upload or enter Image URL" and a search icon. Below the search bar, the results are displayed for the query "DT-Logo.jpg".

14 results
Searched over 36.2 billion images in 1.3 seconds for: DT-Logo.jpg

Using TinEye is private. We do not save your search images. TinEye is free to use for non-commercial purposes. For business solutions, [learn about our technology](#).

Show only stock and collection results:
 1 result found in collections.

Sort by best match | Filter by domain/collection | < 1 of 2 >

The results list includes:

- cross-index.pl**
profil.php - First found on Jul 05, 2019
Filename: [domaintools_logo.png](#) (300 x 250, 24.5 KB)
- twitter.com**
DomainTools - First found on Mar 30, 2016
Filename: [QdAlHW8j_400x400.png](#) (400 x 400, 77.5 KB)
- archive.is**
H9zfB - First found on Apr 10, 2017
Filename: [62c063954aa4213e1471e049b9095aba333154f9.png](#) (101 x 101, 2.6 KB)
3r0Is - First found on Apr 10, 2017
Filename: [62c063954aa4213e1471e049b9095aba333154f9.png](#) (101 x 101, 2.6 KB)
- COLLECTION**
jerryperes.deviantart.com
art/Domain-Transfer-Domain-Domain-Ser... - First found on Jan 25, 2017
Filename: (1024 x 330, 13.5 KB)

Рис. 3.8. Запит у TinEye

Google у цьому випадку показав найкращий результат, зв'язавши зображення з сайтом DomainTools. Яндекс виконав завдання гірше, в першу чергу спираючись на Twitter та показав перелік схожих логотипів. Bing не ідентифікував безпосередньо DomainTools, але у результаті показав зображення блогу DomainTools та сторінки в LinkedIn. TinEye показав, де містяться інші версії цього логотипу.

Зображення Фишера Стівенса з фільму “Хакери”:

