



ПОЛНОТЕКСТОВЫЙ ПОИСК
ФАСЕТНЫЙ ПОИСК

SPHINX

ПОЛНОТЕКСТОВЫЙ ПОИСК

Концепция



**Извлечение
ключевых слов**



**Морфология
Фонетика**



**Удобная
структура**

ФАСЕТНЫЙ ПОИСК

Концепция



Фильтры
Значения
Документы



Сужение
результатов
поиска



Отсутствие
пустой
выборки



НЕМНОГО ОБ ИНСТРУМЕНТЕ

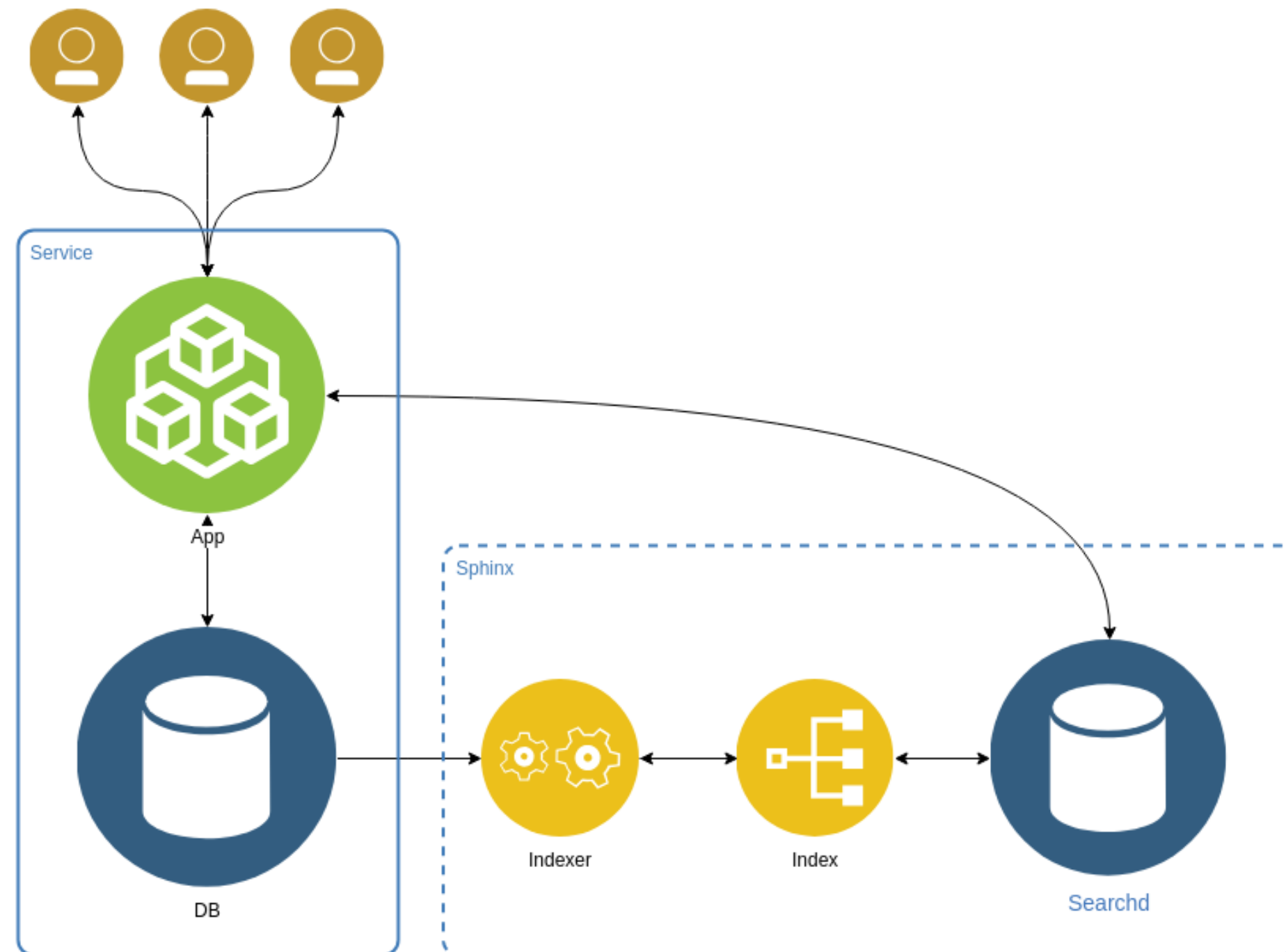
Sphinx это полнотекстовый поисковый сервер с открытым исходным кодом, разработанный с нуля с учетом производительности, качества поиска и простоты интеграции.

