

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Черкаський національний університет
імені Богдана Хмельницького
Черкаський інститут банківської справи
Чорноморський державний університет імені Петра Могили

*Всеукраїнська науково-практична
Інтернет-конференція*

**Автоматизація та комп'ютерно-
інтегровані технології у
виробництві та освіті:
стан, досягнення,
перспективи розвитку**

13-19 березня 2023 року

м. Черкаси

серед вбудованих та користувацьких. Привертає увагу функціонал накопичення термінів, які не потребують перекладу, та усталених фраз чи власних словосполучень з індивідуальним перекладом. Також наявний переклад слів в контексті.

Переважає більшість професійних надавачів послуг з перекладу працюють із середовищем TRADOS. Інтерфейс цієї програми є вже більш складним, проте, якщо докласти зусиль та трохи більше часу, то її також можна освоїти. Зручна система навігації надає функціонал для ефективного опрацювання лінгвістичних баз. Також користувач може створити власний словник термінів.

Таким чином, виконані дослідження програмних засобів для лінгвістичного супроводу студентської наукової діяльності показали, що кожне середовище має свої переваги та недоліки, проте обирати потрібно кожному індивідуально. Звичайно, що студентам варто розпочинати з простіших програм, таких як ABBYY Lingvo чи PRAGMA 6.X. Хто хоче більше заглибитись в сферу перекладу науково-технічної літератури, необхідно ознайомлюватись з такими програмами як PROMT 18 Master чи TRADOS, які містять більше функцій та мають великий обсяг тематичних словників.

1. Гаращук І. Критерії оцінювання комп'ютеризованих середовищ опрацювання іншомовних інформаційних ресурсів. Тези доповідей студентської наукової конференції УАД. Львів, 2022. С. 8.

Гук Віталій, кандидат технічних наук, старший викладач кафедри програмного забезпечення автоматизованих систем, Черкаський національний університет імені Б. Хмельницького, м. Черкаси, України

Наконечна Оксана, кандидат технічних наук, доцент кафедри комп'ютерних наук та інформаційних технологій Житомирського державного університету ім. І. Франка, м. Житомир, України

ТЕМАТИЧНИЙ АНАЛІЗ ІНФОРМАЦІЙНОГО ПОШУКУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Центральна проблема інформаційного пошуку - надання інформації користувачу в конкретній предметній області. Параметри і апарат пошуку представляє класичну задачу інформаційного пошуку.

Звичайно, це пошук документів, що задовольняють запиту, в рамках деякої колекції документів.

Розглядаючи задачу пошуку документів за зразком [1] можна виділити наступні етапи: визначення тематики документа з призначенням вагових коефіцієнтів і обчислення тематичної близькості документів.

Реалізація задач пошуку документів за зразком дозволяє підвищити якість і ефективність такого пошуку. Розглядаючи пошук документів в Інтернеті, сформулюємо актуальність даної прикладної задачі з урахуванням специфічності середовища пошуку, а саме: об'єм доступних інформаційних ресурсів Інтернету, високий ступінь їх оновлення, взаємозв'язок сторінок між собою, відсутність централізованого адміністрування інформаційних ресурсів, надмірність інформаційних ресурсів, об'єднання в Інтернет численних груп користувачів, різних за кваліфікацією.

До найпростіших моделей пошуку відноситься модель дескрипторного пошуку і модель, заснована на Дублінському ядрі.

Дублінське ядро (Dublin Core) [2 - 3] – це набір елементів метаданих, значення яких зафіксовано в специфікації стандарту, що його визначає. В термінах значень цих елементів можна описувати зміст різного роду текстових документів.

У моделі, заснованій на класифікаторах, документ представляється у вигляді сукупності асоційованих з ним атрибутів. Атрибути є ідентифікатори класів, до яких відноситься даний документ. Класи формують ієрархічну структуру класифікатора.

У булевих моделях пошуку користувач може формулювати запит у вигляді булевого виразу, використовуючи для цього оператори І, АБО, НІ. Терми запиту залежать від конкретного варіанту моделі пошуку.

Векторні моделі в даний час є найпоширенішими і найчастіше використовуються в практиці пошуку. Векторні моделі, на відміну від булевих, легко дозволяють ранжирувати результуючу множину документів запиту. Суть таких моделей зводиться до представлення документів і запитів у вигляді векторів.

