

Re: faster search engine for fulltext search

Source: <http://coding.derkeiler.com/Archive/Perl/perl.beginners/2006-06/msg00024.html>

- *From:* orasnita@xxxxxxxxxx (Octavian Rasnita)
 - *Date:* Thu, 1 Jun 2006 18:03:43 +0300
-

Hi,

Here is the SQL query I have used in a table with less than 200.000 records, and I have searched for a single word named "deputat" which was found in 2208 records, but only 20 of them were returned.

Here is the table:

```
CREATE TABLE `articles` (  
  `id_newspapers` smallint(3) unsigned NOT NULL default '0',  
  `id_sections` smallint(3) unsigned NOT NULL default '0',  
  `id` int(6) unsigned NOT NULL auto_increment,  
  `hash` varchar(16) NOT NULL default "",  
  `url` varchar(255) NOT NULL default "",  
  `full_url` varchar(255) NOT NULL default "",  
  `pre_title` varchar(255) NOT NULL default "",  
  `title` varchar(255) NOT NULL default "",  
  `post_title` varchar(255) NOT NULL default "",  
  `body` text NOT NULL,  
  `body_hash` varchar(32) NOT NULL default "",  
  `article_ident` varchar(255) NOT NULL default "",  
  `date` date NOT NULL default '0000-00-00',  
  `time` time NOT NULL default '00:00:00',  
  `id_categories` tinyint(2) unsigned NOT NULL default '0',  
  `active_view` tinyint(1) unsigned NOT NULL default '1',  
  `size` smallint(5) unsigned NOT NULL default '0',  
  PRIMARY KEY (`id_newspapers`,`id_sections`,`id`),  
  UNIQUE KEY `hash` (`hash`),  
  UNIQUE KEY `id_newspapers` (`id_newspapers`,`body_hash`),  
  KEY `date` (`date`),  
  KEY `id_categories` (`id_categories`),  
  KEY `size` (`size`),  
  KEY `title` (`title`),  
  KEY `time` (`time`),  
  FULLTEXT KEY `full` (`pre_title`,`title`,`post_title`,`body`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1
```

And this is the SQL query:

Re: faster search engine for fulltext search

```
select sql_calc_found_rows hash, title, date, substr(body, 1, 250) as
preview, match(pre_title, title, post_title, body) against('deputat') as
rank from articles where match(pre_title, title, post_title, body)
against('deputat' in boolean mode) order by date desc, rank desc limit 0,20;
```

This query took more than 12 seconds the first time when I have used it. After using it a second time, it takes much less because it was cached. I have used 2 times the match() function, because I want to be able to search in boolean mode (in order to be able to use +, -, and * operators) but I have seen that the queries are same as slow if I use it only once, so this is not a problem.

I have tried to "explain" this query, and here is the result:

```
explain select sql_calc_found_rows hash, title, date, substr(body, 1, 250)
as preview, match(pre_title, title, post_title, body) against('deputat') as
rank from articles where match(pre_title, title, post_title, body)
against('deputat' in boolean mode) order by date desc, rank
desc limit 0,20\G
```

```
id: 1
select_type: SIMPLE
table: articles
type: fulltext
possible_keys: full
key: full
key_len: 0
ref:
rows: 1
Extra: Using where; Using filesort
```

So, what do you think, for such a query, can I do something to make MySQL work faster, or do you think that there are other ways to do it faster in perl?

(I am thinking that maybe there are some CPAN modules that can create a fulltext index and search faster with it... creating my own personalised database maybe...)

If these queries work so slow, I am wondering how slow it will work when the database will have a million records.

Thank you.

Teddy

----- Original Message -----

From: "JupiterHost.Net" <mlists@xxxxxxxxxxxxxxxxxx>

To: <beginners@xxxxxxx>

Sent: Thursday, June 01, 2006 4:37 PM

Subject: Re: faster search engine for fulltext search

Re: faster search engine for fulltext search

Perhaps doing your queries to return only some instead of all records will help, a great module for doing this is:

<http://search.cpan.org/perl/doc/Data::Pagate>

I am using a limit clause and the sql query returns at most 20 records,

but

when very many records are found, the search is very slow, even if it returns only 20.

I know how many of them are found, because I use found_rows() in order

to

```
perldoc -f found_rows  
perldoc DBI  
perldoc DBD::mysql
```

have no refernce to that function?

Care to reference found_rows()'s docs?

show how many records would be found if no limit clause would be used.

Depending on the DB it may or may not matter "speed"-wise(MySQL's LIMIT is much more efficient than Postgre's for example) the point for the OP was this:

pagination will help your script be faster in the sense that the array of records its working with is 20 instead of 20,000,000 (IE independant of how efficient the DB is with a proper query, which is not a perl issue in the least)

it also makes it easier to use for your end user, but thats a UI issue.

--

To unsubscribe, e-mail: beginners-unsubscribe@xxxxxxxx
For additional commands, e-mail: beginners-help@xxxxxxxx
<<http://learn.perl.org/>> <<http://learn.perl.org/first-response>>

Re: faster search engine for fulltext search