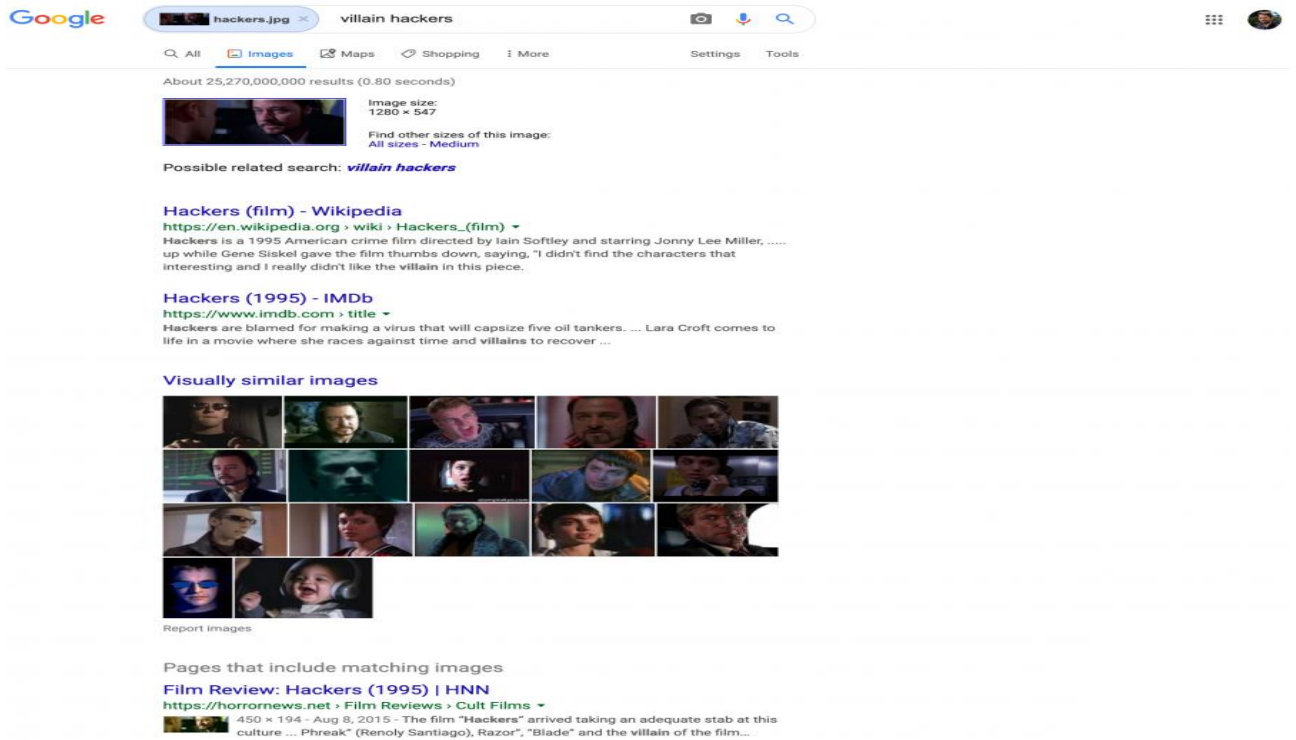


Рис. 3.9. Запит у Google

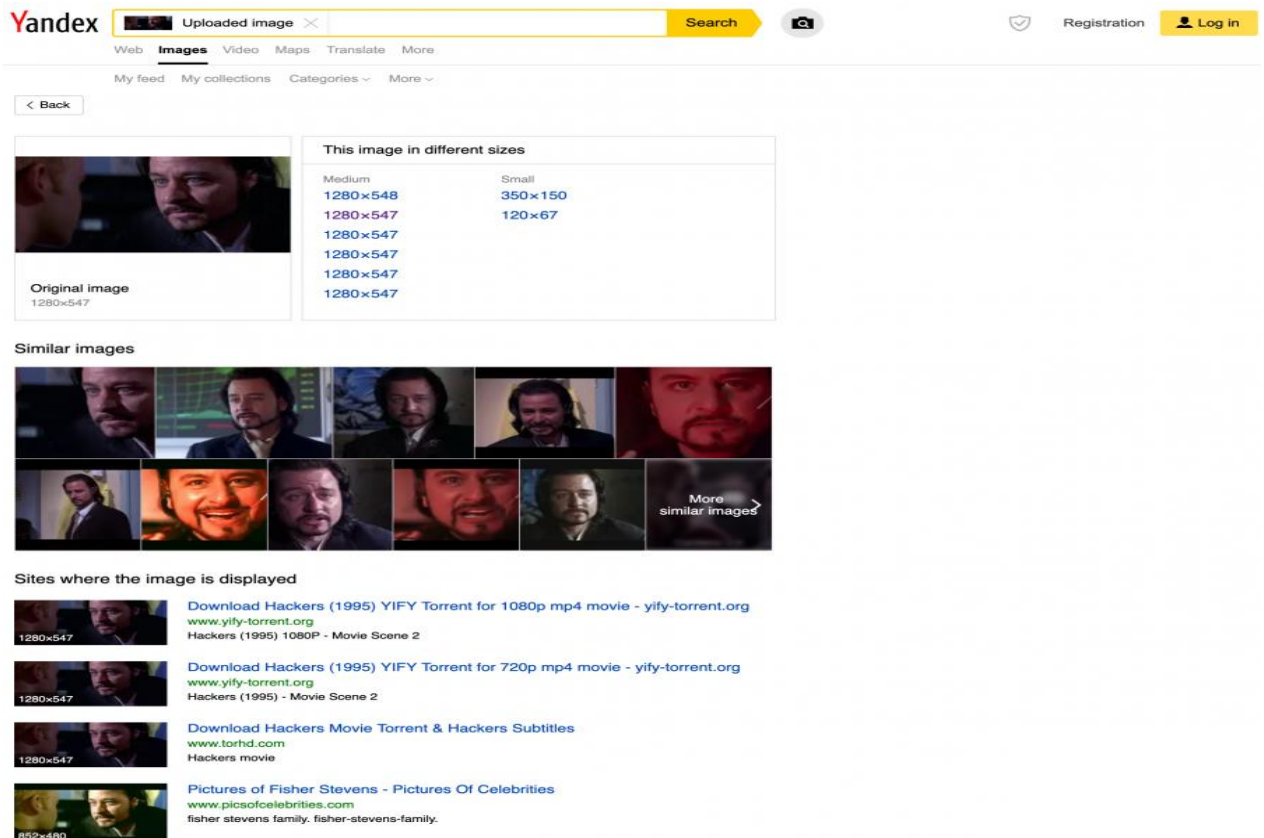


Рис. 3.10. Запит у Яндекс

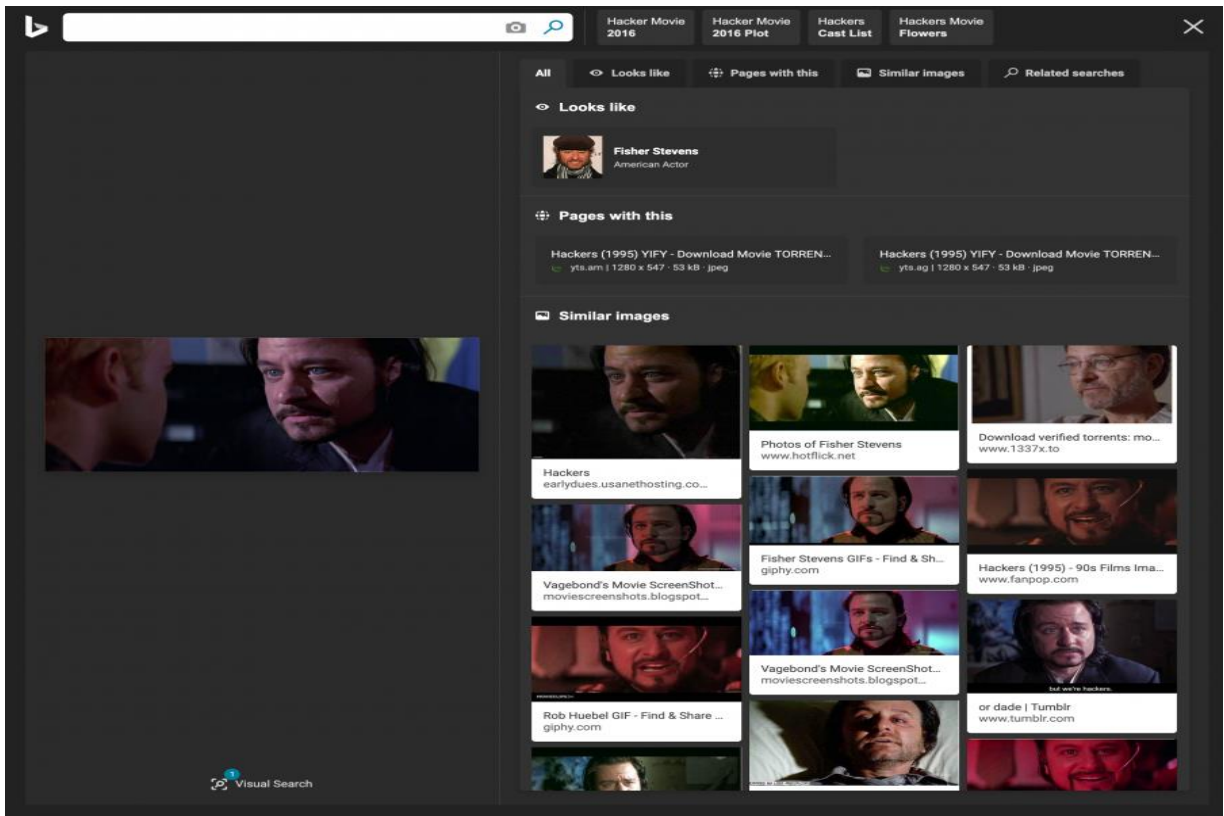


Рис. 3.11. Запит у Bing

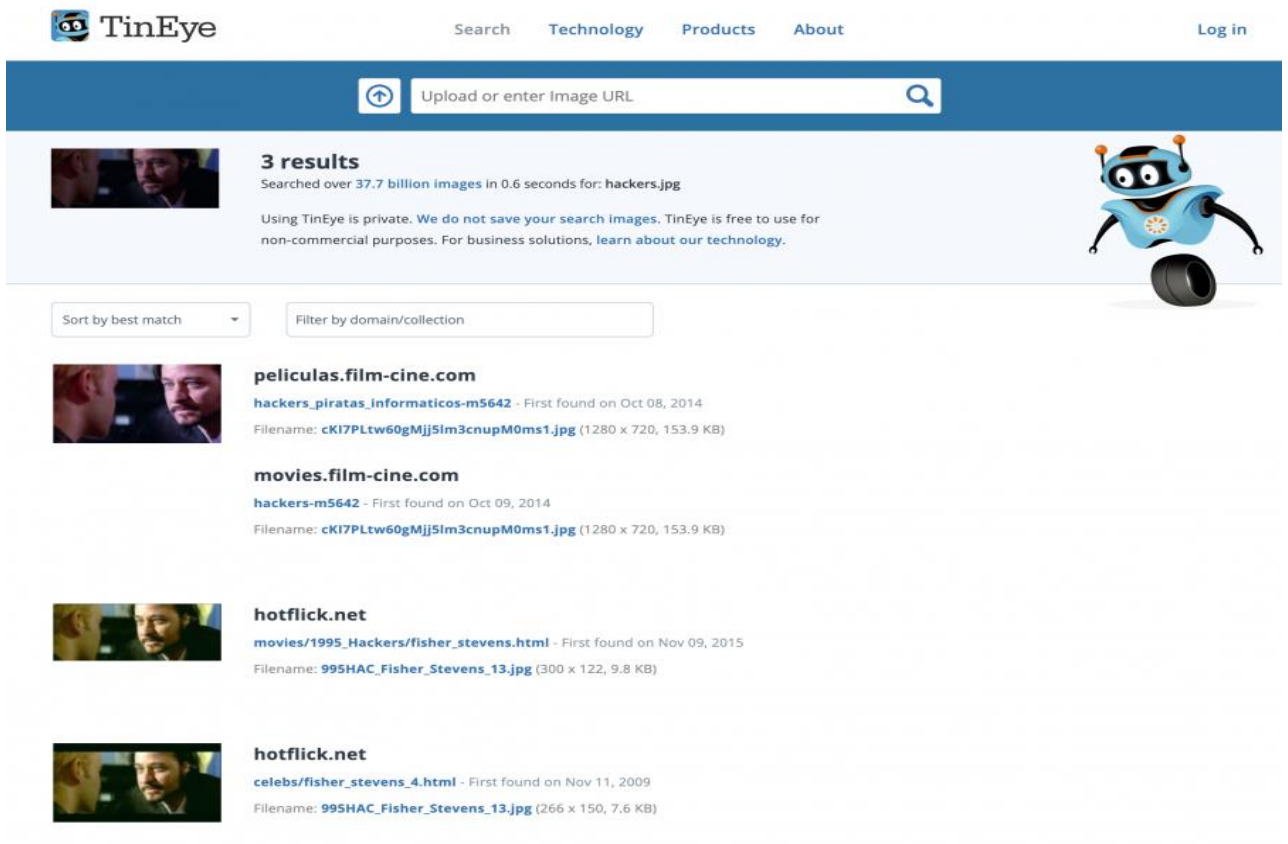


Рис. 3.12. Запит у TinEye

Google ідентифікував зображення як «злочинці-хакери», та видав посилання на фільм “Хакери”, що є чудовим результатом. Яндекс показав в основному однакові зображення з торрент сайтів, а також перелік інших зображень Фішера Стівенса, ідентифікував фільм “Хакери”. Bing також виконує завдання та ідентифікує Фішера Стівенса як актора на зображенні, помітив зображення як “хакерський фільм”. TinEye видав результати з торрент сайтів, що містять дане зображення у різних розмірах.

Логотип Культу Мертвої Корови:

