

Mr oonga開発者が 来たぞ！

須藤功平

クリアコード

三木会
2016-11-17





Mroonga

- 読み方：むるんが
- ストレージエンジン
- MySQL 5.7対応
もちろん5.5, 5.6も対応
- MariaDB 10.2対応
もちろん5.5, 10.0, 10.1も対応
 - 10.0以降はバンドルもされている



特徴

- **高速日本語全文検索** (全言語OK)
MySQL 5.6以前は日本語未対応
5.7以降は日本語対応しているが遅い
- **カラムストアによる高速処理**
- **全文検索初心者でも使える**
- **全文検索上級者は活用できる**



高速日本語全文検索

1. ベンチマーク
2. 速さの秘密



ベンチマーク環境

- 対象：Wikipedia日本語版
- レコード数：約185万件
- データサイズ：約7GB
- メモリー4GB・SSD250GB (ConoHa)



補足

- MySQL 5.7を使用
- 他人のベンチマークは参考程度
 - 検討時は実環境でベンチマークを！

詳細：

<https://github.com/groonga/wikipedia-search/issues/4>



検索1

キーワード：テレビアニメ

(ヒット数：約2万3千件)

InnoDB ngram	3m2s
InnoDB MeCab	6m20s
Mroonga: 1	0.11s



検索2

キーワード：データベース

(ヒット数：約1万7千件)

InnoDB ngram	36s
InnoDB MeCab: 1	0.03s
Mroonga: 2	0.09s



検索3

キーワード：PostgreSQL OR MySQL
(ヒット数：約400件)

InnoDB ngram	N/A(Error)
InnoDB MeCab: 1	0.005s
Mroonga: 2	0.028s



検索4

キーワード：日本

(ヒット数：約63万件)

