

НТУУ «КПІ ім. Ігоря Сікорського»

Кафедра «Автоматизації проектування енергетичних процесів та систем»

# КОНСОЛІДАЦІЯ ДАНИХ З ВИКОРИСТАННЯМ СТОХАСТИЧНИХ МЕТОДІВ ОПТИМІЗАЦІЇ

Хомицький В.С. студент, гр. ТР-61м

Кузьмініх В.О. к.т.н., доцент

# Актуальність

- Кількість інформації що генерує людина зростає із року в рік в геометричній прогресії, що призводить до проблем зберігання та пошуку необхідних даних. Пропорційно обсягам інформації зростає складність задачі обробки. Без використання спеціальних підходів та інструментів керування даними користування сучасними джерелами інформації стає практично неможливим.

# Наукова новизна

Найбільш суттєвими науковими результатами магістерської дисертації є:

- Удосконалення способу проведення консолідації даних та пошуку інформації;
- Застосування методів консолідації для організації пошуку наукових статей з відкритих джерел.

## Мета

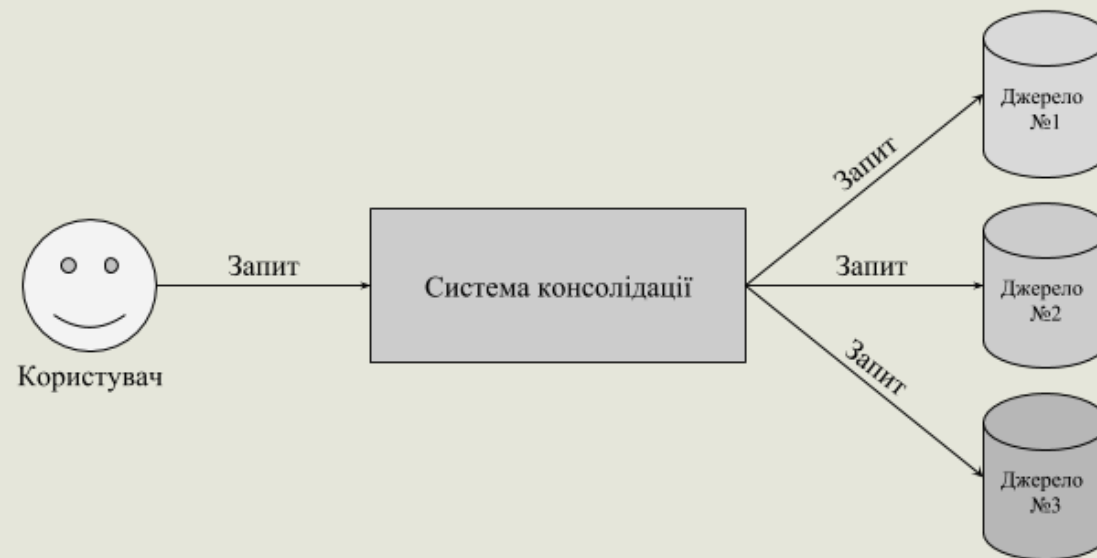
- Створення засобів та виявлення нових підходів в проведені операції консолідації даних.

# Постановка задачі

- Проаналізувати існуючі системи консолідації даних;
- Ознайомитися з існуючими алгоритмами консолідації даних та методами оптимізації;
- Розробити структуру бази даних яка буде оптимально зберігати необхідну інформацію для проведення розрахунків;
- Розробити тестове середовище для відладки алгоритмів;
- Розробити програмне забезпечення для проведення консолідації даних використовуючи тестове середовище та реальні джерела інформації;

# Консолідація даних

Консолідація даних – процес пошуку, відбору, структурування, перетворення, зберігання та надання споживачу інформації по заданим темам.



# Інтеграція з джерелами

- Бібліотека НТУУ «КПІ»

Для інтеграції з бібліотекою НТУУ «КПІ» було використано підключення до бази даних системи ALEPH у форматі MARC XML, що являється всесвітньо відомим і розповсюдженим форматом машиночитного каталогізаційного запису.

- Видавнича служба УРАН

Для інтеграції з видавничою службою УРАН було використано підключення до прикладного програмного інтерфейсу системи Journals, що використовує протокол OAI (Open Archives Initiative).

# Програмні засоби



Java™



hazelcast



elasticsearch



VERT.X



## Пропорційний метод

$$ratio_n = \frac{found_n}{requested_n}$$

$$totalRatio = ratio_1 + ratio_2 + \dots + ratio_n$$

$$k_n = \frac{ratio_n}{totalRatio}$$

# Скінчений та повний автомат

- Скінчений автомат пам'яті – вид абстрактного автомату, кількість можливих внутрішніх станів якого обмежений.
- Повний автомат пам'яті – вид абстрактного автомату, кількість можливих внутрішніх станів якого є необмеженим

# Тестове середовище

Містить три джерела, з різним наповненням та однаковою кількістю документів, відсоток наповнення вказує яка кількість документів містить слово «interesting», яке буде використовуватись для тесту.

