

Одеський національний університет імені І. І. Мечникова

Інститут математики, економіки і механіки

Кафедра методів математичної фізики

Дипломна робота

магістра

на тему: «Нечіткий пошук в базах даних»

«Fuzzy search in the databases»

Виконала: студентка денної форми навчання
спеціальності 113 Прикладна математика

Триморуш Ірина Валентинівна

Керівник: к. ф.-м. н., доц. Шуміхін С. А.

Рецензент: к. ф.-м., доц. Процеров Ю. С.

Рекомендовано до захисту:

Протокол засідання кафедри

№ ____ від «____» _____ р.

Завідувач кафедри

Захищено на засіданні ЕК № _____

Протокол № ____ від «____» ____ р.

Оцінка _____ / _____ / _____

Голова ЕК

Одеса — 2017 р.

Зміст

Вступ	3
1 Нечіткі множини	4
1.1 Основні поняття теорії нечітких множин	4
1.2 Функції приналежності і методи їх побудови	7
1.3 Розширення арифметичних операцій над нечіткими даними	21
1.3.1 Відношення еквівалентності	21
1.3.2 Відношення порядку	22
2 Нечіткі дерева пошуку	26
2.1 Нечіткі двійкові дерева пошуку	26
2.1.1 Основні операції дерев пошуку	27
2.2 Нечітка база даних	29
2.3 Нечіткий пошук у базах даних, нечіткі індекси	32
2.3.1 Побудова нечіткого бінарного дерева з заданим α -рівнем	32
2.3.2 Нечіткий пошук з використанням нечіткого індексу .	34
2.3.3 Зсув функції приналежності	35
2.3.4 Агрегація результатів	40
3 Практична частина	43
3.1 Технічна документація	43
3.2 Практичне застосування - нечіткий пошук	62
Висновки	73
Список літератури	74
Додатки	76
3.3 Додаток А. Таблиця вихідних даних	76
3.4 Додаток Б. Код програми	77

Вступ

Проблема обробки великих обсягів даних супроводжує людство з давніх часів. В останні десятиліття ця проблема набула дещо інший характер. Для зберігання великих обсягів даних з'явився великий вибір різноманітних засобів, в тому числі і сучасні бази даних, які можуть зберігати і обробляти терабайти інформації. Раніше людині вистачало стандартного функціоналу баз даних, який представляє з себе чітку обробку даних. Під чіткою обробкою даних розуміємо свідомо чітко обмежені і певні операції над даними. Наприклад, запити пошуку: "Знайти щось в такій кількості за конкретний проміжок часу". Подібні запити сильно обмежують і обрізають вихідні дані.

Сьогодні все частіше виникає проблема пошуку не тільки чогось конкретного, а й того, що ПОДІБНО того, що шукаємо. Ідея шукати ПОДІБНЕ спричинила розвиток НЕЧІТКИХ баз даних, з нечітким пошуком над чіткими і нечіткими даними. Величезний недолік нечітких баз даних полягає в занадто трудомісткому нечіткому пошуку через пошук ПОДІБНИХ.

Метою даної роботи є знайти спосіб об'єднання продуктивності чіткого пошуку та ефективності нечіткого пошуку, в сенсі пошуку ширшого спектра даних.

Висновки

Основною метою даної роботи є побудова швидкого нечіткого пошуку серед нечітких чи стандартних даних. Для нечітких даних характерна процедура кластеризації, яка допомагає класифікувати набір різноманітних даних за деякими критеріями по кластерам - множинам. Це звичайно допомагає в пошуку цілої категорії даних, але що робити, якщо серед критеріїв кластерів не має того, що шукаємо. В такому випадку потрібно буде перекласифікувати дані. Ідеєю цієї роботи є рішення таких задач, яка полягає в побудові певної упорядкованої структури даних. В цій роботі були розглянуті побудова і пошук по бінарному дереву, вузлами якого є кластери. Це звичайно пришвидчить пошук даних по кластерам, так як самі кластери вже впорядковані якимось заданим чином. А також дозволить шукати не тільки самі кластери, а й дані, які перетинаються у декількох кластерах, або ж, навпаки, тільки частину кластеру. Можна запевнити, що пошук по нечіткому бінарному дереву швидкий, так як в основі залишається структура стандартного бінарного дерева, а складність алгоритму пошуку в середньому $O(\log n)$.

Список літератури

1. Лиу Дик Кхам. Деякі розширення в теорії нечітких реляційних баз даних / Лиу Дик Кхам. — Санкт- Петербург.: Наука, 2003. — 116 с.
2. Ібрагімов В. А. Елементи нечіткої математики / В. А. Ібрагімов. — Баку.: Наука, 2010. — 12 с.
3. Заде Л. А. Fuzzy sets / Л. А. Заде. — Information and Control, 1965. — 68 с.
4. Заде Л. А. Поняття лінгвістичної змінної і його застосування до прийняття наближених рішень. / Л. А. Заде. — М.: Мир, 1976. — 46 с.