

Ваш надёжный партнёр в сфере ИТ

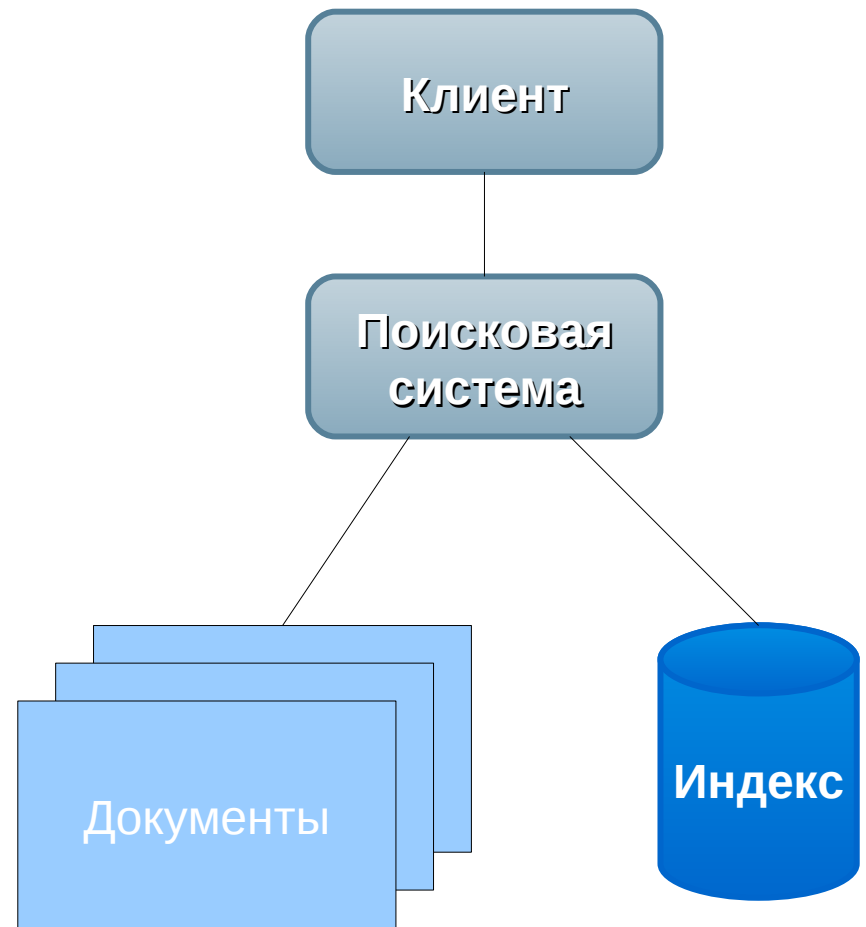
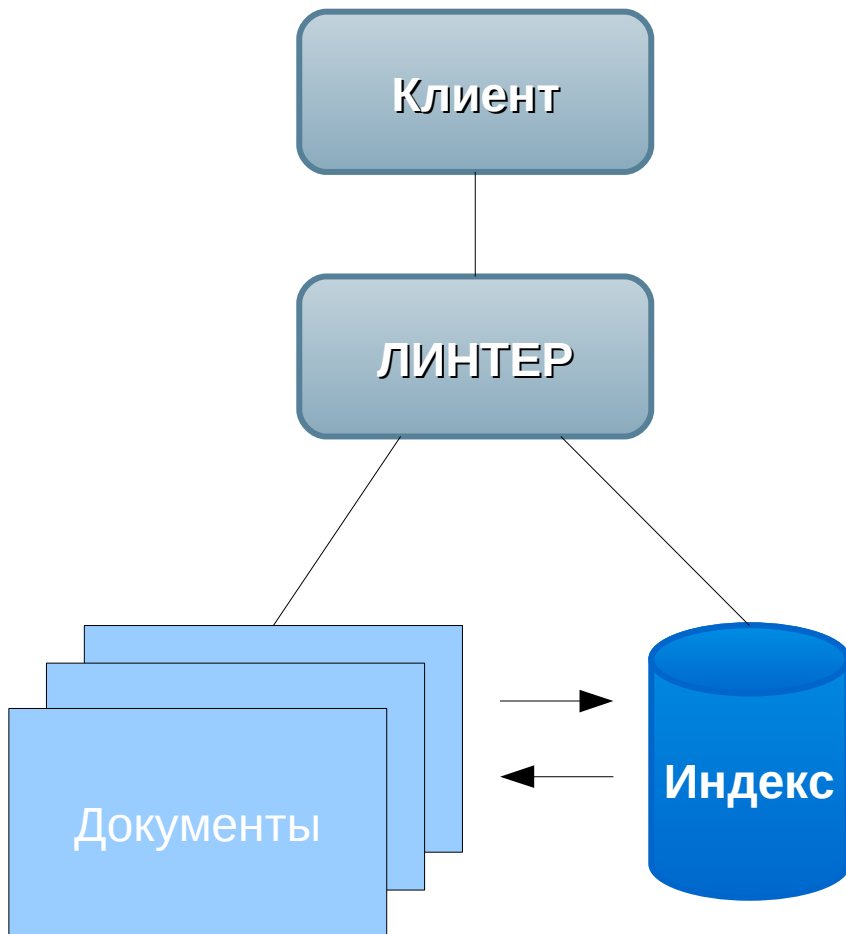


