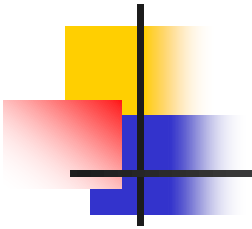
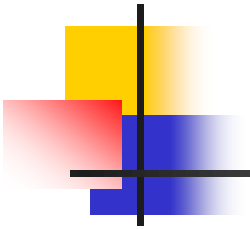


Хранение и анализ хеш-значений файлов в компьютерно-технической экспертизе

Куликов Егор, 344 группа
Математико-механический факультет
Научный руководитель: ст. пр. Губанов Ю.А.
2015

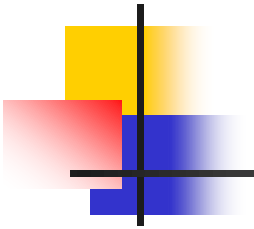


Введение

Компьютерно-техническая экспертиза применяется во всём мире

Исследование файловой системы, отсечение «лишних» файлов

NSRL – крупнейшее хранилище хеш-значений файлов



Постановка задачи

1) Реализовать эффективное хранилище для хеш-значений файлов
2) Реализовать эффективный поиск всех хеш-значений устройства в хранилище
3) Интеграция в Belkasoft Evidence Center

Существующие решения

1) Autopsy

открытый исходный код

применение плоских файлов

2) Internet Evidence Finder

разработка Magnet Forensics

применение SQLite

возможность подключения PhotoDNA

Существующие решения

3) PhotoDNA

исключительно анализ изображений

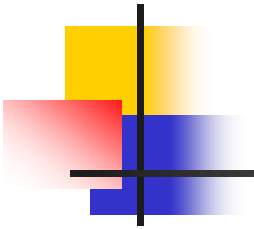
механизм запроса на сервер

применяется NCMEC

4) Foresics Explorer

разработка Get Data

применение собственных баз



Подходы к реализации

Применение реляционных БД

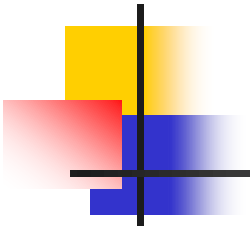
SQLite

скорость сохранения не выше 20000 записей в минуту

MS SQL Server

требует соединения с сервером

до 55 тысяч записей в минуту



Подходы к реализации

Применение NoSQL

Berkeley DB

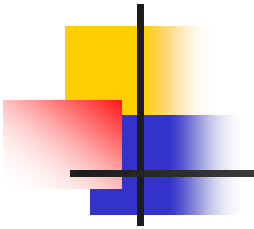
скорость записи в 2 раза ниже, чем у SQLite

наблюдается регрессия скорости

Mongo DB

нет регрессии

но 70000 записей в минуту



Подходы к реализации

Плоские файлы

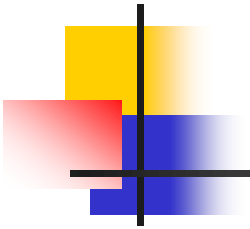
структура в виде двоичного дерева

64 каталога на нижнем уровне

1024 файла в каталоге

достигнута скорость 10^6 записей в минуту

на 420 ГБ тестовых данных



Бинарный поиск

Хорошая асимптотика O(log n)