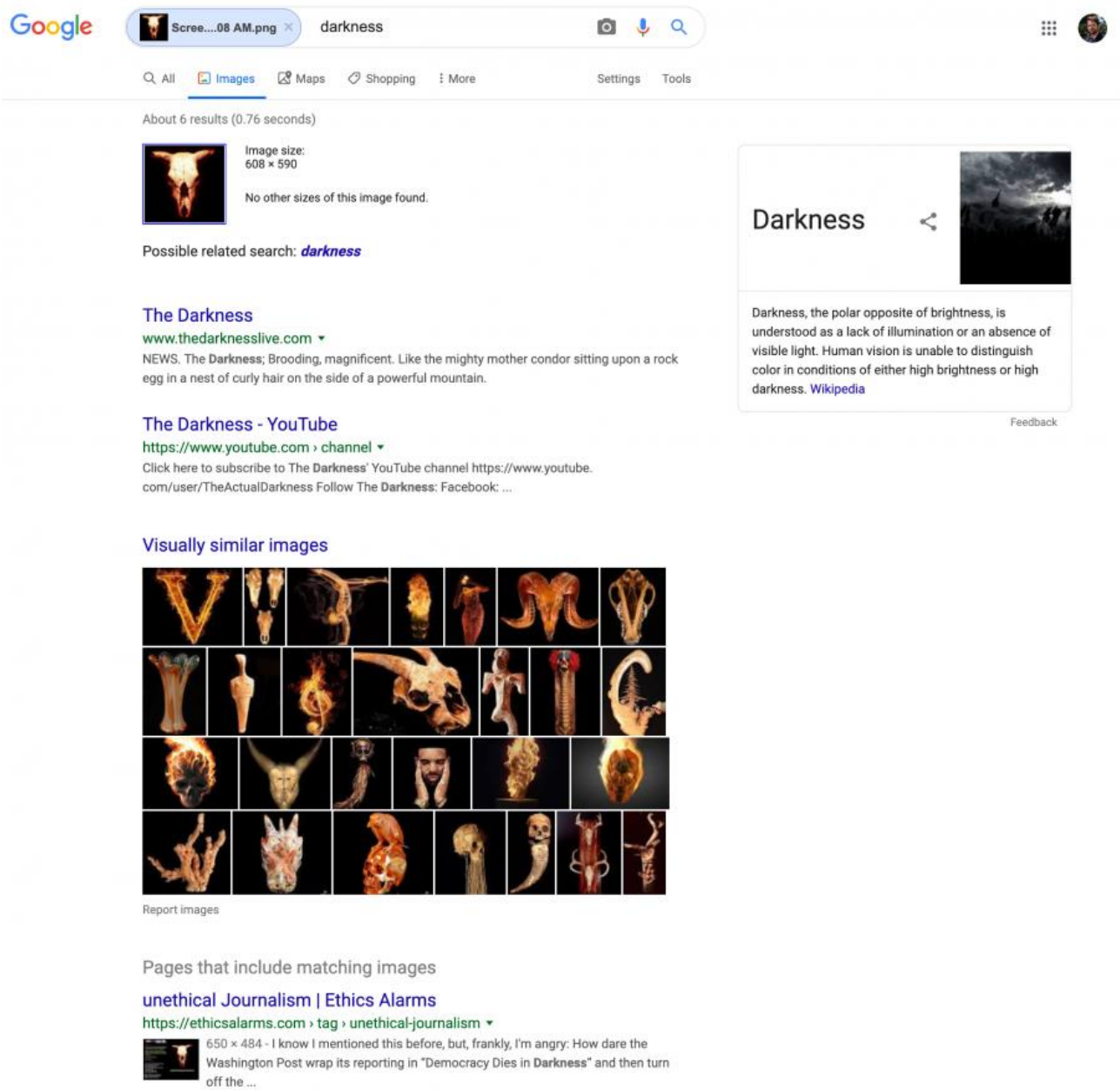


Рис. 3.13. Запит у Google

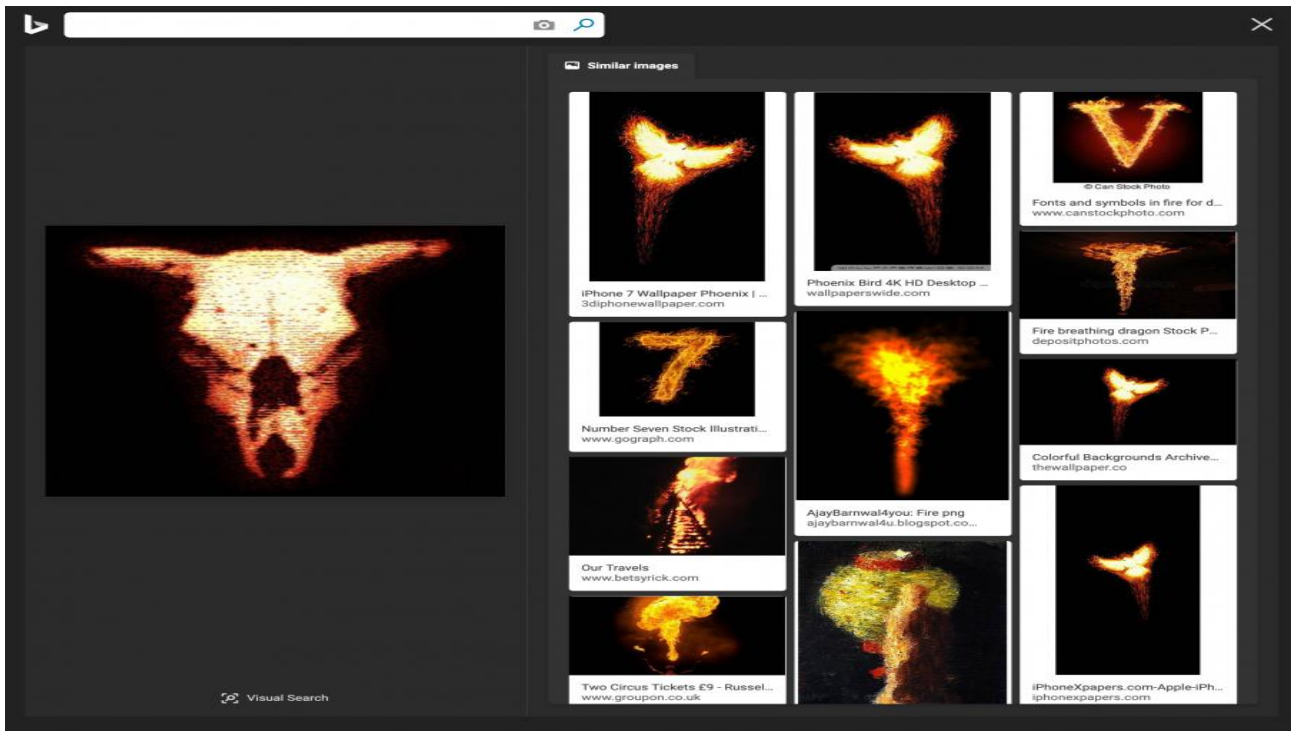


Рис. 3.14. Запит у Bing

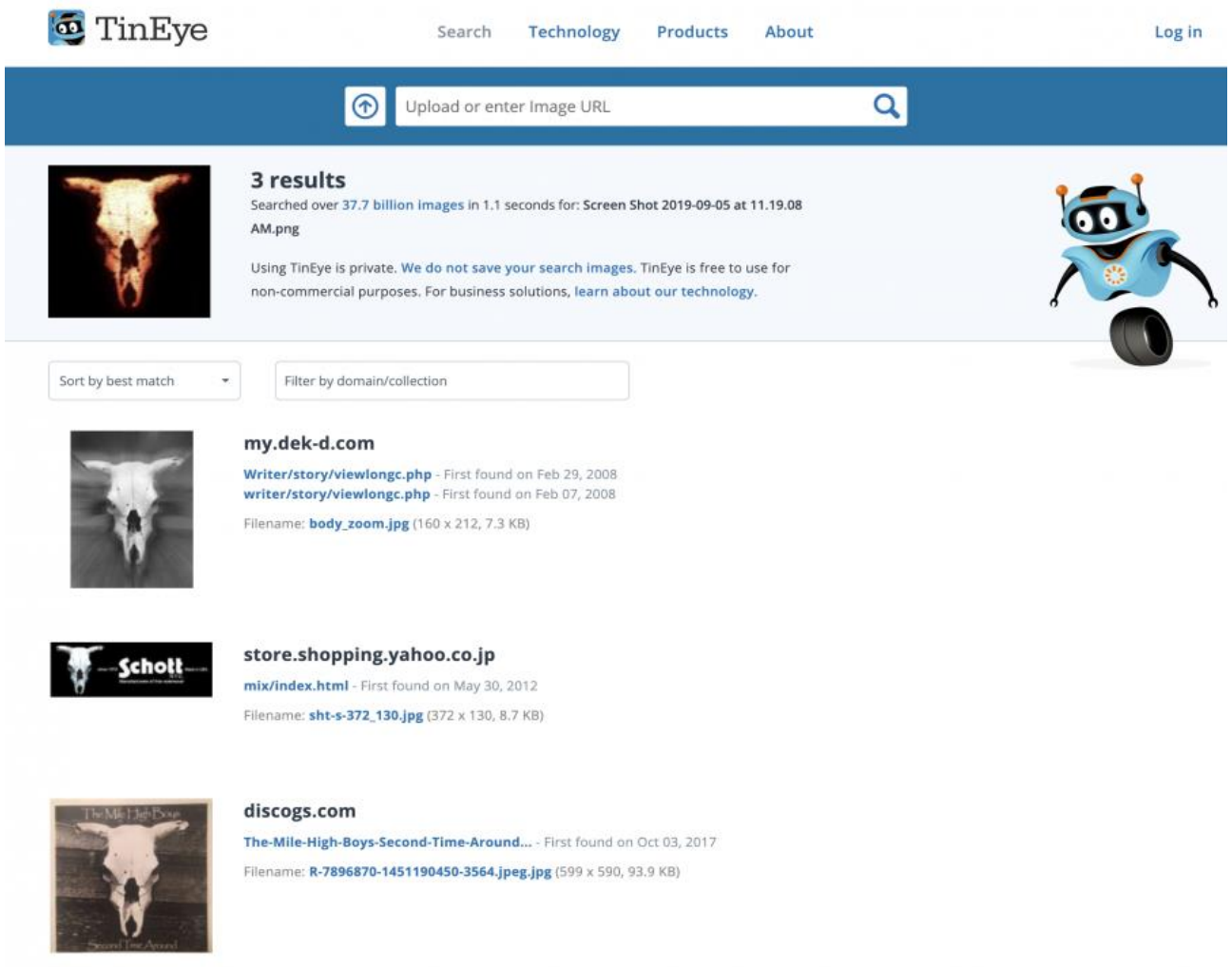


Рис. 3.15. Запит у TinEye

Google ідентифікував зображення як «тьма», та видав зображення у схожій кольоровій гамі. Bing показав у результатах зображення вогню та феніксів. TinEye видав результати зображень на основі заданого, але з певною обробкою.

Фотографія Центрального залізничного вокзала Мілану:

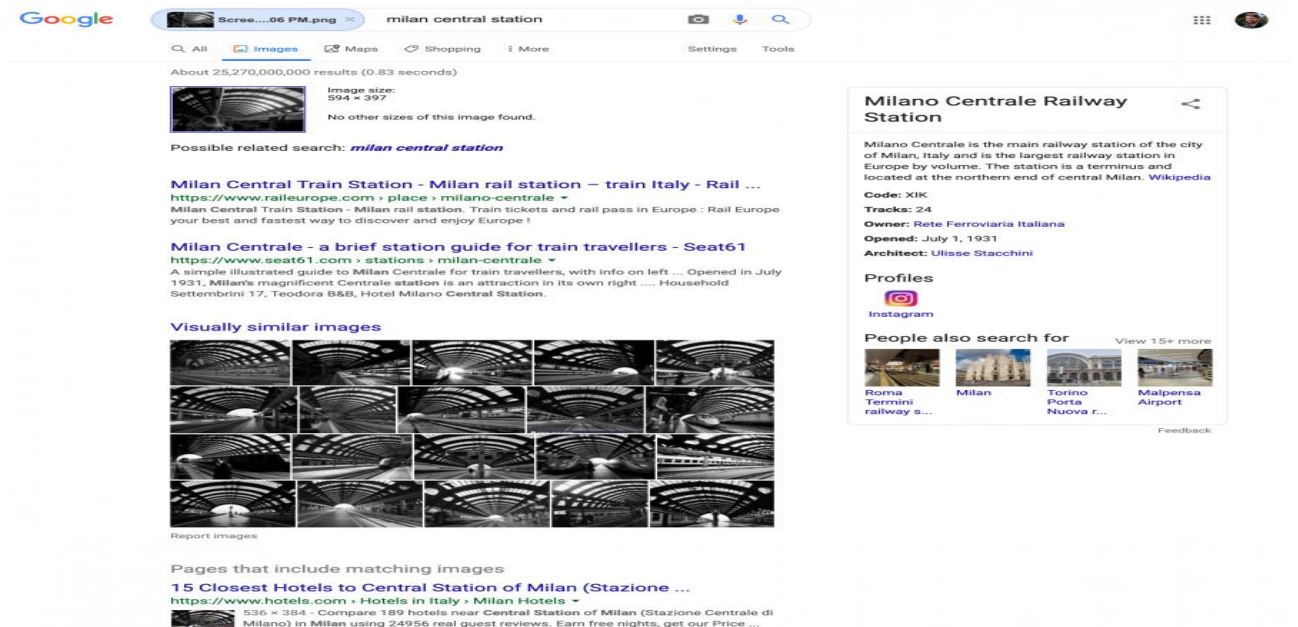


Рис. 3.16. Запит у Google

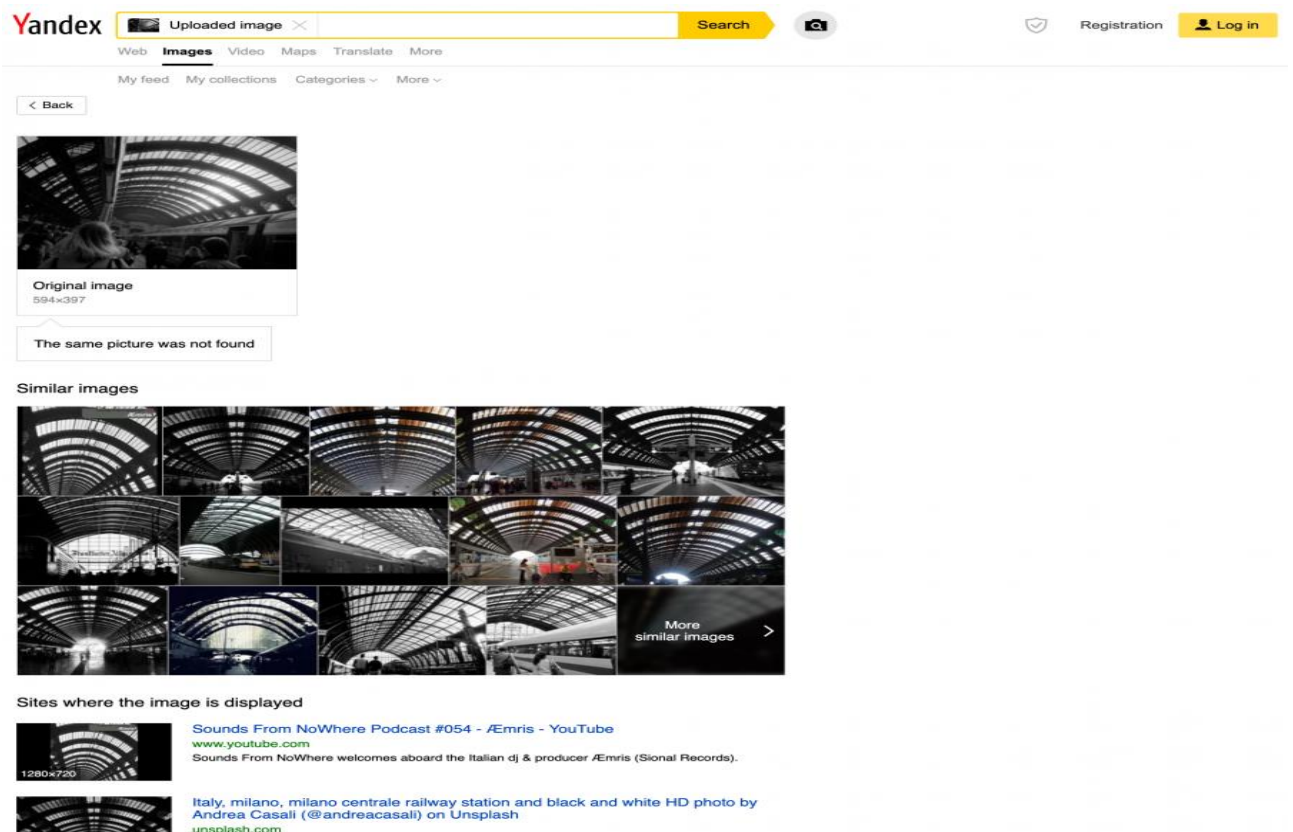


Рис. 3.17. Запит у Яндекс

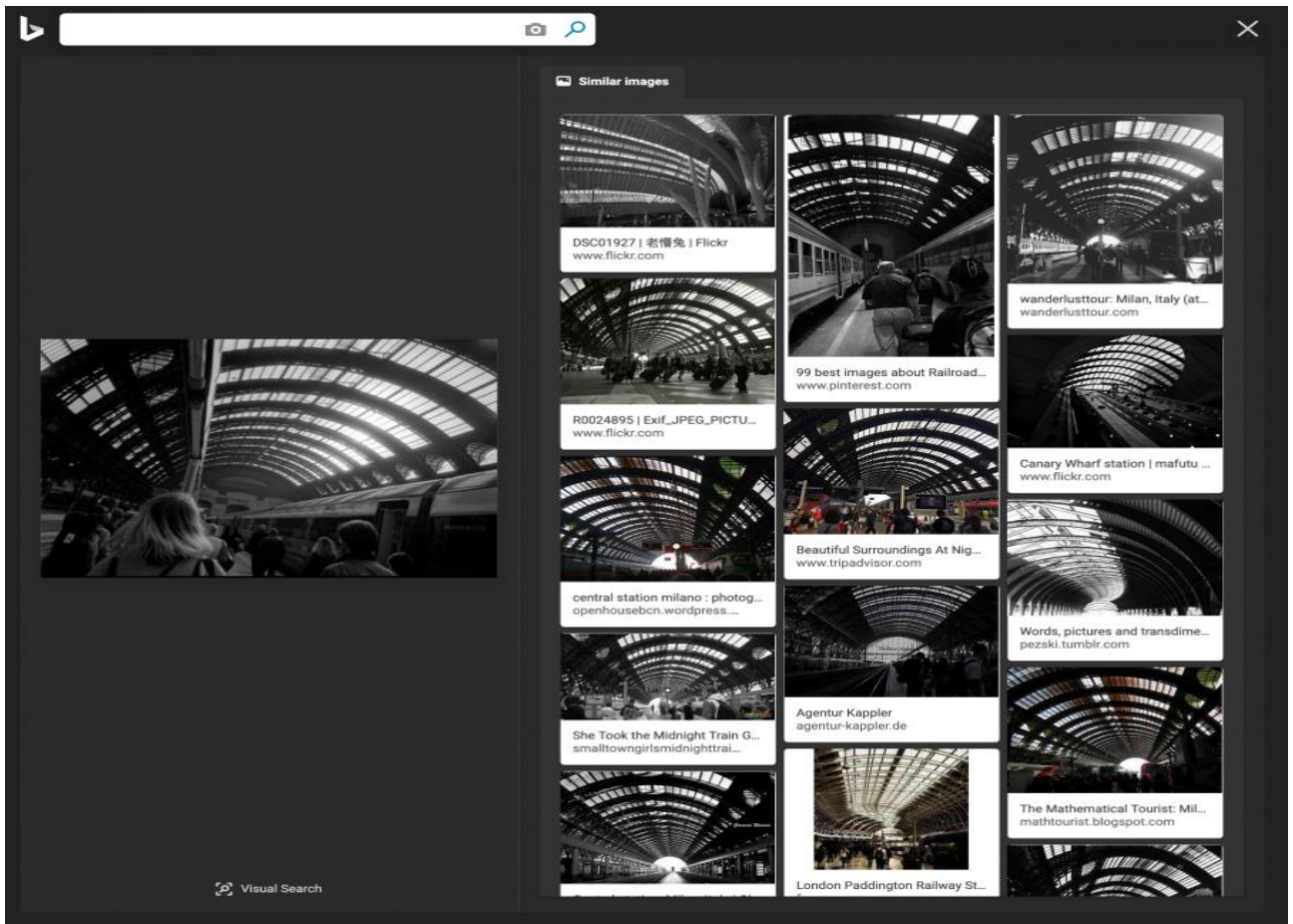
