

# MySQL Full-Text Search Tutorial

— by royalwzy



# Personal Profile

Database technology  
enthusiast

Oracle 10g/11g OCM  
MySQL 5.6 Database Administrator  
MCITP/MCTS DBA(MSSQL2008)  
RHCE  
Java Programmer

Database technology  
researcher

mysqlxp  
mysqlclone

Database technology  
user

More than 5 years experience in database use and  
management  
Proficient in synchronizing data between heterogeneous  
database  
Skilled in designing and optimizing database  
Install and configure MySQL cluster and high-availability  
cluster environment  
Synchronize and merge data from Oracle to MySQL

Database technology  
sharer

Oracle ACEA, Oracle Young Expert  
ACOUG/SHOUG/OCMU core member  
ITPUB Moderator and Community Expert  
Senior lecturer in "ENMO" and "UPLOOKING"  
Active in OTN & ITPUB and other forums  
My blog:royalwzy.com

---

# Table of Contents

---

- ❖ MySQL Full-Text Search Introduction
- ❖ Three Types of Full-Text Searches
- ❖ MySQL Full-Text Stopwords
- ❖ Fine-Tuning MySQL Full-Text Search
- ❖ MySQL Full-Text Restrictions

---

# MySQL Full-Text Search Introduction

---

- ❖ Exciting feature
- ❖ Switch tables to InnoDB

---

# yntax of Full-Text Search

---

MATCH (col1,col2,...) AGAINST (expr [search\_modifier])

search\_modifier:

```
{  
  IN NATURAL LANGUAGE MODE  
| IN NATURAL LANGUAGE MODE WITH QUERY EXPANSION  
| IN BOOLEAN MODE  
| WITH QUERY EXPANSION  
}
```

---

# Three Types of Full-Text Searches

---

- ❖ Natural Language Full-Text Searches
- ❖ Boolean Full-Text Searches
- ❖ Full-Text Searches with Query Expansion

---

# Natural Language Full-Text Searches

---

- ❖ By default or with the `IN NATURAL LANGUAGE MODE` modifier
- ❖ `MATCH()` returns a relevance value
- ❖ `AGAINST` takes a search string and an optional modifier to search for

# Example of Natural Language Full-Text Searches

```
mysql> SELECT * FROM articles;
```

id	title	body
1	MySQL Tutorial	This database tutorial ...
2	How To Use MySQL	After you went through a ...
3	Optimizing Your Database	In this database tutorial ...
4	MySQL vs. YourSQL	When comparing databases ...
5	MySQL Security	When configured properly, MySQL ...
6	Database, Database, Database	database database database
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..

```
8 rows in set (0.00 sec)
```

```
mysql> SELECT id, title, body, MATCH (title,body) AGAINST ('database' IN NATURAL LANGUAGE MODE) AS score  
-> FROM articles  
-> WHERE MATCH (title,body) AGAINST ('database' IN NATURAL LANGUAGE MODE);
```

id	title	body	score
6	Database, Database, Database	database database database	1.0886961221694946
3	Optimizing Your Database	In this database tutorial ...	0.36289870738983154
1	MySQL Tutorial	This database tutorial ...	0.18144935369491577

```
3 rows in set (0.00 sec)
```