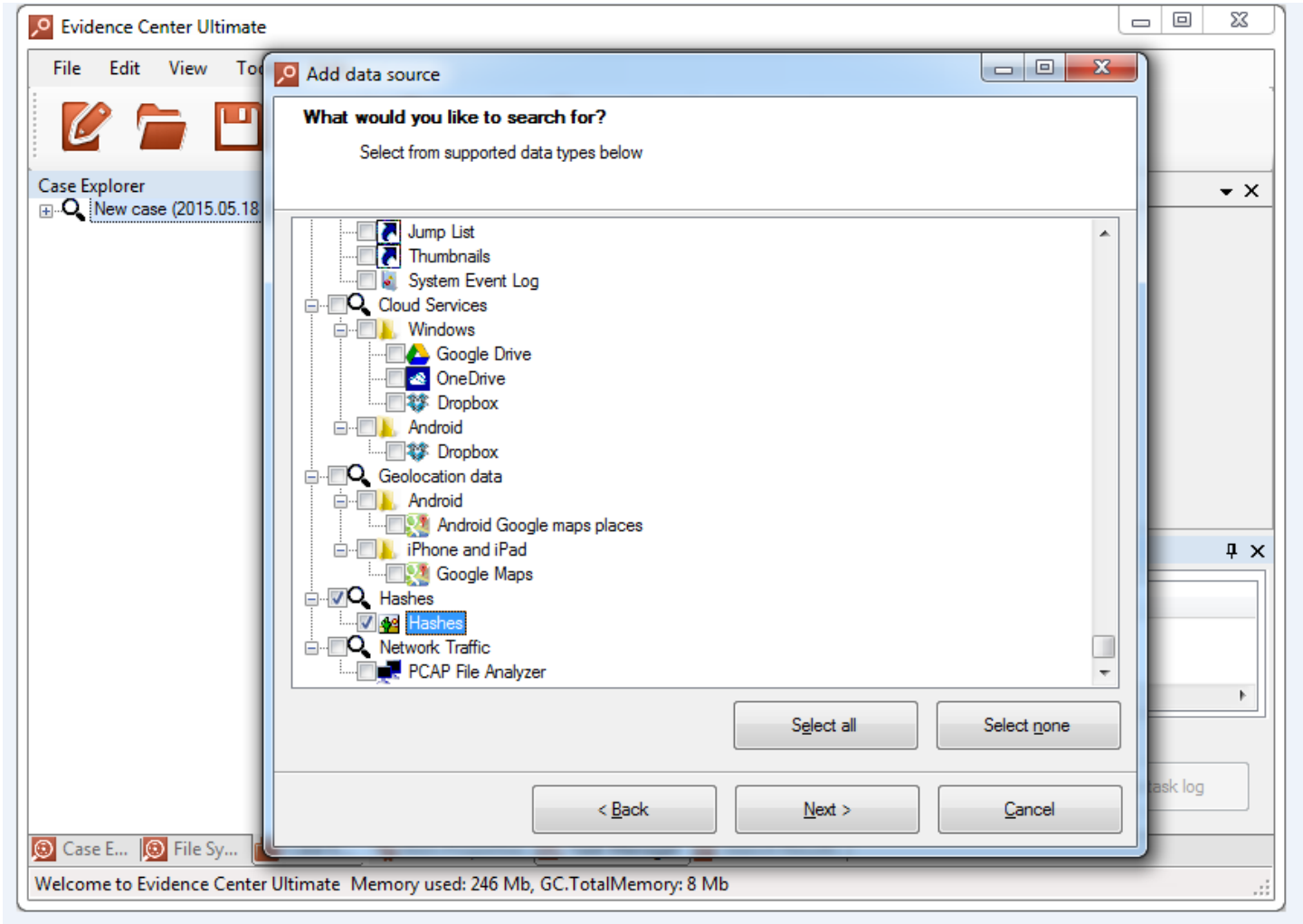
Требует сортировки файлов хранилища

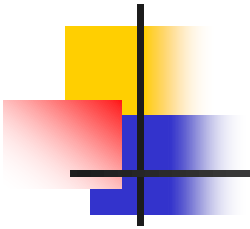
k-way merge sort

Требует специального формата файла

250 символов + пробел на одну запись

Интеграция


Evidence Center Ultimate
File
Edit
View
Case Explorer
New case (2015.05.18
Add data source
What would you like to search for?
Select from supported data types below
Jump List
Thumbnails
System Event Log
Cloud Services
Windows
Google Drive
OneDrive
Dropbox
Android
Dropbox
Geolocation data
Android
Android Google maps places
iPhone and iPad
Google Maps
Hashes
Hashes
Network Traffic
PCAP File Analyzer
Select all
Select none
< Back
Next >
Cancel
Case E...
File Sy...
Welcome to Evidence Center Ultimate Memory used: 246 Mb, GC.TotalMemory: 8 Mb



Заключение

1)Разработано эффективное хранилище для хеш-значений файлов и алгоритм поиска по нему

2)Проведена интеграция в Belkasoft Evidence Center

Возможный путь продолжения работы:

создание аналога PhotoDNA для работы не только с изображениями