Он написан на C++ и работает на Linux, Windows, MacOS, Solaris, FreeBSD и некоторых других системах.



Схема Интеграции

В существующий сервис



Файл Конфигурации

Структура, особенности



Строим индекс

Из героев вселенной Marvel :)

Скачан датасет из 16k героев, немного подправлен и залит в mysql в следующем виде:

```
CREATE TABLE `marvel` (  
  `id` int(11) UNSIGNED NOT NULL,  
  `name` varchar(69) DEFAULT NULL,  
  `character` varchar(18) DEFAULT NULL,  
  `eye` varchar(15) DEFAULT NULL,  
  `hair` varchar(21) DEFAULT NULL,  
  `sex` varchar(22) DEFAULT NULL,  
  `year` year(4) DEFAULT NULL  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```



Файл конфигурации

Вселенной Marvel :)

```
source db
{
    type = mysql

    sql_host = 127.0.0.1
    sql_user = test
    sql_pass = test
    sql_db = test
    sql_port = 3306
}

source marvel:db
{
    sql_query = SELECT * FROM marvel

    sql_field_string = name
    sql_attr_string = character
    sql_attr_string = eye
    sql_attr_string = hair
    sql_attr_string = sex
    sql_attr_uint = year
}

index marvel
{
    source = marvel
    path = /app/data/marvel
    morphology = stem_enru, metaphone
    min_stemming_len = 3
    min_word_len = 3
}

indexer
{
    mem_limit = 32M
}

searchd
{
    listen = 9306:mysql41
    log = /app/logs/searchd.log
    pid_file = /var/run/searchd.pid
    binlog_path =
}
```


Полнотекстовый поиск

По вселенной marvel

```
SELECT id, name
FROM marvel
WHERE MATCH('man')
LIMIT 3;
```

id	name
5092	Moon Man (Earth-616)
1	Spider-Man (Peter Parker)
4	Iron Man (Anthony \"Tony\" Stark)

3 rows in set (0.00 sec)

```
SELECT id, name
FROM marvel
WHERE MATCH('men')
LIMIT 3;
```

id	name
5092	Moon Man (Earth-616)
1	Spider-Man (Peter Parker)
4	Iron Man (Anthony \"Tony\" Stark)

3 rows in set (0.00 sec)

Фасетные фильтры

Сбор значений

```
SELECT id, name FROM marvel WHERE MATCH('spider') LIMIT 3
FACET character
FACET eye
FACET sex;
```

```
+-----+-----+
| id    | name                                     |
+-----+-----+
| 16172 | \"Spider-Girl\" (Mutant\Spider Clone) (Earth-616) |
|      | Spider-Man (Peter Parker)              |
| 2848  | Peter Parker (Spider-Skeleton) (Earth-616) |
+-----+-----+
3 rows in set (0.00 sec)
```

```
+-----+-----+
| character | count(*) |
+-----+-----+
| Neutral   | 3        |
| Good      | 3        |
| Bad       | 8        |
|           | 6        |
+-----+-----+
4 rows in set (0.00 sec)
```

```
+-----+-----+
| eye      | count(*) |
+-----+-----+
| Brown    | 1        |
| Hazel    | 1        |
|          | 17       |
| Red      | 1        |
+-----+-----+
4 rows in set (0.00 sec)
```

```
+-----+-----+
| sex      | count(*) |
+-----+-----+
| Female   | 3        |
| Male     | 14       |
|          | 3        |
+-----+-----+
3 rows in set (0.00 sec)
```

Отдельный Фасет

Перечислимый

```
SELECT
  GROUPBY() as value,
  COUNT(*) as count
FROM marvel
WHERE eye!=''
GROUP BY eye
ORDER BY count DESC;
```

value	count
Blue	1962
Brown	1924
Green	613
Black	555
Red	508
White	400
Yellow	256
Grey	95
Hazel	76
Variable	49
Purple	31
Orange	25
Pink	21
One Eye	21
Gold	14
Silver	12
Violet	11
Amber	10
Multiple	7
No	7

20 rows in set (0.00 sec)

Отдельный Фасет

Интервальный

```
SELECT
    MIN(year),
    MAX(year)
FROM marvel
WHERE MATCH('spider man') AND year>0;
```

```
+-----+-----+
| min(year) | max(year) |
+-----+-----+
|      1962 |      2008 |
+-----+-----+
1 row in set (0.00 sec)
```

Спасибо за внимание



Рассмотренный пример индекса можно скачать и поиграться

<https://github.com/ivankomolin/lesson-sphinx>