Моделі вірогідності вперше були запропоновані в 1960 році [4]. В їх основі лежить принцип ранжирування вірогідності (Probabilistic Ranking Principle, PRP). Цей принцип полягає в наступному -

щонайвища загальна ефективність пошуку досягається у разі, коли результуючі документи ранжируються за зменшенням вірогідності їх релевантності запиту.

Так само, як і моделі вірогідності, мережі висновку засновані на принципі ранжирування вірогідності результуючих документів пошуку [5]. Головна їх відмінність від моделей вірогідності полягає в тому, що використовується оцінка не вірогідності релевантності документа запиту, а вірогідність того, що він задовольняє інформаційним потребам користувача.

Всю сукупність представлених на сьогоднішній день методів тематичного аналізу тексту можна розділити на дві групи: лінгвістичний аналіз; статистичний аналіз.

Перший метод орієнтований на виділення суті тексту по його семантичній структурі. Другий метод аналізує текст по частотному розподілу слів.

Очевидно, що наближеність того або іншого конкретного документа до інформаційних потреб користувача залежить від контексту, в рамках якого відбувається пошук.

Однією з найважливіших складових загального контексту запиту є тематичний. Інформацію про тематичну орієнтованість даних можна використовувати в спеціалізованих методах інформаційного пошуку. Такі методи одержали назву тематико-орієнтованих методів інформаційного пошуку (пошук документів заданої тематики, маршрутизація запитів, пошук за документом-зразком, тематична класифікація документів).

Однією з задач інформаційного пошуку є задача пошуку за документом-зразком. Документ-зразок виступає як одна з форм представлення інформаційних потреб користувача. Метою пошуку є виявлення тематично подібних документів.

Не дивлячись на велику кількість різних рішень, відсутня чітко відпрацьована методологія пошуку за документом-зразком. Існуючі підходи не забезпечують повною мірою рішення цієї задачі. Більшою мірою вони є суміжними по відношенню до даної задачі. Відсутня достатня кількість спеціалізованих досліджень, присвячених рішенню саме цього питання. Більш того, відсутня чітка формалізація постановки задач такого пошуку.

Розглянемо розроблені методи і підходи, що використовуються при рішенні задач пошуку документів за зразком.

Пропонується наступна послідовність дій: для кожного документа визначається деяка відносно невелика множина документів, що представляє його апроксимоване тематичне оточення; побудовані тематичні оточення аналізуються з метою формування множини ключових слів, що характеризують тематику початкового документа щодо інших документів колекції; одержані набори ключових слів використовуються для подальшого обчислення відносних оцінок тематичної подібності.

Не дивлячись на особливості середовища Інтернету, або багато в чому завдяки таким особливостям, існують вельми цікаві і оригінальні варіанти реалізації пошуку документів за зразком в Інтернеті.

Одним з таких варіантів є використання інформації про структуру посилань. В загальному випадку реалізація такого варіанту припускає аналіз структури графа Інтернету (вершинами якого виступають сторінки, а ребрами - посилання). Як документ-зразок виступає деяка сторінка. Посилання на дану сторінку і посилання з неї використовуються в різних алгоритмах локального аналізу структури графа Інтернету.

Ще одним цікавим варіантом рішення задачі пошуку документів за зразком є використання документа на додаток до традиційного запиту.

Загальну схему пошуку за документом-зразком можна представити у вигляді наступної послідовності:

- виконується попередній відбір з колекції документів, і потім для відібраних документів обчислюється тематична подібність;
- обчислені оцінки тематичної близькості w_1, \dots, w_n використовуються при ранжируванні документів за тематичною подібністю до документа-зразка.

У даному випадку рішення цієї проблеми базується на представленні запиту користувача у вигляді документа-зразка і реалізації методу ефективного аналізу тематики документів. Як критерій ефективності виступає точність. Метод тематичного аналізу, що розробляється, повинен точніше ідентифікувати тематику документів.

Задачі пошуку документів за зразком передбачають рішення двох основних задач: виділення тематики документів; обчислення тематичної подібності документів.

Обидві ці задачі відносяться до задач класифікації – віднесення документа за його тематичним представленням до деякого класу і визначення міри подібності між різними класами документів.

Аналіз існуючих досліджень щодо реалізації пошуку документів за зразком виявив незначне число готових і апробованих рішень в даній області, що багато в чому пов'язано з відсутністю достатньо відпрацьованої теорії і практики рішення задач тематичного аналізу неструктурованої природно-мовної текстової інформації довільного змісту.