---

# What does Last Slide Show?

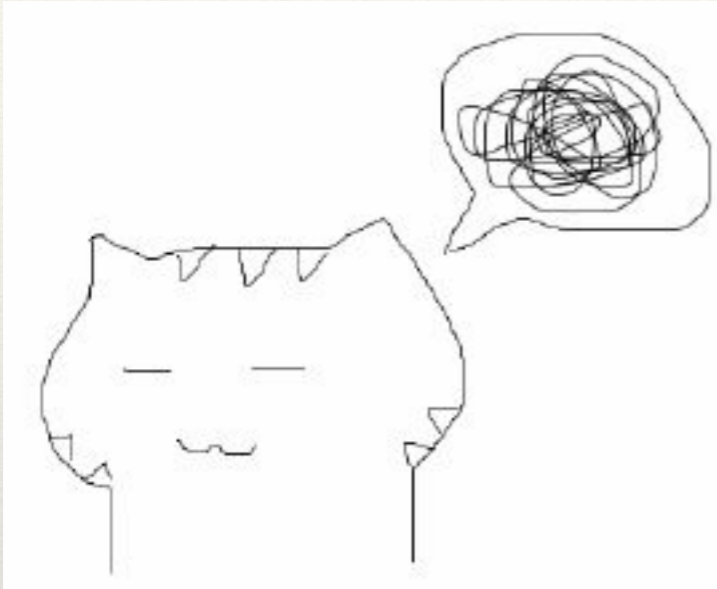
---

- ❖ By default, performed in case-insensitive fashion(latin1->latin1\_bin)
- ❖ The rows returned are automatically sorted with the highest relevance first
- ❖ Relevance is computed based on
  - ❖ The number of words in the row
  - ❖ The number of unique words in that row
  - ❖ The total number of words in the collection
  - ❖ The number of documents (rows) that contain a particular word

---

# Boolean Full-Text Searches

---



What if you want to retrieve records that include word "YourSQL" and exclude word "MySQL"?

```
SELECT * FROM articles  
WHERE MATCH (title,body)  
AGAINST ('+YourSQL -MySQL' IN BOOLEAN MODE);
```



---

# Operators in Boolean Full-Text Search

---

❖ (no operator)

❖ +

❖ -

❖ ~

❖ > <

❖ ( )

❖ \*

❖ “ ”

❖ @distance

---

# Examples Using Boolean Full-Text Operators

---

- ❖ 'apple banana'
- ❖ '+apple +juice'
- ❖ '+apple macintosh'
- ❖ '+apple -macintosh'
- ❖ '+apple ~macintosh'
- ❖ '+apple +(>plus <minus)'
- ❖ 'apple\*'
- ❖ '"some words"'
- ❖ '"word1 word2 word3" @8'

---

# Quizzes

---

❖ Which search string will return more results?

A: '+word +the\*'

B: '+word +the'

❖ What is the relevancy ranking of the following three search strings in row "The Macintosh is a series of personal computers (PCs) designed, developed, and marketed by Apple Inc."

A: '+apple macintosh'

B: '+apple -macintosh'

C: '+apple ~macintosh'

---

# How Relevancy Ranking is Calculated

---

- ❖ InnoDB full-text search's algorithms are based on BM25 and TF-IDF ranking algorithms
- ❖ TF-IDF(Term Frequency-Inverse Document Frequency)
- ❖ May differ from MyISAM relevancy rankings

---

# How Relevancy Ranking is Calculated

---

- ❖  $\{TF\}$
- ❖  $\{IDF\} = \log_{10}( \{total\_records\} / \{matching\_records\} )$ 
  - ❖  $total\_records$  is the number of records in the collection
  - ❖  $matching\_records$  is the number of records that the search term appears in
- ❖  $\{Ranking\} = \{TF\} * \{IDF\} * \{IDF\}$

# Example of Relevancy Ranking for a Single Word Search

```
mysql> SELECT * FROM articles;
```

id	title	body
1	MySQL Tutorial	This database tutorial ...
2	How To Use MySQL	After you went through a ...
3	Optimizing Your Database	In this database tutorial ...
4	MySQL vs. YourSQL	When comparing databases ...
5	MySQL Security	When configured properly, MySQL ...
6	Database, Database, Database	database database database
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..

```
8 rows in set (0.00 sec)
```

```
mysql> SELECT id, title, body, MATCH (title,body) AGAINST ('database' IN BOOLEAN MODE) AS score  
-> FROM articles  
-> ORDER BY score DESC;
```

id	title	body	score
6	Database, Database, Database	database database database	1.0886961221694946
3	Optimizing Your Database	In this database tutorial ...	0.36289870738983154
1	MySQL Tutorial	This database tutorial ...	0.18144935369491577
2	How To Use MySQL	After you went through a ...	0
4	MySQL vs. YourSQL	When comparing databases ...	0
5	MySQL Security	When configured properly, MySQL ...	0
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...	0
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..	0

```
8 rows in set (0.00 sec)
```



