

Параллельный алгоритм фрактального поиска в базе данных*

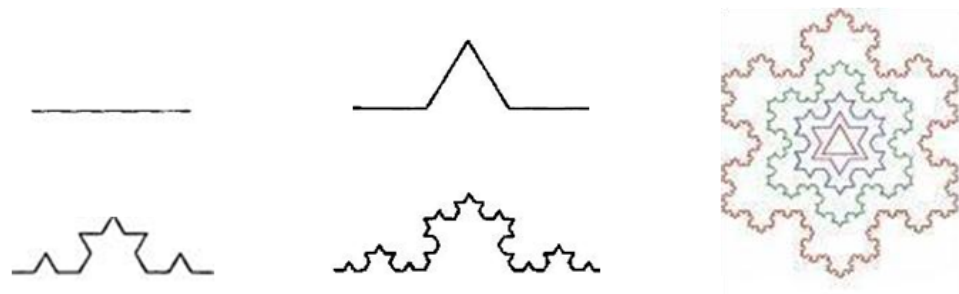
Т.Ю. Лымарь, Н.Ю. Староверова

tatyana.lymar@gmail.com, snu1988@yandex.ru

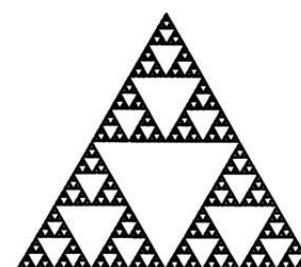
Южно-Уральский государственный университет

Фрактал - структура, состоящая из частей (**доменов**), которые в каком-то смысле подобны целому. Понятие введено Бенуа Мандельбротом в 1975 году.

Кривая (снежинка) Коха



Ковер Серпинского



Самоподобие в реляционных таблицах

A1	A2	A3	A4	A5

Diagram showing a table with 5 columns (A1-A5) and 5 rows. Arrows point from the table to a smaller version of the same table structure below it, illustrating self-similarity.

Имя_Д	Город_Д	Объем_Д	Цвет
Гайка	Тула	5	Крас
Гайка	Тула	10	Син
Болт	Сургут	5	Крас
Болт	Сургут	10	Син
Болт	Сургут	20	Черн

Diagram showing a table with 4 columns (Имя_Д, Город_Д, Объем_Д, Цвет) and 6 rows. Arrows point from the table to two smaller tables below it, illustrating self-similarity. The first smaller table has 2 columns (Гайка, Тула) and 2 rows (Болт, Сургут). The second smaller table has 2 columns (5, Крас) and 3 rows (10, Син, 20, Черн).

Применение теории фракталов к реляционным БД:

1. **Классификация и кластеризация** – автоматическое разбиение отношения на группы кортежей в зависимости от их схожести;
2. **Понижение размерности** - исключение коррелирующих атрибутов;
3. **Сжатие** путем устранения повторяющихся значений доменов.

Параллельный алгоритм выделения доменов

1. На узле P_0 формируется начальное множество D возможных доменов;
2. Элементы множества рассылаются остальным узлам;
3. Каждый узел-получатель пытается соответственно кластеризовать свой фрагмент анализируемой таблицы;
4. В случае неприменимости полученного домена к имеющимся записям, создается новый элемент множества D , для которого процедура рекурсивно повторяется.

Среда реализации – СУБД Oracle

Основа реализации:

словарь СУБД Oracle
+
инструмент сбора статистики

Критерии отбора потенциальных элементов доменов:

1. равное количество различных значений атрибута;
2. накопление информации о совпадающих множествах значений атрибутов.

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 09-07-00241-а).