InnoDB ngram	1.3s
InnoDB MeCab	1.3s
Mroonga: 1	0.21s



検索まとめ

- Mroonga : 安定して速い
- InnoDB FTS MeCab
 - ハマれば速い
- InnoDB FTS ngram
 - 安定して遅い



速さの秘密

- 最適化された転置索引実装
 - 2段階のデータ圧縮
 - 高速なポスティングリスト探索
 - 検索だけでなく更新も速い

11年以上開発が続いている全文検索エンジンGroongaを使用

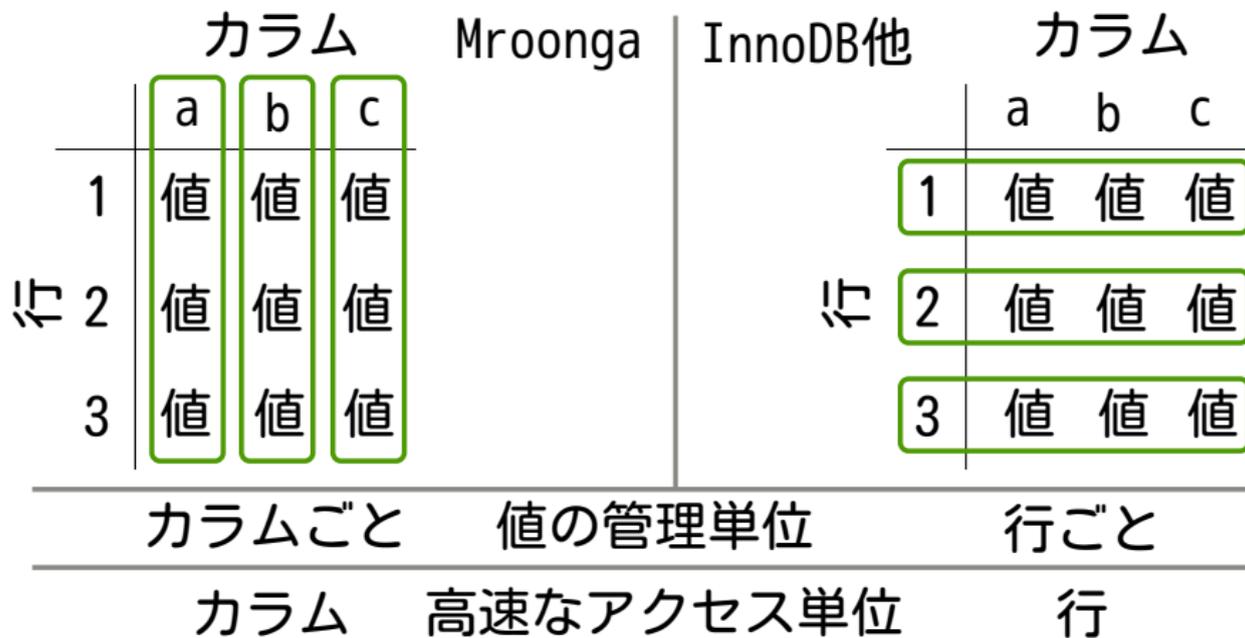


もっと速さの秘密

- カラムストアを活かした最適化
 - ポイント1：余計なI/Oを減らす
 - ポイント2：I/Oを局所化



カラムストア

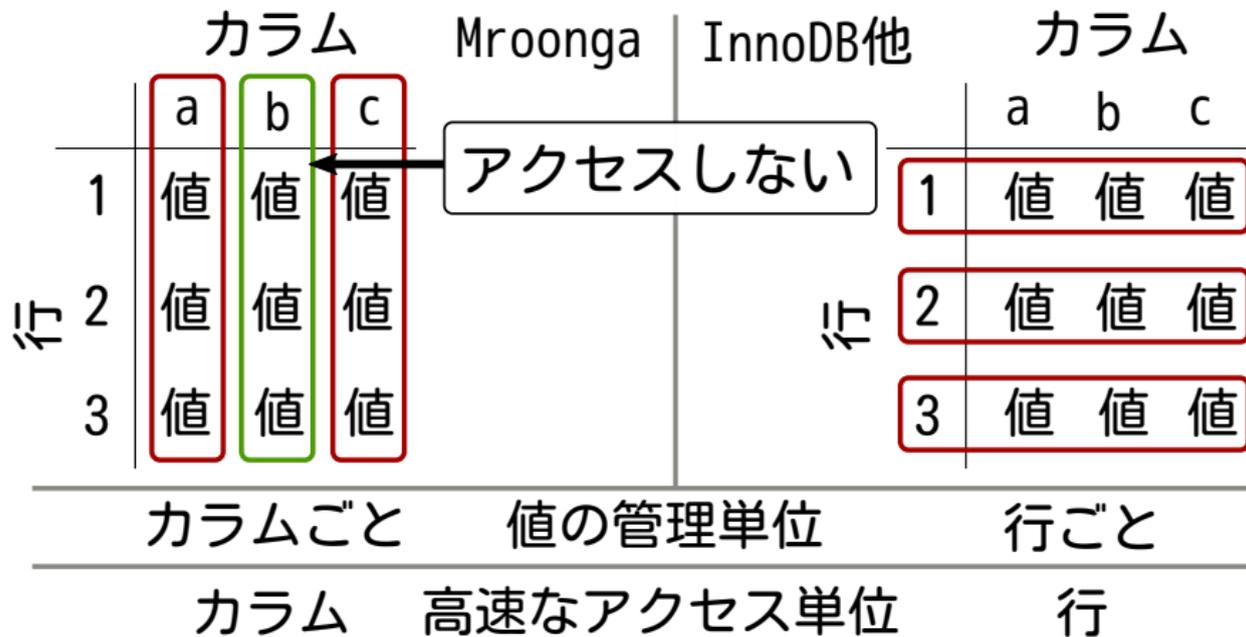


必要なカラムのみアクセス

```
-- aのみにアクセス  
SELECT a  
  FROM table  
  
-- cのみにアクセス  
WHERE c = XXX;  
  
-- bにはアクセスしない
```