---

## Example of Relevancy Ranking for a Single Word Search

---

❖ total\_records:8

```
mysql> SELECT 6 * log10(8/3) * log10(8/3) AS ranking FROM dual;
+-----+
| ranking |
+-----+
| 1.0886961646869382 |
+-----+
1 row in set (0.00 sec)
```

❖  $\text{\${Ranking}}=6*\log_{10}(8/3)*\log_{10}(8/3)$

# Example of Relevancy Ranking for a Multiple Word Search

- ❖ The relevancy ranking value is a sum of the relevancy ranking value for each word
- ❖  $\{\text{rank}\} = \{\text{TF}\} * \{\text{IDF}\} * \{\text{IDF}\} + \{\text{TF}\} * \{\text{IDF}\} * \{\text{IDF}\}$

```
mysql> SELECT id, title, body, MATCH (title,body) AGAINST ('mysql tutorial' IN BOOLEAN MODE) AS score  
-> FROM articles  
-> ORDER BY score DESC;
```

id	title	body	score
1	MySQL Tutorial	This database tutorial ...	0.7405621409416199
3	Optimizing Your Database	In this database tutorial ...	0.3624762296676636
5	MySQL Security	When configured properly, MySQL ...	0.031219376251101494
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..	0.031219376251101494
2	How To Use MySQL	After you went through a ...	0.015609688125550747
4	MySQL vs. YourSQL	When comparing databases ...	0.015609688125550747
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...	0.015609688125550747
6	Database, Database, Database	database database database	0

```
8 rows in set (0.00 sec)
```

---

# Example of Relevancy Ranking for a Multiple Word Search

---

❖ total\_records:8

❖ matching records:6 for "mysql" and 2 for "tutorial"

```
mysql> SELECT (1*log10(8/6)*log10(8/6)) + (2*log10(8/2)*log10(8/2)) AS ranking FROM dual;
+-----+
| ranking |
+-----+
| 0.7405621541938004 |
+-----+
1 row in set (0.00 sec)
```

❖  $\text{\${Ranking}}=(1*\log_{10}(8/6)*\log_{10}(8/6))+(2*\log_{10}(8/2)*\log_{10}(8/2))$

---

# Full-Text Searches with Query Expansion

---

- ❖ Full-text search supports query expansion
- ❖ It's generally useful when a search phrase is too short
- ❖ Relying on implied knowledge that the full-text search engine lacks
- ❖ Enabled by adding `WITH QUERY EXPANSION` or `IN NATURAL LANGUAGE MODE WITH QUERY EXPANSION`
- ❖ For example, a user searching for “database” may really mean that “MySQL”, “Oracle”, “DB2”, and “RDBMS”

# What's is Implied Knowledge?

[https://www.baidu.com/s?wd=梁思成的老婆的情人的老婆是谁%3F&rsv\\_spt=1&rsv\\_iqid=0xc068a67800004033&issp=1&f=8&rsv\\_b...](https://www.baidu.com/s?wd=梁思成的老婆的情人的老婆是谁%3F&rsv_spt=1&rsv_iqid=0xc068a67800004033&issp=1&f=8&rsv_b...)

[https://www.sogou.com/web?query=梁思成的老婆的情人的老婆是谁%3F&\\_asf=www.sogou.com&\\_ast=&w=01019900&p=400401...](https://www.sogou.com/web?query=梁思成的老婆的情人的老婆是谁%3F&_asf=www.sogou.com&_ast=&w=01019900&p=400401...)

**搜狗搜索** [新闻](#) [网页](#) [微信](#) [问问](#) [图片](#) [视频](#) [音乐](#) [地图](#) [购物](#) [更多>>](#)

梁思成的老婆的情人的老婆是谁?