Назва джерела	Scopus	RINZ	КПІ
Наповнення	10%	5%	2%
Кількість документів	10 000	10 000	10 000

## Результат порівняння методів

Метод	Рівномірний	Пропорційний	Пропорційний + Скінченний автомат	Пропорційний + Повний автомат
Найкращій результат	2.5%	4.4%	7.6%	8.1%

Найкращій результат становить 8.1% який надає пропорційний метод з використанням повного автомату, це становить 324% результату у порівнянні зі звичайним рівномірним пошуком коли кількість документів що переглядається ділиться порівну між усіма джерелами.

# Створення пошукового запиту

## Консолідація

### Пошуковий запит

Назва

Текст

Автор

Ключові слова

Починаючи

Закінчуючи

Ліміт

Ітерації

Пошук

Очистити

Статистика

# Відображення результатів пошуку

## Консолідація

[< Назад](#)

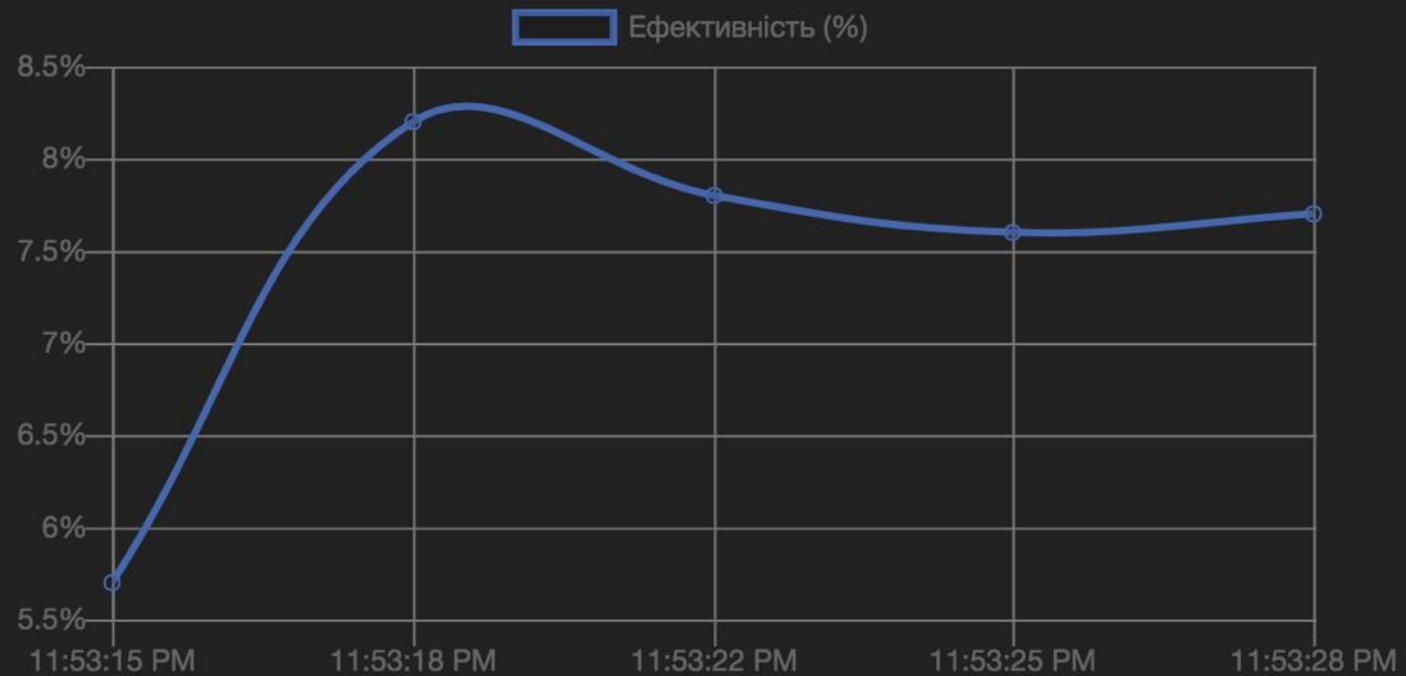
### Результати пошуку

Всього: [41 / 1000](#)      Scopus: [33 / 333](#)      РИНЦ: [0 / 333](#)      КПИ: [8 / 333](#)

