

[www.nbitsearch.com](http://www.nbitsearch.com)

ООО “Новосиб-БИТ”

# Индексирование и быстрый **поиск**

ДВИЖОК

***NBITSearch***

характеристики

***версия 1.03.3***



# Система NBITSearch

## *NBITSearch*

это поисковый движок с открытым API.

---

*NBITSearch* это программное ядро для

- Систем Управления Базами Данных,
- Хранилищ Данных Большого Объёма,
- Поисковых Систем Любых Объектов.

# Назначение системы

***Компактное*** индексирование ***больших*** массивов данных на жёстком диске,

организующее

***скоростной*** чёткий и нечёткий поиск объектов при ***минимальных*** затратах оперативной памяти.

# Чёткий и нечёткий поиск

Нечёткий (неточный) поиск реализуется с помощью

**диапазонных запросов.**

Чёткий (точный) поиск является частным случаем нечёткого.

# Индексируемые объекты

**Объекты S**  
**произвольных**  
**типов T**

# Многофункциональность

Система индексирует  
объекты  $S$   
произвольных типов  $T$   
одновременно  
по набору  
любых функций  $F(S)$ .

# Индексируемые объёмы

Наибольший эффект в скорости поиска проявляется на массивах объектов,

число которых для одного индекса  
 **$\approx 50 \div 100$  миллионов и более.**

Размеры массивов индексируемых объектов могут исчисляться  
**десятками и сотнями терабайт.**

# Пределы индексирования

Один индекс поддерживает  
 **$\approx 2$  миллиарда**  
своих объектов.

**Ограничения**  
на число индексов являются  
искусственными.



# Что такое миллиард?

**1 миллиард секунд это  
≈ 32 года.**

**1 миллиард листов  
для лазерного принтера это  
пачка высотой ≈ 100 км.**

# Скорость индексирования

Асимптотика:

$$T \sim (N) * LOG (N)$$

**T** – время создания одного индекса,  
**N** – число индексируемых объектов.

# Компактность индексов

Размер одного индекса  
может варьировать  
в пределах

**0,1 % ÷ 5,0 %**

от размера  
индексируемых объектов.

# Скорость поиска

Оценка времени  
вычисления адреса первого  
потенциального блока данных:

$$T \sim \text{LOG}(N)$$

**T** – время “логического зондирования”,  
**N** – число объектов индекса.

# Скорость поиска

Скорость выборки результатов диапазонных запросов с жёсткого диска может быть

**в 10 ÷ 100 раз выше**

(в большом массиве данных),

чем аналогичная скорость в стандартных реляционных СУБД.

# Скорость поиска

Скорость выборки результатов диапазонных запросов с жёсткого диска может быть

**в 1000 раз (и более) выше**

(в большом массиве данных),

чем аналогичная скорость в решениях полным перебором.

# Скорость поиска

**Время выборки  
результатов диапазонных запросов  
с жёсткого диска**

**линейно зависит**

**от числа объектов, попадающих  
в результирующее множество.**

# Поисковые затраты памяти

**Ввиду  
КОМПАКТНОСТИ ИНДЕКСОВ,  
КАЖДЫЙ ИЗ НИХ  
перед началом запросов  
полностью загружается  
в оперативную память.**



# Поисковые затраты памяти

**Размер  
буферов памяти**

**для выборки данных зависит  
от потребностей пользователя.**

**Часто это мизерный размер,  
исчисляемый десятками мегабайт.**

# Чтение результатов поиска

Чтение  
в оперативную память

результатов поиска  
с жёсткого диска

**происходит оптимально:  
ОСЦИЛЛЯЦИИ МАГНИТНОЙ ГОЛОВКИ НЕТ.**

# *СПАСИБО ЗА ВНИМАНИЕ!*

[www.nbitsearch.com](http://www.nbitsearch.com)

Технология разработана при поддержке [Фонда СР МФП НТС](#),  
образованного Правительством Российской Федерации  
ООО "Новосиб-БИТ" © 2004 - 2011  
Запатентовано