Рішення задач тематичного аналізу є актуальним не тільки в області інформаційно-пошукових систем, але і взагалі в системах обробки і аналізу інформації. Це широкий спектр різних задач інтелектуальної обробки інформації, у тому числі задач добування, ідентифікації і розпізнавання смислового змісту мови. Все це обумовлює актуальність і значущість досліджень в області тематичного аналізу і обробки неструктурованої інформації.

Список використаних джерел

1. Вавіленкова А. І. Теоретичні основи аналізу електронних текстів: монографія. К.: ТОВ «СІК ГРУП УКРАЇНА», 2016. 192 с.
2. Федько В. В. Організація баз даних та знань : навч.-практ. посіб. для самост. підготовки студ. / В. В. Федько, О. В. Тарасов, М. Ю. Лосев. Харків: ХНЕУ, 2013. 198с.
3. Maron M.E., Kuhns J.L. On relevance, probabilistic indexing and information retrieval. Journal of the ACM, No. 7, 1960, pp. 216-244.
4. Singhal A. Modern Information Retrieval: A Brief Overview. Data Engineering Bulletin, IEEE Computer Society, Vol. 24, No. 4, December 2011, pp. 35-43.
5. Введення у пошукову оптимізацію. Режим доступу: <https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=uk>.

<i>Гаращук Ірина Василівна</i> ДОСЛІДЖЕННЯ СЕРЕДОВИЩ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ ДЛЯ ОПРАЦЮВАННЯ ІНШОМОВНИХ ДЖЕРЕЛ	79
Гук Віталій, Наконечна Оксана ТЕМАТИЧНИЙ АНАЛІЗ ІНФОРМАЦІЙНОГО ПОШУКУ ТЕКСТОВОЇ ІНФОРМАЦІЇ	81
Старанчук Остап Ігорович, Боровик Олег Васильович АКТУАЛЬНІСТЬ ЗАДАЧІ ТА МОЖЛИВИЙ ПІДХІД ЩОДО УДОСКОНАЛЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ ФРАЗЕОЛОГІЧНИХ ОДИНИЦЬ В АНГЛОМОВНИХ ТЕКСТАХ	86
<i>Боровик Олег Васильович, Боровик Людмила Володимирівна</i> СТРУКТУРА СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА СПРОЩЕННЯ КОНТРОЛЬНИХ ПРОЦЕДУР У ПУНКТАХ ПРОПУСКУ	88
<i>Бойківська Галина Миколаївна, Замулко Ірина Василівна</i> АВТОМАТИЗОВАНЕ КЕРУВАННЯ БІЗНЕС-ПРОЦЕСАМИ В УПРАВЛІННІ ПЕРСОНАЛОМ	90
<i>Малиновський Микита Ігорович, Міхєєнко Денис Юрійович</i> РОЗРОБКА ВЕБ- ОРІЄНТОВАНОЇ СИСТЕМИ ДЛЯ ОПТИМІЗАЦІЇ ТА АНАЛІТИКИ РОБОТИ ВОЛОНТЕРІВ З ВИДАЧІ ГУМАНІТАРНОЇ ДОПОМОГИ ЦИВІЛЬНОМУ НАСЕЛЕННЮ	92
<i>Косолапов Анатолій Аркадійович</i> ПЕРСПЕКТИВИ ПРОЕКТУВАННЯ СОЦІО- КОМП'ЮТЕРНО-ІНТЕГРОВАНИХ СИСТЕМ АВТОМАТИЗАЦІЇ ПІДПРИЄМСТВ В УКРАЇНІ.....	94
<i>Гітис Веніамін Борисович, Ляхов Богдан Андрійович</i> РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ОПТИМІЗАЦІЇ БІБЛІОТЕЧНИХ ФОНДІВ НАВЧАЛЬНИХ ЗАКЛАДІВ	96
<i>Разживін Олексій Валерійович, Люта Анастасія Володимирівна, Ільїнський Михайло Ігорович</i> ДОСЛІДЖЕННЯ МЕРЕЖІ ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБЛІКУ СПОЖИВАННЯ ЕНЕРГЕТИЧНИХ РЕСУРСІВ МІКРОРАЙОНУ	99
Секція 5. Комп'ютерне проектування та моделювання технологічних процесів 102	
Кисельова Олеся Борисівна, Губіна Вікторія Костянтинівна ВІЗУАЛЬНИЙ СТОРИТЕЛІНГ В ОСВІТНЬОМУ ПРОЦЕСІ.....	103