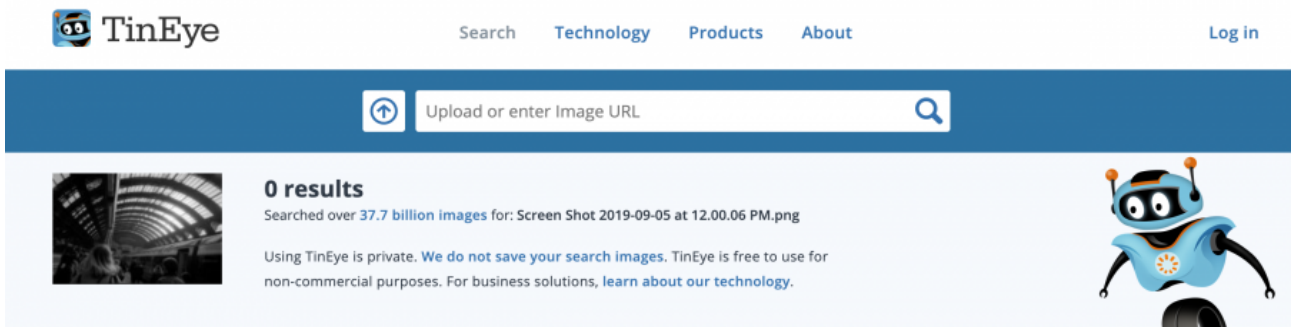


Рис. 3.18. Запит у Bing



Your image may be unique, and doesn't appear anywhere on the web. This happens with personal photographs, artwork, or other original images.

