

SYSTEMY BAZ DANYCH

SQL



TEMAT 21-08: Wyszukiwanie pełnotekstowe.

Autor dokumentu: Wojciech Galiński

poniedziałek, 21 listopada 2016 r.

351203 Technik informatyk

ŹRÓDŁA WIEDZY: <http://dev.mysql.com/doc/>.

Zagadnienia obowiązkowe

1. **Wyszukiwanie pełnotekstowe** – to wyszukiwanie wykorzystujące sztuczną inteligencję (na wzór wyszukiwarki „Google”). Posiada następujące cechy podstawowe:

Oto wybrane operatory wyszukiwania pełnotekstowego:

- **brak operatora** – fragment opcjonalny, ale wiersze zawierające taki fragment oceniane są wyżej;
- **operator +** – fragment wymagany;
- **operator -** – fragment niedozwolony;
- **operator ()** – grupowanie;
- **operator "** – traktowanie jako pojedynczy ciąg

Cechy charakterystyczne wyszukiwania pełnotekstowego:

- zaczyna sensownie działać, gdy w bazie znajduje się więcej niż 10 rekordów (wśród nich powinny znaleźć się podobne rekordy);
 - działa nawet dla tekstu o rozmiarze 1 GB;
 - wartość semantyczna wiersza jest tym większa, im więcej razy wzorzec występuje w wierszu;
 - wartość semantyczna wiersza jest tym większa, im wiersz jest krótszy;
 - wartość semantyczna rekordu przyjmuje wartość zero, jeśli wzorzec nie występuje w wierszu;
 - wartość semantyczna wyrazów krótkich lub występujących w co najmniej połowie wierszy wynosi 0;
 - dla mało popularnych fraz mogą nie zostać zwrócone żadne wyniki (porównywana jest zawartość bazy danych);
 - może zwrócić rekordy, które nie zawierają szukanej frazy, ale ich tekst jest w jakimś sensie podobny do niej.
2. **Wyszukiwanie pełnotekstowe w trybie „BOOLEAN”** – w tym trybie indeks „FULLTEXT” nie jest wymagany.

SELECT * FROM tabela

WHERE MATCH (kolumny) AGAINST (wzorzec_tekstowy IN BOOLEAN MODE);

PRZYKŁAD:

-- Poniższe zapytanie może zwrócić zarówno imiona, jak i nazwiska z ciągiem 'Be'

```
SELECT * FROM osoba
```

```
WHERE MATCH (imie,nazwisko) AGAINST ('+Beata -Adam' IN BOOLEAN MODE);
```

```
SELECT * FROM osoba
```

```
WHERE MATCH (imie,nazwisko) AGAINST ('+(Beata Adam) -"Bea ta"' IN BOOLEAN MODE);
```

3. **Wyszukiwanie pełnotekstowe w trybie „NATURAL”** – wymaga zdefiniowania indeksu „FULLTEXT”. Oto wzorzec definicji takiego indeksu:

```
FULLTEXT(kolumna_tekstowa_1, kolumna_tekstowa_2, kolumna_tekstowa_3)
```

Następnie możliwe jest wyszukiwanie danych tekstowych za pomocą następującej składni:

```
SELECT * FROM tabela  
WHERE MATCH (kolumny) AGAINST (wzorzec_tekstowy IN NATURAL MODE);
```

PRZYKŁAD:

```
-- Poniższe zapytanie zwraca wartość semantyczną porównywanych rekordów  
SELECT imie, nazwisko, MATCH(imie,nazwisko)AGAINST('Alan') AS 'Wartość' FROM osoba;
```

4. **Wyszukiwanie pełnotekstowe w trybie „WITH QUERY EXPANSION”** – to specjalny tryb, w którym występują 2 przebiegi wyszukiwania i znajdowane są wyrazy o podobnym znaczeniu.

Oto wzorzec:

```
SELECT * FROM tabela  
WHERE MATCH (kolumny) AGAINST (wzorzec_tekstowy WITH QUERY EXPANSION);
```

PRZYKŁAD:

```
-- Poniższe zapytanie może zwrócić zarówno imiona, jak i nazwiska z ciągiem 'Be'  
SELECT * FROM osoba WHERE MATCH (imie,nazwisko) AGAINST ('prof' WITH QUERY EXPANSION);
```

5. **Wymagania dla tabel do obsługi wyszukiwania pełnotekstowego** – wyszukiwanie pełnotekstowe można stosować jedynie w tabelach opartych o silnik „MyISAM” na kolumnach ze zdefiniowanym indeksem „FULLTEXT”.

PRZYKŁAD:

```
CREATE TABLE Artykul  
(  
  Id INT PRIMARY KEY AUTO_INCREMENT,  
  Tytuł VARCHAR(255) NOT NULL,  
  Tresc VARCHAR(255) NOT NULL,  
  FULLTEXT(Tytuł, Tresc)  
) ENGINE=MyISAM;
```

Pytania kontrolne

1. Czym jest wyszukiwanie pełnotekstowe?
2. Jakie znasz operatory wyszukiwania pełnotekstowego? Wymień je i opisz.
3. Jakie cechy charakterystyczne wyszukiwania pełnotekstowego można wymienić?
4. Czym różni się wyszukiwanie w trybie „BOOLEAN” od wyszukiwania w trybie „NATURAL”?
5. Czym charakteryzuje się wyszukiwanie pełnotekstowe w trybie „WITH QUERY EXPANSION”?
6. Jakie są wymagania dla tabel, żeby obsługiwały wyszukiwanie pełnotekstowe?

Zadania

1. Utwórz tabelę umożliwiającą wyszukiwanie pełnotekstowe, wypełnij ją min. 20 rekordami i sprawdź jak działa wyszukiwanie pełnotekstowe.