#	Назва	Автор	Джерело	Дата	Ключові слова
32	No Highway	Alisha Barton	Scopus	03/11/2010	assumenda,rum,autem
45	That Good Night	Stevie Farrell	Scopus	27/08/1999	quos,et,quis,dolorem,voluptas
53	An Evil Cradling	Christophe Morar	КПИ	29/05/2011	maxime,rem,maiores,et
55	Blithe Spirit	Johnson Schroeder	Scopus	29/11/2010	necessitatibus,omnis,autem
61	A Summer Bird-Cage	Miss Flo Graham	Scopus	13/02/2013	qui,a,reiciendis
64	The Last Enemy	Peyton Bartoletti	Scopus	18/06/1993	inventore,distinctio,ut
76	The Skull Beneath the Skin	Ms. Arthur MacGyver	Scopus	28/03/2002	aut,tenetur,ad,quasi,explicabo
77	The Daffodil Sky	Antonia Kris	Scopus	29/11/2012	quod,qui,reprehenderit
80	Noli Me Tangere	Sage Mraz	Scopus	19/01/2006	eum,sapiente,aut,cumque,dicta
93	The Moving Finger	Angel Boyle	Scopus	12/11/2016	aut,et,nisi
98	Behold the Man	Jaime Hoeger	Scopus	28/08/2002	ullam,pariatur,deserunt,quos
104	I Sing the Body Electric	Lesley Murray	Scopus	07/03/2008	et,ut,fugiat,et
105	Carrion Comfort	Clinton Schroeder III	Scopus	21/04/2004	sed,pariatur,blanditiis,veritatis
111	Time of our Darkness	Wallace Reichert	Scopus	06/07/1989	dolorem,aperiam,corporis,reprehenderit

