

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

Харківський національний університет радіоелектроніки
Кафедра ЕОМ

**ТЕХНОЛОГІЯ СТВОРЕННЯ КРОСПЛАТФОРМНОЇ
ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ПЕРСОНАЛЬНОЇ
БАГАТОФУНКЦІОНАЛЬНОЇ ЕЛЕКТРОННОЇ БІБЛІОТЕКИ**

Кваліфікаційна робота
Другий (магістерський) рівень

Автор:
Скоромний О.Г.,
студ. гр. СПМ-20-1

Керівник:
Носик А.М.
доц. кафедри ЕОМ

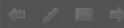
Актуальність теми

В наш час майже всі книги переведені до електронного виду. Для організації та використання цих документів створюються електронні бібліотеки.

Одною з проблем при використанні сучасних електронних бібліотек є пошук документів за фразою або реченням, що містяться в тексті документу, а не в мета інформації.

Мета роботи

Метою атестаційної роботи є аналіз технологій створення багатоплатформової інформаційної системи для персональної багатофункціональної електронної бібліотеки та розробка власної інформаційної системи для персональної електронної бібліотеки з доповненням методу повнотекстового пошуку.



3

Пошукові системи

Пошукова система - алгоритми і реалізуюча їх сукупність комп'ютерних програм, що надає користувачеві можливість швидкого доступу до необхідної йому інформації за допомогою пошуку в великій колекції доступних даних.



Xapian



Sphinx



4

Пошуковий індекс

Пошуковий індекс - структура даних, яка містить інформацію про документи і використовується в пошукових системах.

Індексування, що здійснюється пошуковою машиною, - процес збору, сортування та зберігання даних з метою забезпечити швидкий і точний пошук інформації.

Створення індексу включає міждисциплінарні поняття з лінгвістики, когнітивної психології, математики, інформатики та фізики.



5

Інвертований індекс

Інвертований індекс - структура даних, в якій для кожного слова колекції документів у відповідному списку перераховані всі документи в колекції - в яких воно зустрілося.

Багато пошукових систем використовують інвертований індекс при оцінці пошукового запиту, щоб швидко визначити місце розташування документів, що містять слова із запиту, а потім рангувати ці документи за релевантністю.



6

Порівняння систем повнотекстового пошуку

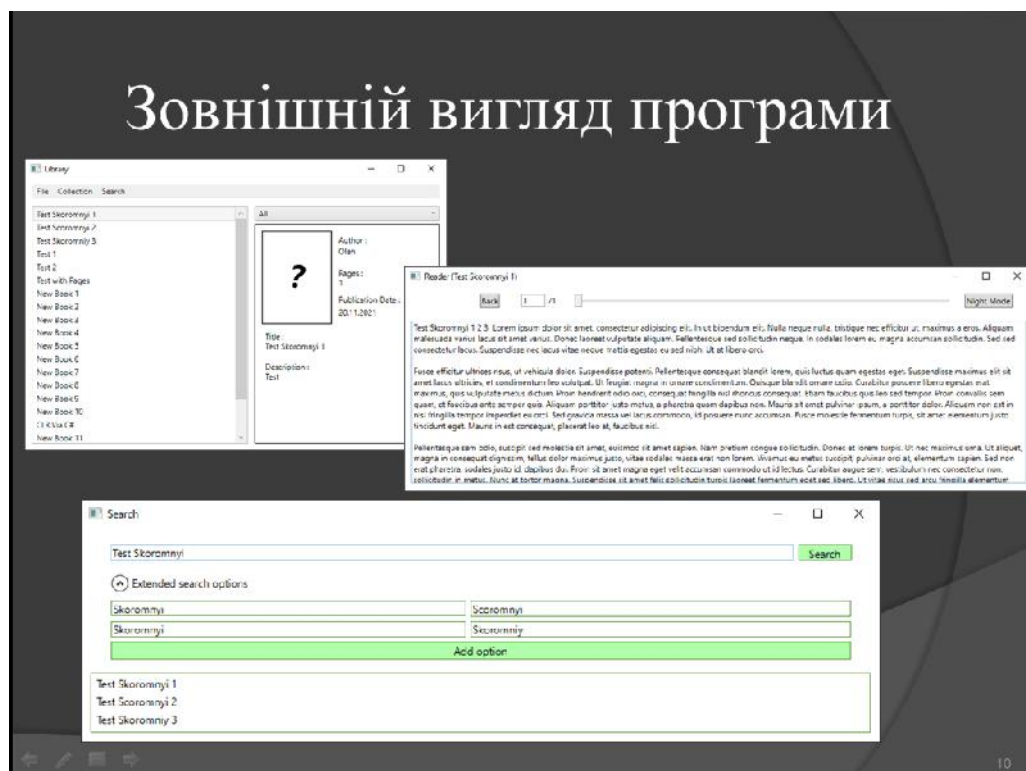
	MySQL	PostgreSQL	Харіан	Sphinx	Solr	CLucene
Швидкість індексації	314 КБ/с	522 КБ/с	1.36 МБ/с	4.5 МБ/с	2.75 МБ/с	3.8 МБ/с
Швидкість пошуку	175 мс / 3.46 сек	28 мс / 2.1 сек	14 мс / 135 мс	7 мс / 75 мс	25 мс / 212 мс	10 мс / 212 мс
Розмір індексу	150%	150%	200%	30%	20%	20%
Реалізація	СУБД	СУБД	Бібліотека	Сервер	Сервер	Бібліотека
Кількість інтерфейсів	4	4	9	6-4	8	4
Оператори пошуку	&*"	&*	&*N-~	&*N-<Z	&*N-~	&*N-~
Стемери	-	15	15	15	31	15
Стоп-слова, синоніми	-	+	+	+	+	+
Soundex	-	-	-	+	+	+

Оператори пошуку

- & - булеві операції
- * - префіксний пошук
- " - точна фраза
- N - слова поблизу
- - діапазони
- ~ - приблизний пошук
- < - порядок слів
- Z - зони

Вибір програмного забезпечення

- ❖ У якості платформи було обрано .Net Core через його кросплатформенність.
- ❖ У якості мови програмування було обрано C#, через простоту та швидкість написання програм.
- ❖ У якості середовища розробки було обрано Visual Studio, через його вбудовані функції, що спрощують розробку.
- ❖ У якості інструменту створення графічного інтерфейсу було обрано WPF, через його великий набір можливостей та сумісність з Xamarin, що використовується у розробці мобільних інтерфейсів.
- ❖ У якості пошукового двигуна було обрано Sphinx через його велику швидкість індексації та малий час пошуку.



Особливості програмного продукту

- ❖ Наявність повнотекстового пошуку з використанням синонімів та транслітерацій
- ❖ Підвищена швидкодія методу пошуку при використанні додаткових опцій
- ❖ Наявність вбудованої функція перегляду документів
- ❖ Можливість додавання електронних документів до колекцій

11

Висновки

За результатами проведеного аналізу та досліджень було створено спеціалізований програмний продукт, що призначений для індивідуального наповнення, організації власного користування персональною електронною бібліотекою читачами. Створена персональна бібліотека має особисте заповнення документами, зручний інтерфейс та потужний пошук.

Наукова новизна результатів роботи полягає в модифікації методу пошуку в інформаційних системах персональних електронних бібліотек при використанні синонімів та транслітерацій за рахунок додання проміжного рівня між користувачем та пошуковою системою.

Практичне значення отриманих результатів полягає в запропонованих технологічних рішеннях удосконалення методів пошуку та процесів розробки інформаційних систем, які можуть бути використані розробниками програмного забезпечення при розробці власних рішень подібного спрямування.

12