

MySQL Fulltext Search Score Explained

Asked 16 years, 2 months ago Modified 6 years, 10 months ago Viewed 27k times



I've been experimenting with fulltext search lately and am curious about the meaning of the Score value. For example I have the following query:

15



```
SELECT table. * ,  
MATCH (  
col1, col2, col3  
)  
AGAINST (  
'+(Term1) +(Term1)'  
) AS Score  
FROM table  
WHERE MATCH (  
col1, col2, col3  
)  
AGAINST (  
'+(Term1) +(Term1)'  
)
```

In the results for Score I've seen results, for one query, between 0.4667041301727 to 11.166275978088. I get that it's MySQLs idea of relevance (the higher the more weight).

What I don't get is how MySQL comes up with that score. Why is the number not returned as a decimal or something besides ?

How come if I run a query "IN BOOLEAN MODE" does the score always return a 1 or a 0 ? Wouldn't all the results be a 1?

Just hoping for some enlightenment. Thanks.

mysql

full-text-search

Share Improve this question Follow

asked Oct 23, 2008 at 15:16



Eric Lamb

1,434 ● 5 ● 17 ● 29

2 Answers

Sorted by: Highest score (default)



Take the query "word1 word2" as an example.

12

BOOLEAN mode indicates that your entire query matches the document (e.g. it contains both word1 AND word2). Boolean mode is a strict match.



The formula normally used is based on the Vector Space Model of searching. Very simplified, it figures out two measures to determine how important a word is to a query. The term frequency (terms that occur often in a document are more important than other terms) and the inverse document frequency (a term that occurs in many documents is weighted lower than a term that occurs in few documents). This is known as [tf-idf](#), and is used as a basis for the vector space model. These scores form the basis for the [Vector Space Model](#), which someone else can explain thoroughly. :)

Share

edited Mar 19, 2017 at 19:40

answered Oct 24, 2008 at 11:57

Improve this answer



[Vegard Larsen](#)

13k ● 14 ● 60 ● 102

Follow



9

Generally relevance is based on how many matches each row has to the words given to the search. The exact value will depend on many things, but it really only matters for comparing to other relevance values in the same query.



If you really want the math behind it, you can find it at the [internals manual](#).



Share

edited Feb 2, 2018 at 14:40

answered Oct 23, 2008 at 15:36

Improve this answer



[Mike Doe](#)

17.5k ● 11 ● 73 ● 94



[Harrison Fisk](#)

7,144 ● 3 ● 26 ● 14

Follow



May I display to client the value 11.166275978088 as "relevance 11%"? – [se_pavel](#) May 18, 2009 at 17:37

1 that would be a bad idea...its not accurate that way...no – [johnnietheblack](#) Jan 12, 2010 at 21:53

@se_pavel rather I think what you could do instead is get the sum of the score, divide it by 11.1662xx.. and multiply it by 100. If my math is not haywire, you should be able to get the relevance percentage easily. Example: $11/159.399 \times 100 = 6.90092158671\%$ – [lhsan](#) Jan 28, 2020 at 3:45