搜狗搜索

全部时间

## 梁思成老婆情人老婆



陆小曼  
第二任妻子



张幼仪  
第一任妻子

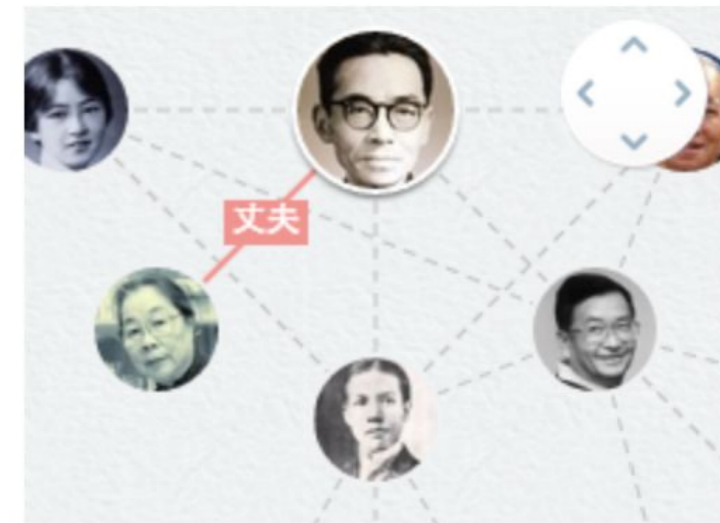
推理说明：梁思成的第一任妻子是林徽因。林徽因的前男友是徐志摩。梁思成的第一任妻子是林徽因。林徽因的前男友是徐志摩。徐志摩的第一任妻子是张幼仪。徐志摩的第二任妻子是陆小曼。

来自搜狗知识图谱 [反馈](#)

## 梁思成的关系图谱

● 亲情 ● 友情 ● 爱情

[查看更多](#)



# Example of Query Expansion

```
mysql> SELECT * FROM articles
-> WHERE MATCH (title,body) AGAINST ('database' IN NATURAL LANGUAGE MODE);
```

id	title	body
6	Database, Database, Database	database database database
3	Optimizing Your Database	In this database tutorial ...
1	MySQL Tutorial	This database tutorial ...

3 rows in set (0.00 sec)

```
mysql> SELECT * FROM articles
-> WHERE MATCH (title,body) AGAINST ('database' WITH QUERY EXPANSION);
```

id	title	body
3	Optimizing Your Database	In this database tutorial ...
1	MySQL Tutorial	This database tutorial ...
6	Database, Database, Database	database database database
5	MySQL Security	When configured properly, MySQL ...
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..
2	How To Use MySQL	After you went through a ...
4	MySQL vs. YourSQL	When comparing databases ...
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...

8 rows in set (0.00 sec)

---

# MySQL Full-Text Stopwords

---

- ❖ Stopwords for InnoDB Search Indexes
  - ❖ INFORMATION\_SCHEMA.INNODB\_FT\_DEFAULT\_STOPWORD table
  - ❖ innodb\_ft\_server\_stopword\_table/innodb\_ft\_user\_stopword\_table
- ❖ Stopwords for MyISAM Search Indexes
  - ❖ storage/myisam/ft\_static.c
  - ❖ ft\_stopword\_file

# Fine-Tuning MySQL Full-Text Search

```
mysql>
mysql>
mysql>
mysql> SELECT * FROM ft_myisam
      -> WHERE MATCH (title) AGAINST ('Security' IN NATURAL LANGUAGE MODE);
```

id	title	body
5	MySQL Security	When configured properly, MySQL ...

1 row in set (0.00 sec)

```
mysql> SELECT * FROM ft_myisam;
```

id	title	body
1	MySQL Tutorial	This database tutorial ...
2	How To Use MySQL	After you went through a ...
3	Optimizing Your Database	In this database tutorial ...
4	MySQL vs. YourSQL	When comparing databases ...
5	MySQL Security	When configured properly, MySQL ...
6	Database, Database, Database	database database database
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..