Levkin D., Kotko Ya., Levkin A. MODERN MATHEMATICAL MODELING OF BIOTECHNOLOGY PROCESSES.....	105
<i>Павленко Тетяна Степанівна</i> ДИСЕРТАЦІЙНІ ДОСЛІДЖЕННЯ З ПРОБЛЕМ ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У НАВЧАННІ ТА УПРАВЛІННІ ОСВІТНІМ ПРОЦЕСОМ.....	107
Секція 6. Інформаційні технології в навчанні та управлінні навчальним процесом	110
<i>Кривонос О.М.</i> ВИКОРИСТАННЯ МІКРО-МОВ В ПРОЦЕСІ ВИВЧЕННЯ ПРОГРАМУВАННЯ.....	111
Мельник Руслан, Гладка Людмила Іванівна, Дідук Віталій Андрійович РОЗРОБКА МОБІЛЬНОГО ДОДАТКУ ДЛЯ ІНТЕРАКТИВНОГО ВИВЧЕННЯ АНГЛІЙСЬКОЇ МОВИ.....	114
<i>Новицька Тетяна Леонідівна</i> SEMANTIC SCHOLAR – НОВА НАУКОВА БІБЛІОТЕКА НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ.....	116
Держевецька М.А., Рекова Н.Ю. ВИКОРИСТАННЯ СИСТЕМ КОМП'ЮТЕРНОЇ АЛГЕБРИ В НАУКОВИХ ДОСЛІДЖЕННЯХ.....	119
<i>Svitlana Starikova</i> FEATURES OF THE USE OF INFORMATION TECHNOLOGIES IN EDUCATION OF SCHOOL STUDENTS IN WARTIME.....	121
<i>Павленко Тетяна Степанівна</i> ДИСЕРТАЦІЙНІ ДОСЛІДЖЕННЯ З ПРОБЛЕМ ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У НАВЧАННІ ТА УПРАВЛІННІ ОСВІТНІМ ПРОЦЕСОМ.....	124
Постернак І.М., Постернак О.С. ІНФОРМАЛЬНА ОСВІТА ЗА ДОПОМОГОЮ ПЛАТФОРМИ COURSERA	126
<i>Регета Любов Максимівна</i> ЗАСТОСУВАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У НАВЧАННІ УКРАЇНСЬКОЇ МОВИ.....	128
Нечволода Людмила Володимирівна, Трофименко Данил Дмитрович ОЦІНКА ЕФЕКТИВНОСТІ РОБОТИ WEB-САЙТУ	130
<i>А.В. Ткаченко, Т.В. Ткаченко</i> ОНОВЛЕННЯ ЗМІСТУ ПРАКТИКУМУ З «МЕТОДИКИ НАВЧАННЯ ФІЗИКИ У СУЧАСНІЙ ШКОЛІ» З ВИКОРИСТАННЯМ ІКТ	132