Новые средства
СУБД ЛИНТЕР

Полнотекстовый
ПОИСК

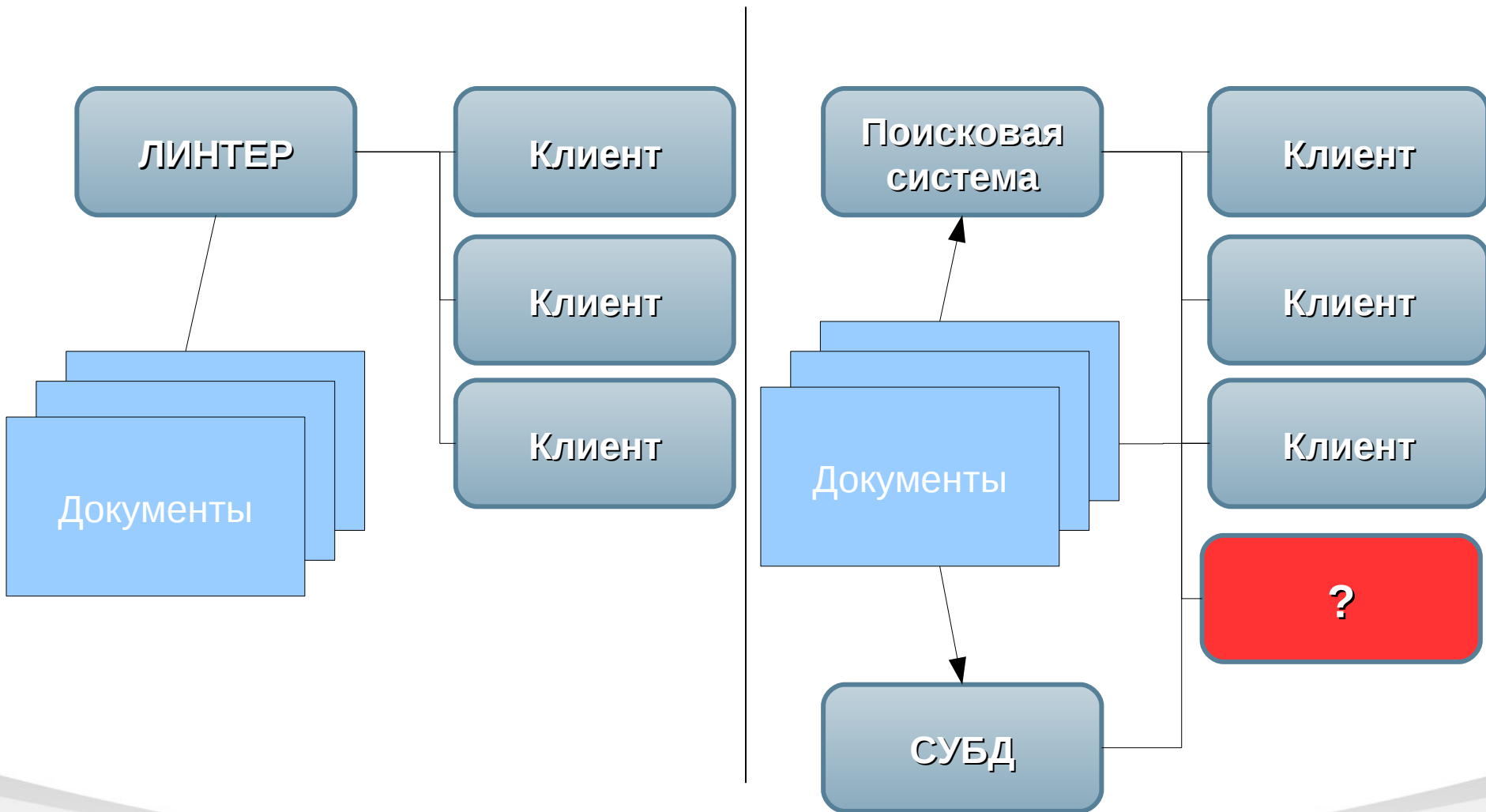
Полнотекстовый поиск

Синхронность



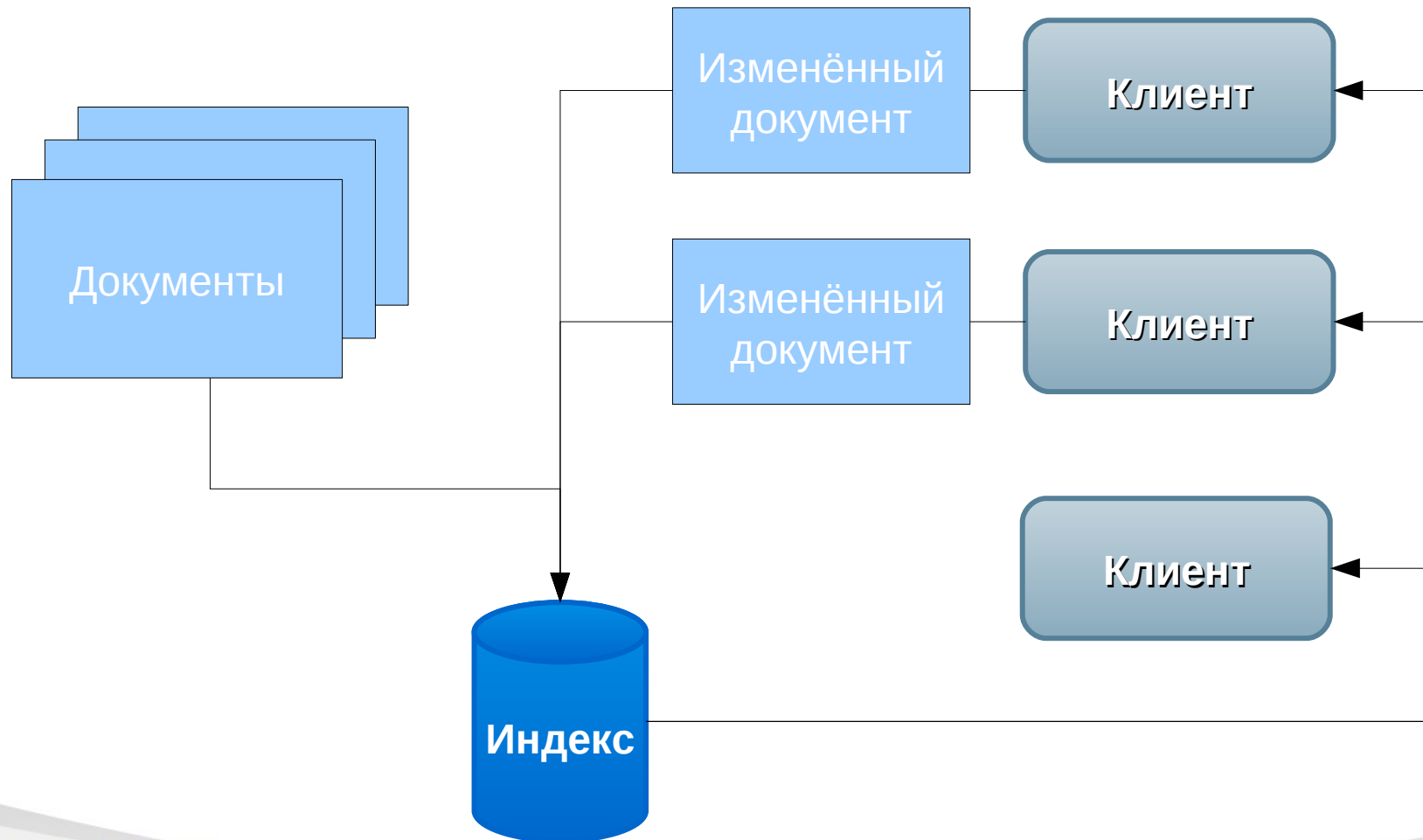
Полнотекстовый поиск

Доступ



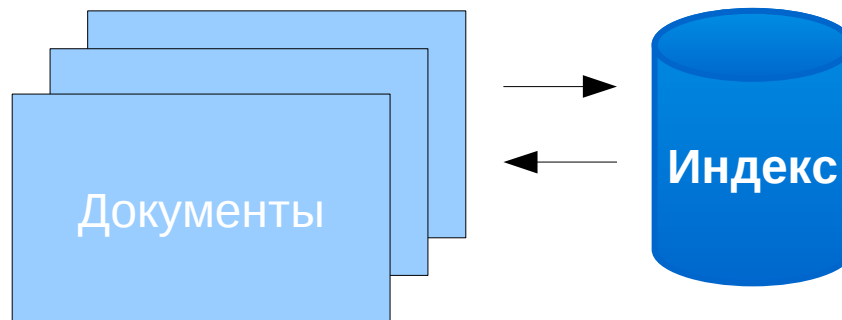
Полнотекстовый поиск

Многоверсионность

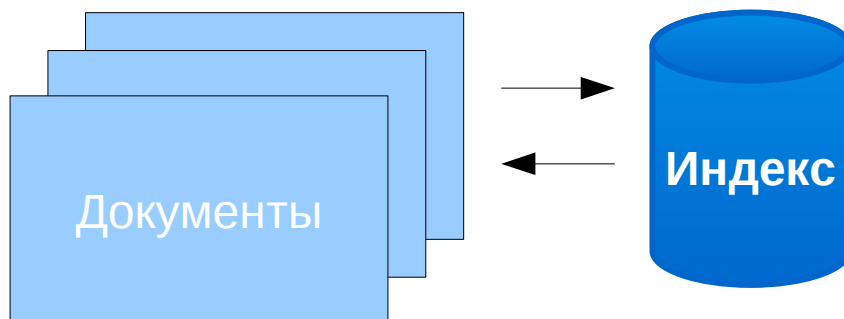


Полнотекстовый поиск

Целостность



Сбои, наводнения, цунами, войны, ...



Полнотекстовый поиск

Использование

- » Система управления проектами Project Tracking
- » Система обработки рекламаций Ticket Tracking
- » ИАС НЕВОД-КРЕДИТ
- » Поисковая система сайта www.linter.ru

Память

- » Постоянный объём используемой памяти
- » Минимальные требования к памяти
(для работы достаточно нескольких десятков килобайт)
- » Возможность настройки
- » Возможность временного отключения подсистемы

Снимки

Снимок состояния индекса позволяет:

- » READ ONLY транзакции
- » Полнотекстовый поиск остаётся доступным во время индексации
- » Быстрое восстановление после сбоя
(не нужна полная индексация)

Полнотекстовый поиск

Расстояние между словами

» Поиск цитаты:

TEXT CONTAINS '"*Errata humanum est*"'

» Поиск близко расположенных слов:

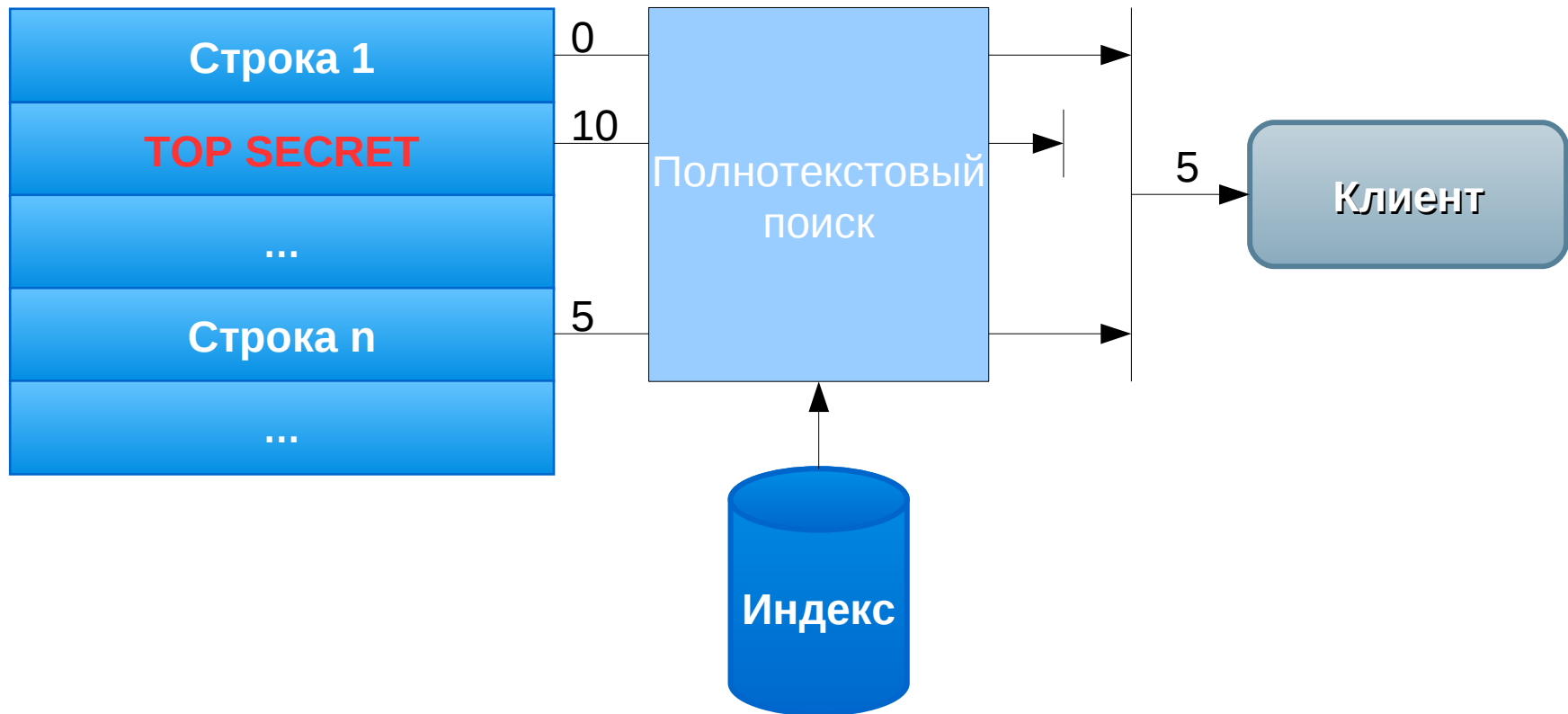
TEXT CONTAINS '"Александр |2| Пушкин"'

Александр Сергеевич Пушкин
Пушкин Александр
...