8 rows in set (0.00 sec)



# Case Study

```
mysql> EXPLAIN SELECT COUNT(*) FROM articles
-> WHERE MATCH (title,body)
-> AGAINST ('database' IN NATURAL LANGUAGE MODE);
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| id | select_type | table | type | possible_keys | key | key_len | ref | rows |
Extra |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | SIMPLE | NULL | NULL | NULL | NULL | NULL | NULL | NULL |
Select tables optimized away |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)
FROM articles,
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| id | select_type | table | type | possible_keys | key | key_len | ref | rows |
s | Extra |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | SIMPLE | articles | ALL | NULL | NULL | NULL | NULL |
6 | NULL |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)
```

---

# Case Study

---

```
mysql> SELECT id, title, body, MATCH (title,body) AGAINST ('Security implications of running MySQL as root' IN NATURAL LANGUAGE MODE) AS score
-> FROM articles
-> WHERE MATCH (title,body) AGAINST('Security implications of running MySQL as root' IN NATURAL LANGUAGE MODE)
-> ORDER BY MATCH (title,body) AGAINST('Security implications of running MySQL as root' IN NATURAL LANGUAGE MODE);
```

id	title	body	score
1	MySQL Tutorial	This database tutorial ...	0.015609688125550747
2	How To Use MySQL	After you went through a ...	0.015609688125550747
4	MySQL vs. YourSQL	When comparing databases ...	0.015609688125550747
8	MySQL Full-Text Indexes	MySQL fulltext indexes use a ..	0.031219376251101494
7	1001 MySQL Tricks	1. Never run mysqld as root. 2. ...	0.8311812281608582
5	MySQL Security	When configured properly, MySQL ...	0.8467909097671509

```
6 rows in set (0.00 sec)
```

# MySQL Full-Text Monitor

```
mysql> SHOW TABLES LIKE '%FT%';
+-----+
| Tables_in_information_schema (%FT%) |
+-----+
| INNODB_FT_DELETED                    |
| INNODB_FT_DEFAULT_STOPWORD          |
| INNODB_FT_INDEX_TABLE                |
| INNODB_FT_BEING_DELETED              |
| INNODB_FT_INDEX_CACHE                |
| INNODB_FT_CONFIG                     |
+-----+
6 rows in set (0.00 sec)
```

❖ Variable:innodb\_ft\_aux\_table

```
mysql> SET GLOBAL innodb_ft_aux_table = 'dbname/tablename';
Query OK, 0 rows affected (0.00 sec)
```

---

# MySQL Full-Text Restrictions

---

- ❖ Full-text searches are supported for InnoDB and MyISAM tables only
- ❖ Full-text searches are not supported for partitioned tables
- ❖ Full-text indexes can be created only for CHAR, VARCHAR, or TEXT columns
- ❖ Ideographic languages limitations

---

# MySQL Full-Text Restrictions

---

- ❖ All columns in a FULLTEXT index must use the same character set and collation.
- ❖ The MATCH() column list must match exactly the column defined in FULLTEXT index
- ❖ The argument to AGAINST() must be a string value
- ❖ Index hints are more limited for FULLTEXT searches than Non-FULLTEXT

---

# Best Practices

---

- ❖ Drop FULLTEXT index before load large data sets
- ❖ InnoDB Full-Text Plugin
- ❖ InnoDB N-gram parser
- ❖ MeCab Parser

*“Do NOT alter the MySQL sources unless you know what you are doing!!!”*

---

# Variables for MyISAM

---

- ❖ `ft_boolean_syntax`
- ❖ `ft_max_word_len`
- ❖ `ft_min_word_len`
- ❖ `ft_query_expansion_limit`
- ❖ `ft_stopword_file`

---

# Variables for InnoDB

---

- ❖ `innodb_ft_aux_table`
- ❖ `innodb_ft_cache_size`
- ❖ `innodb_ft_enable_diag_print`
- ❖ `innodb_ft_enable_stopword`
- ❖ `innodb_ft_max_token_size`
- ❖ `innodb_ft_min_token_size`



---

# Variables for InnoDB

---

- ❖ `innodb_ft_num_word_optimize`
- ❖ `innodb_ft_result_cache_limit`
- ❖ `innodb_ft_server_stopword_table`
- ❖ `innodb_ft_sort_pll_degree`
- ❖ `innodb_ft_total_cache_size`
- ❖ `innodb_ft_user_stopword_table`
- ❖ `innodb_optimize_fulltext_only`

Q & A