減ったI/O



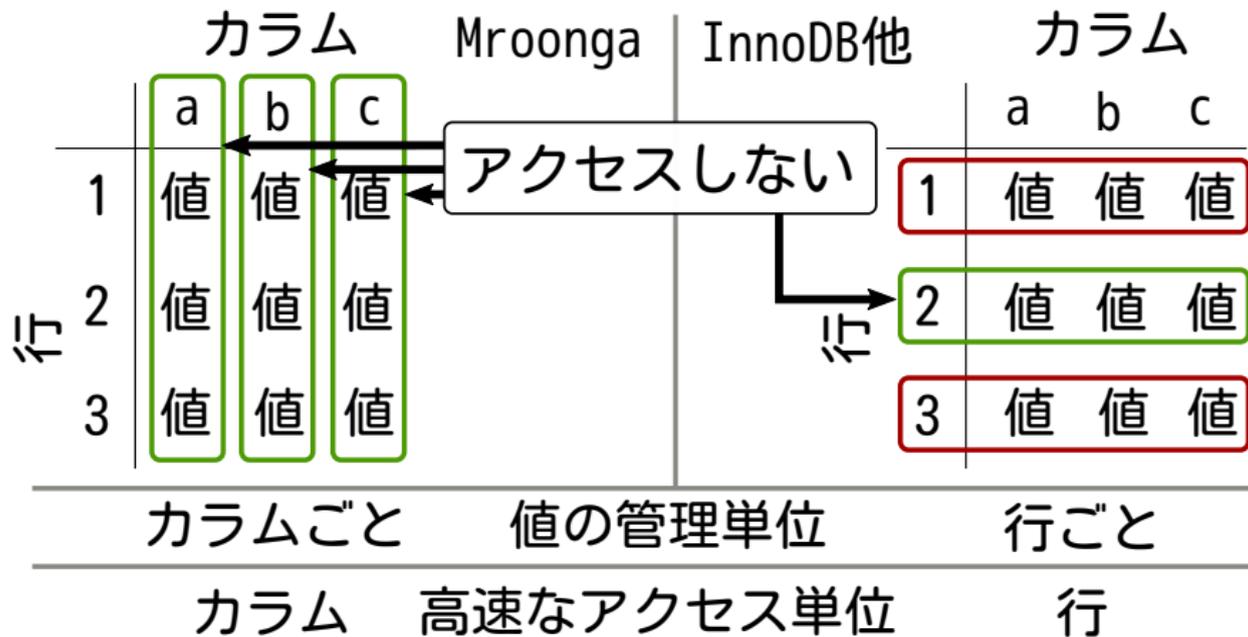


行カウント

```
-- カラムの値は必要ない
SELECT COUNT(*)
  FROM table
-- cの全文検索インデックスにだけアクセス
  WHERE MATCH(c)
        AGAINST('+keyword' IN BOOLEAN MODE);
-- a, b, cはアクセスしない
```



減ったI/O





ORDER BY LIMIT

```
SELECT *  
  FROM table  
 WHERE MATCH(c)  
        AGAINST('+keyword' IN BOOLEAN MODE)  
-- MySQLではなくMroongaがORDER BY LIMITを処理  
-- →Mroongaは10レコードだけMySQLに返す  
--   マッチしたレコードすべては返さない  
ORDER BY a LIMIT 10;
```

ORDER BY LIMITの最適化

- Mrroongaが検索
 - カラム毎の処理でI/Oを局所化
(索引非使用時)
- Mrroongaがソート
 - カラム毎の処理でI/Oを局所化
- MrroongaがOFFSET/LIMITを処理



カラム毎の処理は速い

	カラム			Mroonga	InnoDB他	カラム			
	a	b	c			a	b	c	
1	値	値	値			1	値	値	値
2	値	値	値			2	値	値	値
3	値	値	値			3	値	値	値

カラムごと 値の管理単位 行ごと

カラム 高速なアクセス単位 行



最適化のまとめ

- 転置索引実装が速い
 - 検索も更新も速い
- カラムストアで速い
 - ポイント：I/O削減・I/O局所化

全文検索初心者でも使える

- インストールが簡単
- MySQLの標準機能のみで使える



インストールが簡単

- APT/Yumリポジトリ
- MariaDBバンドル
- MariaDB込みのWindowsバイナリ



標準機能のみで使える

```
-- 作成  
CREATE TABLE table (  
    -- ...,  
    FULLTEXT INDEX (column)  
) ENGINE=Mroonga;
```



標準機能のみで使える

-- 変換

```
ALTER TABLE table  
  ADD FULLTEXT INDEX (column)  
  ENGINE=Mroonga;
```



標準機能のみで使える

```
SELECT * FROM table
WHERE
  MATCH(column)
  AGAINST('+keyword'
           IN BOOLEAN MODE);
```

全文検索上級者向け機能

- カスタマイズ
 - デフォルト値はいい感じ
→初心者はカスタマイズなしでよい
- Groongaの機能をもっと使える
(高速・高機能)



文字正規化ルール変更

```
CREATE TABLE table (  
    -- ...,  
    FULLTEXT INDEX (column)  
    --  
    -- コメントでパラメーターを指定  
    COMMENT='normalizer "NormalizerAuto"  
) ENGINE=Mroonga;
```



文字正規化ルール変更

```
CREATE TABLE table (  
  -- ....,  
  FULLTEXT INDEX (column)  
  -- MariaDBの場合：  
  -- カスタムパラメーターを使える  
  NORMALIZER='NormalizerAuto'  
) ENGINE=Mroonga;
```

Groongaの検索機能を使う

```
SELECT * FROM table  
WHERE
```

```
-- 「*SS」プラグマ使用時は「c1」は無視される  
MATCH(c1)
```

```
-- 「*SS」はGroongaの全検索機能を使うためのプラグマ  
-- 1つのクエリーで複数のインデックスを使用可能
```

```
AGAINST('*SS c1 @ "keyword" && c2 < 100'  
        IN BOOLEAN MODE);
```



今後

- 最新機能サポート
 - JSONを全文検索
(JSON型のデータの読み書きは対応済み)
 - 仮想カラム・生成カラム
 - パーティショニング+全文検索対応
- 最新版をMariaDBにバンドル



最新版をバンドル

- Mroongaは毎月リリース
- MariaDB 10.2.1はMroonga **5.04**をバンドル
 - Mroongaの最新版は6.10
 - Mroongaは**6.03**からMariaDB 10.2をサポート
 - 現在テスト中→テスト対応後マージ



まとめ1

- 高速日本語全文検索 (全言語OK)
- カラムストアによる高速処理
- 全文検索初心者でも使える
- 全文検索上級者は活用できる



まとめ2

- 今後もMrOongaは便利になる
 - MariaDBで最新MrOongaを使える
- MySQLで全文検索ならMrOonga！