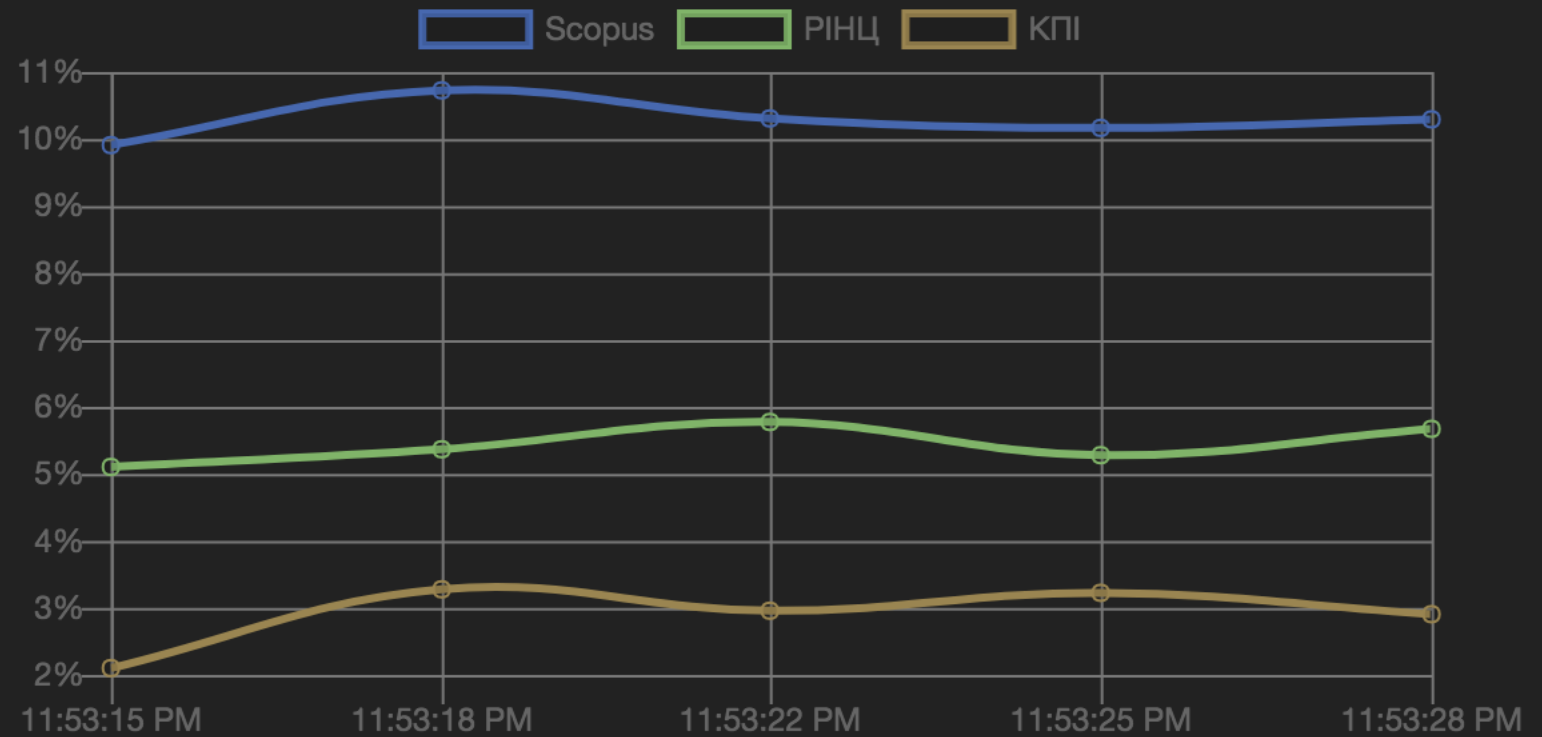
# Відображення статистики

**Ефективність  
Загальна**



# Відображення статистики

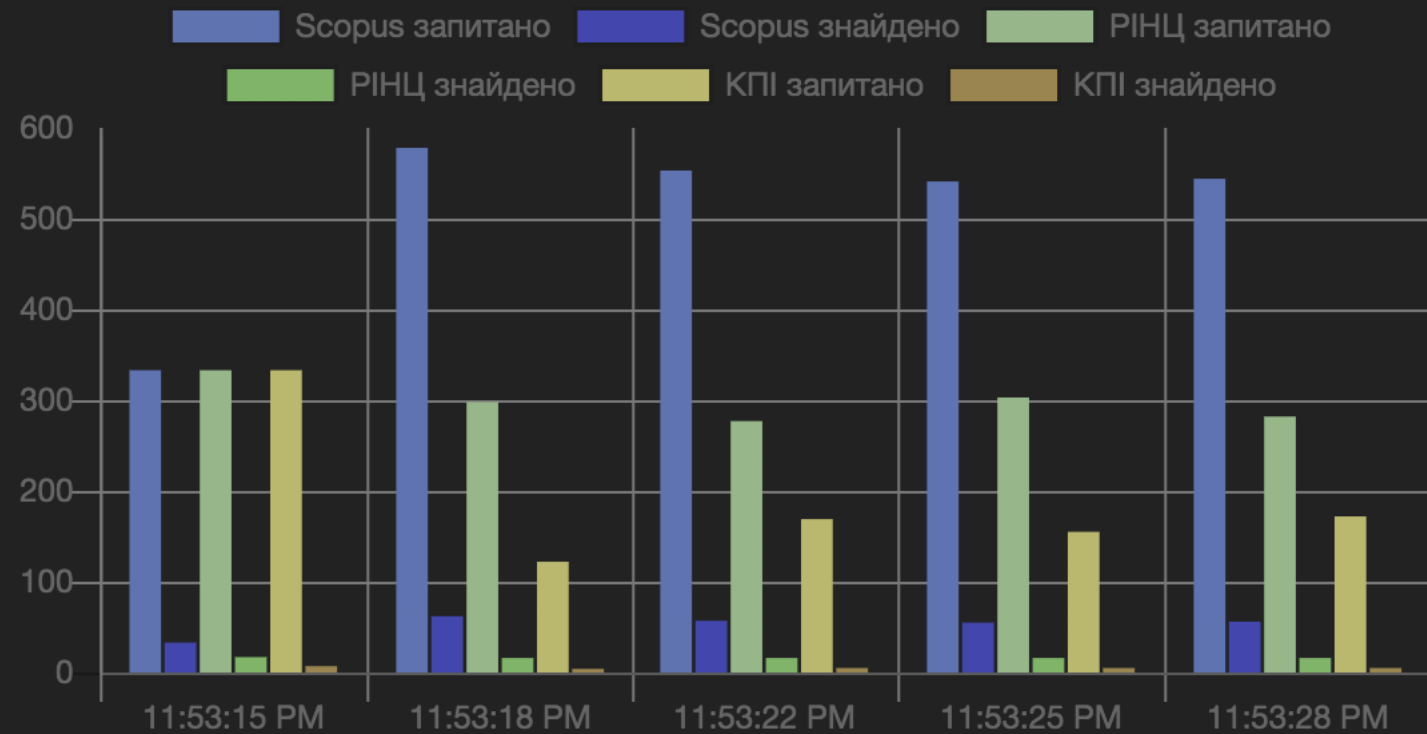
**Ефективність  
Джерел**





# Відображення статистики

## Джерела



# Висновки

- Проаналізовано існуючі системи консолідації даних;
- Дослідження алгоритми консолідації;
- Розроблено та реалізовано модель бази даних пошуку та досвіду;
- Розроблено тестове середовище для відладки алгоритмів консолідації;
- Розроблено систему консолідації з підключенням до тестового середовища та реальних джерел.

Дякую за увагу!