Полнотекстовый поиск

Мандатная защита

Таблица



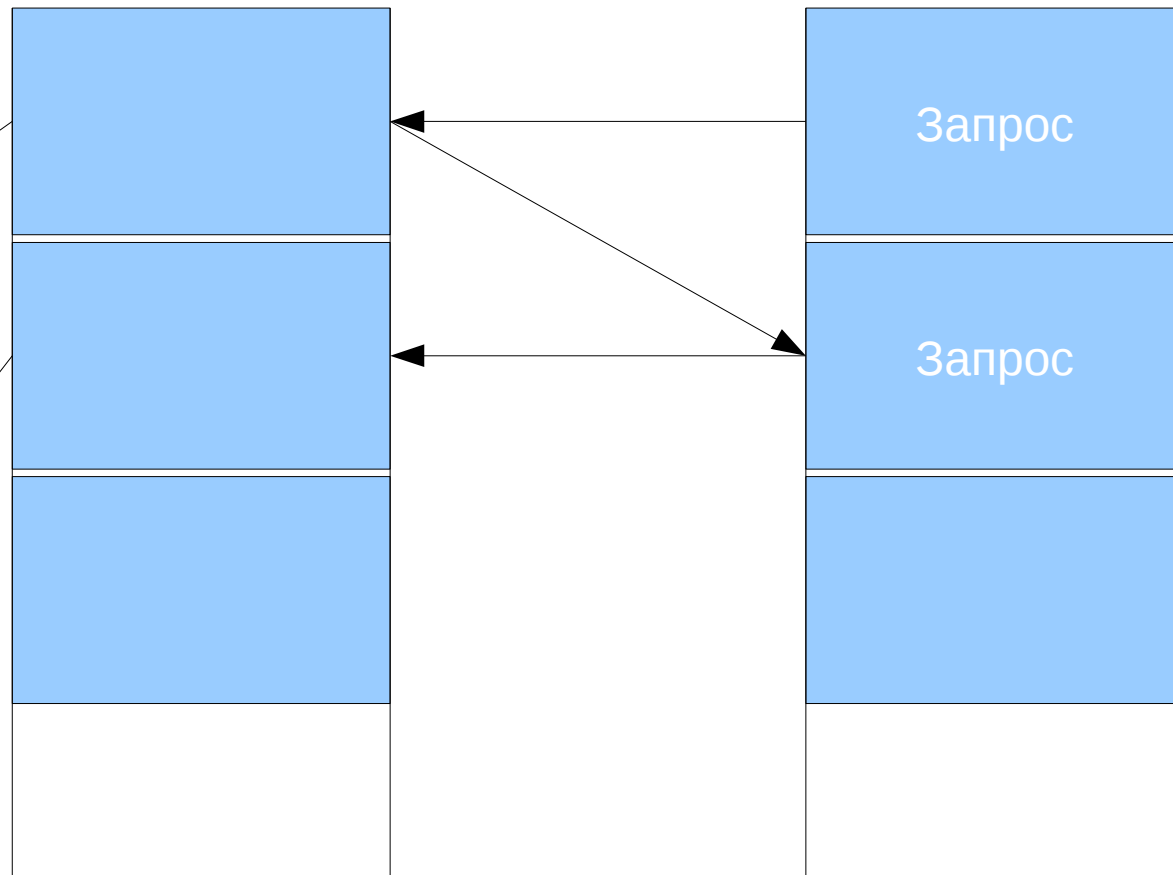
Полнотекстовый поиск

Асинхронность

Устройства

Пул страниц I/O

Очередь запросов



Полнотекстовый поиск

SIMILAR TO vs. CONTAINS

TEXT CONTAINS 'WORD'

TEXT SIMILAR TO '

`[^a-zA-Z][wW][oO][rR][dD][^a-zA-Z] |`

`^[wW][oO][rR][dD][^a-zA-Z] |`

`[^a-zA-Z][wW][oO][rR][dD]$ |`

`^[wW][oO][rR][dD]$'`

Полнотекстовый поиск

Документы

- » Текстовое поле (CHAR, VARCHAR) в однобайтной или многобайтной кодировке, поддерживаемой ядром
- » Текстовое поле (NCHAR, NVARCHAR) в формате UNICODE