Хміль Н.А., Хміль А.М. ЦИФРОВИЙ ІНСТРУМЕНТАРІЙ ДЛЯ СТВОРЕННЯ МУЛЬТФІЛЬМІВ В КУРСІ «ЦИФРОВИЙ ДИЗАЙН»	139
Гладка Людмила Іванівна, Дідук Віталій Андрійович, Гладкий Антон Андрійович ВИКОРИСТАННЯ ЗМІШАНОГО НАВЧАННЯ ПРИ ВИВЧЕННІ ДИСЦИПЛІНИ “АЛГОРИТМІЗАЦІЯ І ПРОГРАМУВАННЯ” У ЗАКЛАДАХ ВИЩОЇ ОСВІТИ. 141	
Шинкура Л.М. ВИКОРИСТАННЯ ЦИФРОВИХ ІНСТРУМЕНТІВ	143
<i>Безух Дмитро Станіславович</i> СИСТЕМА ПІДТРИМКИ ПРОЦЕСУ РОЗРОБКИ РОБОЧИХ ПРОГРАМ НАВЧАЛЬНИХ ДИСЦИПЛІН ДЛЯ ЗАКЛАДІВ ВИЩОЇ ОСВІТИ	144
<i>Романенко Т. В., Русіна Н. Г.</i> ПЕРЕВАГИ ТА НЕДОЛІКИ ВИКОРИСТАННЯ CHATGPT У НАВЧАЛЬНОМУ ПРОЦЕСІ	146
<i>Іванова Світлана Миколаївна</i> РЕФЕРАТИВНО-АНАЛІТИЧНА БАЗА ДАНИХ DIMENSIONS –	148
<i>Кільченко Алла Віленівна</i> ВИКОРИСТАННЯ ВЕБРЕСУРСУ ЕЛЕКТРОННОЇ БІБЛІОТЕКИ НАПН УКРАЇНИ В УМОВАХ ВОЄННОГО СТАНУ	151
Васильєва А. А., Жуковський С.С. PROCESSING—ІНСТРУМЕНТ ДЛЯ ПРОГРАМУВАННЯ ІНТЕРФЕЙСІВ, АНІМАЦІЙ ТА ЗОБРАЖЕНЬ	153
Секція 7. Інтелектуальні системи та машинне навчання.....	157
<i>Hanna Kostrova</i> METHODS OF ADAPTIVE FUZZY CONTROL OF ROBOTS BASED ON NEURAL NETWORKS.....	158
<i>Gitis Veniamin, Sliedniev Lev, Gitis Iryna</i> STUDY OF THE EFFICIENCY OF MACHINE LEARNING METHODS FOR SOLVING HOTEL BUSINESS PROBLEMS	160
<i>Рощенко Станіслав Ігорович, Міхеєнко Денис Юрійович</i> ПЕРЕТВОРЕННЯ МОВЛЕННСЬКОГО СИГНАЛУ В ТЕКСТОВИЙ ПОТІК У РЕАЛЬНОМУ ЧАСІ З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖ	163
Любченко К.М. ЗАДАЧА ПОШУКУ МІНІМАЛЬНОГО (МАКСИМАЛЬНОГО) АРГУМЕНТУ ФАКТІВ БАЗИ ДАНИХ У МОВІ PROLOG	165

Секція 8. Проблеми підготовки фахівців у галузі автоматизації та інформаційних технологій..... 168

Онищенко І. В., Бурма Т. О., Пархоменко В. Ю. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ЯК ЗАСІБ ПІДВИЩЕННЯ ЯКОСТІ ОСВІТНЬОГО ПРОЦЕСУ В УМОВАХ ДИСТАНЦІЙНОГО НАВЧАННЯ..... 169

Рижов О.А., Строїтелева Н.І., Пишнограєв Ю.М. МЕТОДИКА ВИКЛАДАННЯ ДИСЦИПЛІНИ «МЕДИЧНА ІНФОРМАТИКА» ДЛЯ СТУДЕНТІВ СПЕЦІАЛЬНОСТІ «ЛАБОРАТОРНА ДІАГНОСТИКА»..... 171

Кашина Г. С., Писаренко Н.В. ІНФОРМАЦІЙНО-ТЕХНОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ ФОРМУВАННЯ ПРОФЕСІЙНОЇ КОМПЕТЕНТНОСТІ МАЙБУТНІХ ФАХІВЦІВ ТУРИСТИЧНОЇ ГАЛУЗІ..... 173

Ковальчук М.В, Усата О.Ю. ПРОБЛЕМИ ПІДГОТОВКИ ФАХІВЦІВ У СФЕРІ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ..... 176

Нечволода Людмила Володимирівна, Макаров Станіслав Ігорович ОЦІНКА НАВЧАЛЬНИХ КУРСІВ ЗА ДОПОМОГОЮ МЕТОДА ДІАГНОСТИКИ КОМПЕТЕНТНОСТЕЙ..... 178

Нечволода Людмила Володимирівна, Мальцева Тетяна Миколаївна МЕТОДИЧНІ ПІДХОДИ ДО ОЦІНКИ РІВНЯ ЗДАТНОСТІ ДО ГРИ У ШАХИ 181

Сафонов Денис Олегович ВИКОРИСТАННЯ GOOGLE SITES ДЛЯ СТВОРЕННЯ ВЛАСНОГО САЙТУ ВЧИТЕЛЯ ІНФОРМАТИКИ..... 183

А. С. Лисун ПІДВИЩЕННЯ МОТИВАЦІЇ УЧНІВ ДО ВИВЧЕННЯ ІНФОРМАТИКИ ЗАСОБАМИ ГЕЙМІФІКАЦІЇ 185

Лєдок М. В. ВИКОРИСТАННЯ ХМАРО ОРІЄНТОВАНИХ ТЕХНОЛОГІЙ В СИСТЕМІ ПІДГОТОВКИ МАЙБУТНІХ УЧИТЕЛІВ 187