If you expected to find matches, TinEye may not have crawled the website or page where your image appears. TinEye constantly crawls the web and adds tens of millions of new images to the [TinEye index](#) every month.

TinEye finds exact and altered copies of images by looking at the actual pixels, but it doesn't know what the image represents (a chair, a waterfall, et). TinEye doesn't identify [people or objects](#) in an image. Learn more about [how TinEye works](#).

Рис. 3.19. Запит у TinEye

Google чітко ідентифікував місцезнаходження, та видав зображення об'єкту з інших ракурсів. Яндекс також впорався з завданням, але не прямо, а через фото у блозі. Bing показав у результатах зображення об'єкту з інших ракурсів, але не видав назви об'єкту. TinEye не знайшов подібних фото.

3.1. Висновки до третього розділу

Дивлячись на результати роботи, можна скласти наступну таблицю:

	Деталі	Обличчя	Будівля	Локація	Лого	Ресайз
Google	середньо	середньо	середньо	чудово	чудово	добре
Яндекс	добре	чудово	чудово	чудово	добре	добре
Bing	чудово	добре	добре	добре	добре	середньо
TinEye	середньо	середньо	середньо	середньо	чудово	чудово

За таблицею можна зробити висновок, що для пошуку локації та логотипа найкращим варіантом буде Яндекс та Google. У роботі з деталями на зображенні краще за інших показав себе Bing. Для ідентифікації людини за обличчям та пошуку певної будівлі безумовним лідером є Яндекс. Для специфічних завдань з пошуку інших варіантів зображення підійде TinEye.

Результати проведеного дослідження можуть бути використані для ефективного пошуку зображень, в залежності від типу зображення та бажаного результату. Знаючи, яка пошукова система виконує певну роботу краще, можна досягти значного скорочення часу, що витрачається на дослідження об'єкту, а також досягти нових результатів, що були недоступні раніше.

РОЗДІЛ 4. ЕКОНОМІЧНА ЧАСТИНА

4.1 Маркетингові дослідження

В рамках роботи було розглянуто пошукові системи, так як вони мають ряд значних переваг:

- низький рівень порогових знань для ефективного застосування;
- використання пошукових систем є безкоштовним;
- великий функціонал;
- велика кількість пошукових систем.

Матеріали, отримані внаслідок використання пошукових систем не є конфіденційними, бо знаходяться у відкритому доступі, їх використання не порушує діюче законодавство України.

Завдяки цим особливостям пошукові системи є сильним інструментом здобуття інформації для корпоративної розвідки, на основі яких можна створити певне програмне забезпечення для монетизації.

Корпоративна розвідка є дуже актуальною сферою зараз, оскільки перевірка потенційних партнерів та визначення бекграунду компанії або людини є гарантією вдало побудованих корпоративних відносин, що принесуть прибуток. Таким чином, виконане дослідження може підвищити ефективність роботи компаній, які спеціалізуються на корпоративній розвідці, що, у свою чергу, призведе до значного зросту ринку надання послуг корпоративної розвідки.

4.2 Економічний ефект

Оскільки в даній роботі була розглянута тільки можливість використання пошукових систем, але не було розроблено відповідне програмне забезпечення,

розрахунок економічного ефекту, який складається з витрат до впровадження ПО і після нього, неможливий. Тому розглядається тільки соціальний ефект.

В експериментальній частині було проведено дослідження відповідей пошукових систем на різні типи зображень. Це показало можливість використання даного методу як додатковий засіб ідентифікації об'єктів дослідження, що призводить до отримання нової інформації та економії часу.

4.3 Висновок

До теперішнього часу розроблені елементи теорії і методика розрахунку ефективності обробки зображень пошуковими системами. Дослідження в цій області продовжуються і зараз. Їх метою є підвищення ефективності ідентифікації об'єктів на зображенні, отримання нових точних результатів та зменшення часу, що витрачається на пошук інформації.

Всебічне вивчення існуючих технологічних процесів дозволяє встановити основні їх недоліки, слабкі сторони і намітити раціональні шляхи їх усунення.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. An Introduction to Computer Security: The NIST Handbook. Draft. National Institute of Standards and Technology, Technology Administration, U. S. Department of Commerce, 1994. - 310 p.
2. Cheswick W. R., Bellovin S. M. Firewalls and Internet Security: Repelling the Wily Hacker. - Addison-Wesley, 1994. - 275 с.
3. Браун С. Мозаика и Всемирная паутина для доступа к Internet: Пер. с англ. М.: Мир, Малип, СК Пресс, 1996. - 167с.
4. Гайкович В., Першин А. Безопасность электронных банковских систем. М.: Единая Европа, 1994. - 264 с.
5. Гилстер П. Новый навигатор Internet: Пер с англ. Киев: Диалектика, 1996. - 495 с.
6. Игер Б. Работа в Internet Под ред. Тихонова А. М.: БИНОМ, 1996. - 313 с.
7. Кент П. Internet / Пер. с англ. В. Л. Григорьева. М.: Компьютер, ЮНИТИ, 1996. - 267 с.
8. Колесников О. Э. Интернет для делового человека. М.: МЦФ. Издат. фирма Яуза, 1996. - 281 с.
9. Крол Эд. Все об Internet: Руководство и каталог. Киев: BNV, 1995. - 591 с.
10. Левин В. К. Защита информации в информационно-вычислительных системах и сетях // Программирование. - 1994. - N5. - С. 5-16.
11. Нольден М. Ваш первый выход в Internet: Для начинающих пользователей Internet и широкого круга пользователей РС. Гл. ред. Кондукова Е. В. Спб.: ИКС, 1996. - 238 с.
12. Грушо А.А. Тимонина Е.Е. Теоретические основы защиты информации - М.: Издательство "Яхтсмен", 1996. - 192 с.
13. Диева С.И., Шаваева А.Г. Организация и современные методы защиты информации. - М.: Концерн "Банковский Деловой Центр", 1998. - 472с.