- » Поле BLOB
- » Внешний файл (EXTFILE). В БД хранится только ссылка

Фильтры

Фильтр может быть назначен:

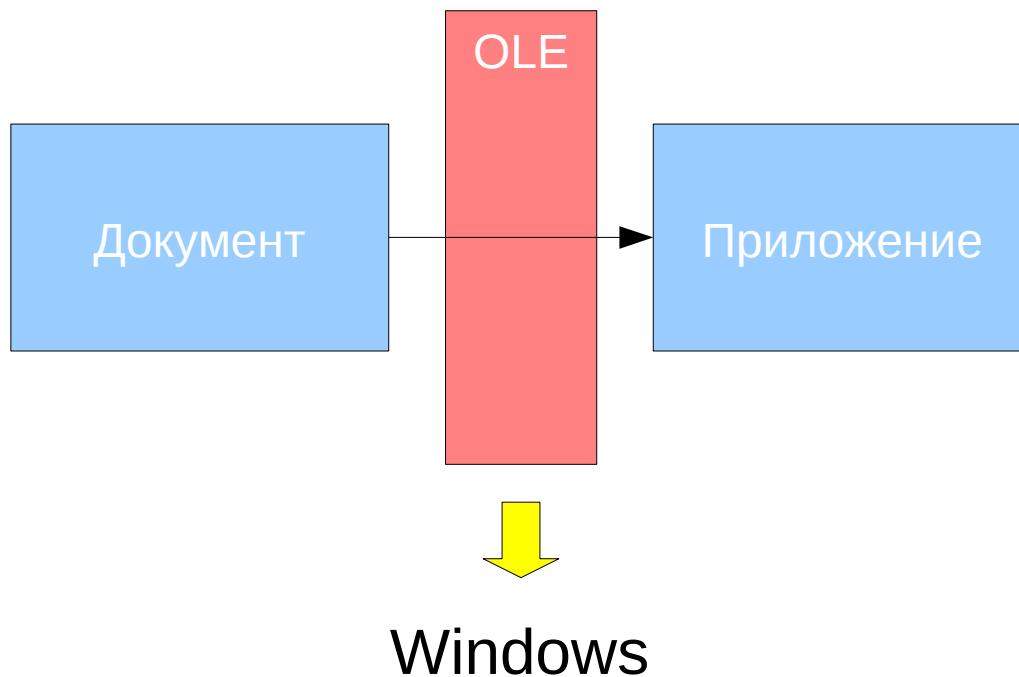
- » Отдельному документу (BLOB, EXTFILE)
- » Целому столбцу
- » На основании расширения (внешнего файла)
- » На основании сигнатуры (Byte Order Mark для UTF-8 и UNICODE, сигнатура MS-CDF для документов MS Office, фиксированный заголовок для PDF и пр.)
- » Документ может быть исключён из индекса
- » Для текстовых документов (BLOB, EXTFILE) возможно автоматическое определение кодировки

Форматы

- » Текст без разметки (одно- и многобайтные кодировки, UTF-8, UNICODE)
- » Форматированный текст (RTF, HTML)
- » Бинарные форматы (PDF, DOC, XLS, PPT)
- » Структурированные данные (XML)

Полнотекстовый поиск

MS Office: только Windows?



СУБД ЛИНТЕР

Linux	QNX
MCBC	USIX
Solaris	VxWorks
Mac OS X	OS-9
BSD	OS-9000
UnixWare	OC2000
IRIX	INTROS
AIX	VMS
SINIX	Windows

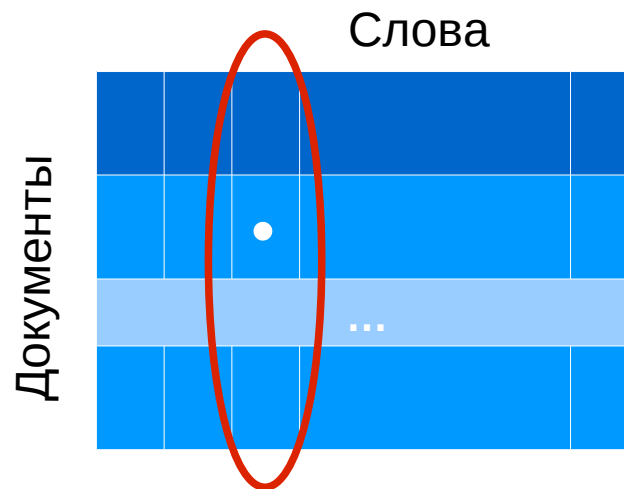
Полнотекстовый поиск

Полнотекстовый индекс

AbC (ID=3)

A/1		•	...
B/2		•	...
C/3		•	...
∅/4		•	...

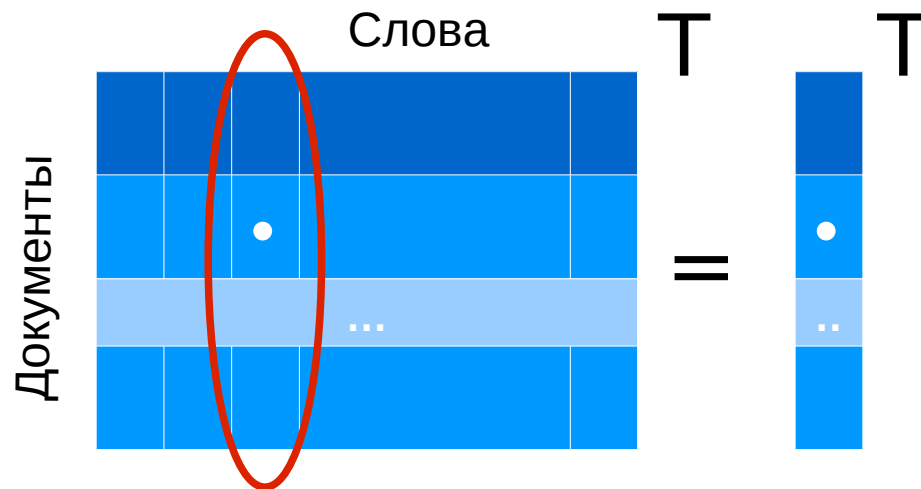
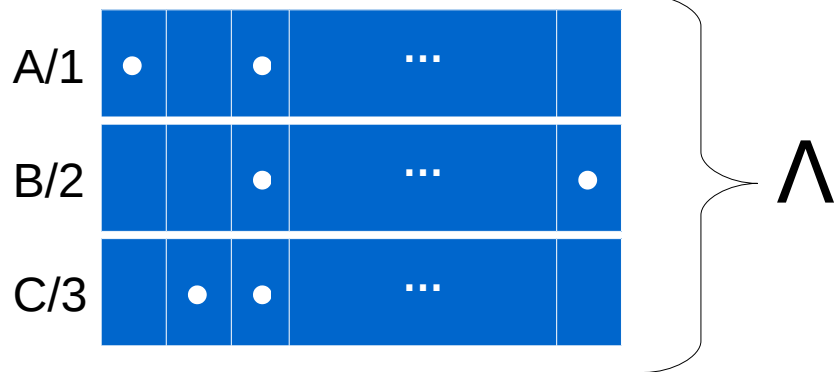
'A'/1		•	...
'b'/2		•	...
'C'/3		•	...