14. Завгородний В.И. Комплексная защита информации в компьютерных системах -
15. Мельников В.П., Клейменов С.А., Петраков А.М. Информационная безопасность - М.: Академия, 2005. -332 с.
16. Садердинов А.А., Трайнев В.А., Федулов А.А. Информационная безопасность предприятия - М.: Издательский дом "Дашков и К", 2005.- 335 с.
17. Семкин С.Н., Беляков Э.В., Гребенев С.В. и др. Основы организационного обеспечения информационной безопасности объектов информатизации - М.: Издательство "Гелиос АРВ", 2005. - 186 с.
18. Степанов Е.А., Корнеев И.К. Информационная безопасность и защита информации - М.: ИНФРА-М, 2001. - 304 с
19. В.М. Вдовин, Л.Е. Суркова. Информационные технологии в финансово-банковской сфере. – М.: Дашков и Ко, 2012. – 304 с.
20. Г.Сенкевич. Информационная система малого предприятия "с нуля". Самое необходимое. – СПб.: БХВ-Петербург, 2012. – 400 с.
21. С.В. Мхитарян. Маркетинговая информационная система. – М.: Эксмо, 2006. – 336 с.
22. Т.Л. Партыка, И.И. Попов. Информационная безопасность. – М.: Форум, Инфра-М, 2002. – 368 с.
23. Т.П. Барановская, В.И. Лойко, М.И. Семенов, А.И. Трубилин. Информационные системы и технологии в экономике 2-е изд. – М.: Финансы и статистика, 2006. – 416 с.
24. В.К. Чаадаев, И.В. Шеметова, И.В. Шibaева. Информационные системы компаний связи. Создание и внедрение. – М.: Эко-Трендз, 2004. – 256 с.
25. А.Г. Серго, В.С. Пуцин. Основы права интеллектуальной собственности. Курс лекций. Учебное пособие. – М.: Интернет-университет информационных технологий, 2005. – 344 с.
26. О.Е. Филинова. Информационные технологии в рекламе. – М.: КУДИЦ-Образ, 2006. – 240 с.

27. В.А. Грабауров, С.В. Грабауров, В.Н. Гулин, В.В. Лабозкий. Информационные технологии. – М.: Современная школа, 2006. – 432 с.
28. А.Н. Бородулин, А.Ю. Заложнев, Е.Л. Шуремов. Внутрифирменное управление, учет и информационные технологии. – М.: ПМСОФТ, 2006. – 344 с.
29. Завельский М.Г. (Ред.). Экономическая политика и ее инструменты. Системное моделирование и информационные технологии. – М.: КомКнига, 2006. – 152 с.
30. О.О. Борисова. Рекламно-информационные технологии библиотечной деятельности. – М.: Профессия, 2006. – 320 с.
31. В.С. Алиев. Информационные технологии и системы финансового менеджмента. – М.: Инфра-М, 2007. – 320 с.
32. В.А. Галатенко. Стандарты информационной безопасности. – М.: Интернет-университет информационных технологий, 2006. – 264 с.
33. Экономико-математические исследования. Математические модели и информационные технологии. – М.: Наука, 2000. – 343 с.
34. Модели и информационные технологии хозяйственной политики. Том 30. – М.: ЛКИ, 2007. – 162 с.
35. В.Ф. Мартыненко, Г.М. Вялкова, В.А. Полесский, Е.Н. Беляев, В.А. Гройсман, И.Ф. Серегина. Информационные технологии в управлении здравоохранением Российской Федерации. – М.: ГЭОТАР-Медиа, 2009. – 248 с.
36. Т.В. Макарова, О.Н. Ткаченко, О.Г. Капустина. Основы информационных технологий в рекламе. – М.: Юнити-Дана, 2009. – 272 с.
37. М.Ф. Меняев. Информационные технологии управления. Книга 2: Информационные ресурсы. – М.: Омега-Л, 2003. – 432 с.
38. Аркадий Воловник. Знакомьтесь, информационные технологии. – СПб.: БХВ-Петербург, 2002. – 352 с.
39. М.А. Борисов, О.А. Романов. Основы организационно-правовой защиты информации. Учебное пособие. – М.: Ленанд, 2016. – 248 с.
40. А.А. Бирюков. Информационная безопасность. Защита и нападение. – М.: ДМК Пресс, 2016. – 474 с.

41. Борисов М.А., Романов О.А. Основы организационно-правовой защиты информации. – М.: , 2016. – с.
42. А.А. Бирюков. Информационная безопасность. Защита и нападение. – М.: ДМК Пресс, 2017. – 434 с.
43. А.А. Малюк. Информационная безопасность. Концептуальные и методологические основы защиты информации. Учебное пособие. – М.: Горячая Линия - Телеком, 2004. – 280 с.
44. А.А. Анисимов. Менеджмент в сфере информационной безопасности. – М.: Интернет-университет информационных технологий, Бином. Лаборатория знаний, 2010. – 176 с.
45. А.А. Малюк, С.В. Пазизин, Н.С. Погожин. Введение в защиту информации в автоматизированных системах. – М.: Горячая Линия - Телеком, 2005. – 147 с.

ДОДАТОК А

ВІДГУК**керівника економічного розділу****на кваліфікаційну роботу магістра****на тему: « Розробка програмного забезпечення для дослідження
ефективності обробки запитів пошукових систем та їх функції у****корпоративній розвідці.»****студента групи 121м-19-1 Ігнашева Василя Олександровича****Керівник економічного розділу
доцент каф. ПЕП та ПУ, к.е.н.****Л. В. Касьяненко**

ПЕРЕЛІК ДОКУМЕНТІВ НА ОПТИЧНОМУ НОСІЇ

Ім'я файла	Опис
Пояснювальні документи Диплом_Ігнашев.doc	Пояснювальна записка до магістерської роботи. Документ Word.
Диплом_Ігнашев.pdf	Пояснювальна записка до магістерської роботи в форматі PDF
Презентація Презентація_Ігнашев.pptx	Презентація до магістерської роботи