Полнотекстовый поиск

Поиск по маске

TEXT CONTAINS 'ABC*'



Нечёткий поиск

TEXT CONTAINS FUZZY 'ABC'

» Пропущенная буква:

BC, AC, AB

» Лишняя буква:

?ABC, A?BC, AB?C, ABC?

» Неверная буква (опечатка):

?BC, A?B, AB?

Полнотекстовый поиск

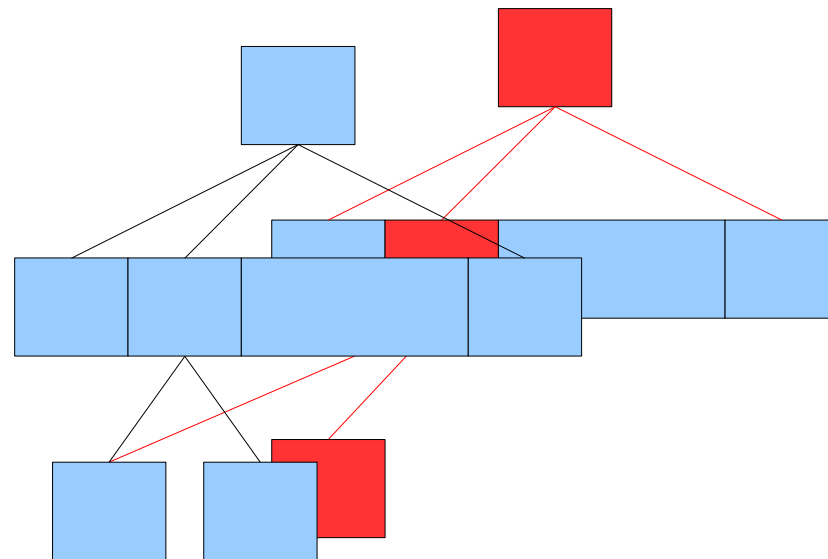
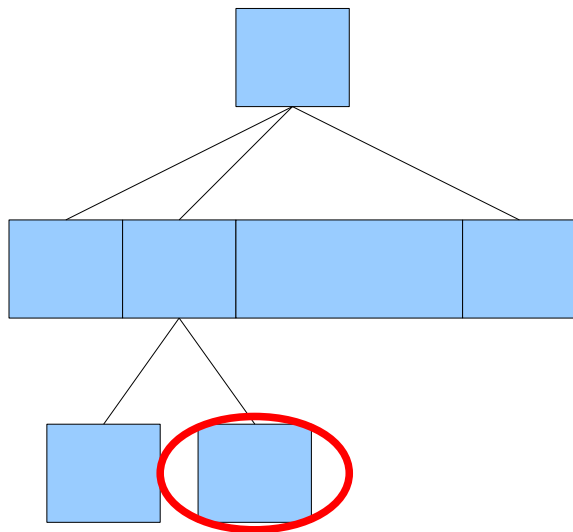
Атрибутный поиск

```
<?xml version="1.0" encoding="utf-8"?>  
<poet name="Александр" surname="Пушкин">  
  <poem title="Медный всадник">  
    На берегу пустынных волн  
    Стоял он, дум великих полн,  
    . . .  
  </poem>  
</poet>
```

ТЕХТ CONTAINS 'title="Медный всадник"'

Полнотекстовый поиск

Модификация индекса



Полнотекстовый поиск

Параллельность



Расширения SQL

- » Функции для работы с внешними файлами (поле EXTFILE)
- » Операторы для работы с фильтрами (создание, назначение и пр.)
- » Функции и операторы для работы с полнотекстовым индексом (INDEXTIME, CREATE PHRASE INDEX, ...)
- » Предикат полнотекстового поиска CONTAINS
- » Извлечение текста документа с помощью парсера (GETTEXT)
- » Определение позиций слов, удовлетворяющих шаблону (GETTEXTPOS)

Полнотекстовый поиск

Перспективы

- » Внешние фильтры
- » Стемминг
- » Параллельные операции в рамках обработки одного поискового запроса

Спасибо за внимание

Вопросы?