

Groonga族 2014

須藤功平

クリアコード

全文検索エンジンGroongaを囲むタベ5
2014/11/29





この時間の目的 (1)

Groongaの 特徴を紹介

(まだGroongaを使っていない人向け)



この時間の目的 (2)

2014年の
Groonga関連の
最新情報を
ざっくり紹介



Groonga族と私

- 開発・リリース・サポート
 - Groonga、Mroonga、Rroonga、Droonga、他各種ツール
- イベント運営
 - Groongaを囲む夕べ、ドキュメント読書会の解説者枠



内容

- Groonga族の概要
- Groonga族の最新情報
- Groonga関連情報の見つけ方



Groonga族の概要

- Groongaの特徴
- Groonga族とは



Groongaの特徴 (1)

即時更新



即時更新

- データ更新時：索引も更新
 - 1つの索引を随時更新
 - ×サブ索引を作成→マージ
 - →すぐに検索できる
- 新鮮な情報をすぐに検索可能に
 - 大事にしていること



Groongaの特徴 (2)

安定した
高い検索性能



安定した高い検索性能

- 索引更新中も検索可能
 - 更新処理が検索処理を阻害しない
 - 更新中も参照可能にすることで実現
- 定期的な重い処理がない
 - ×フルGCによるstop the world
 - ×索引のマージ



Groongaの特徴 (3)

日本語特化の
検索機能

日本語特化の検索機能 (1)

- ローマ字前方一致検索
 - 「to」 → 「トウキョウ」
 - 「tyo」 → 「チョウフ」
 - 「cho」 → 「チョウフ」
(ヘボン式もいける)
 - 「と」 → 「トウキョウ」
(ひらがなもいける)

日本語特化の検索機能 (2)

- ハイブリッドBigram
(↓は正確じゃないけどだいたいあっている説明)
 - 日本語はBigram
 - 英語は空白区切り
 - "日本" "本語" "語は" "Bigram"
- 日英混じりでもBigramを使える
 - 英語Bigramの検索性能の悪さを解消
(英単語のステミングやストップワードもできる)



Groongaの特徴 (4)

他システムと
連携



他システムと連携

- 連携
 - Cライブラリーとして利用可
 - 既存システムに全文検索機能を追加
- メリット
 - Groonga：全文検索以外はお任せ
 - 既存システム：全文検索機能追加
 - 使う側：追加で覚えることが少ない



連携例

- MySQL (Mroonga (むるんが))
- Groonga
 - 通信・レプリケーションなどお任せ
- 既存システム (MySQL)
 - 日本語全文検索対応
- 使う側
 - SQLを使って日本語全文検索できる



Groongaの特徴 (5)

定期リリース



定期リリース

- 毎月肉の日にリリース
(肉の日 == 29日)
- メリット
 - 問題を報告してから修正版が
手に入るまでの時間が短い
 - 最新ディストリビューションへの
対応が速い



リリース！

Groonga 4.0.8
リリース！



リリース！！

Mroonga 4.08
リリース！



リリース！！！！

Droonga 1.0.8
リリース！



Groongaの特徴

- 即時更新
- 安定した高い検索性能
- 日本語特化の検索機能
- 他システムと連携
- 定期リリース



Groonga族の概要

- Groongaの特徴
- Groonga族とは



Groonga族

- Groongaと
関連プロダクトの総称
- 関連プロダクト
 - Xroonga (○るんが)
という名前が多い



関連プロダクトの分類

詳細：<http://groonga.org/ja/related-projects.html>

- データベース
- 言語バインディング
- クライアントライブラリー
- ユーティリティ



データベース

- 既存データベースに全文検索機能を追加
 - 例：MySQL→Mroonga（むるんが）
- 既存データベースのインターフェイスで使える
 - 例：SQL



言語バインディング

- 好きな言語で全文検索機能付きアプリケーションを書ける
 - 例：Ruby→Rroonga（るるんが）
- アプリケーションがDBを開く
 - SQLite3やDBMを使う感じ
 - 1マシンで扱える量向き

クライアントライブラリー

- Groongaサーバーへのアクセスを支援
 - 例：Go→goroo（ごろう）
- クライアント・サーバーモデル
 - MySQLやMongoDBを使う感じ
 - アプリとDBでマシンを分ける構成用



ユーティリティ

- 便利ツールがいろいろある
 - Madhand : Groongaレプリケーション
<https://github.com/yappo/ruby-madhand>
 - groonga-query-log :
スロークエリー調査とか
 - zshの補完関数
 - grnline :
groongaコマンドを便利にする



関連プロダクトの分類

詳細：<http://groonga.org/ja/related-projects.html>

- データベース
- 言語バインディング
- クライアントライブラリー
- ユーティリティ



Groonga族の最新情報

- Groonga族の概要
- Groonga族の最新情報
- Groonga関連情報の見つけ方



2013/12

- Groonga Advent Calendar 2013
 - <http://qiita.com/advent-calendar/2013/groonga>
 - 今年もやるので参加してね！
<http://qiita.com/advent-calendar/2014/groonga>
- geo_in_rectangleが北半球だけでなく全世界をサポート
 - 去年のこのイベントでバグレポート



2013/12

- `between()`関数を追加
 - `1 <= X && X <= 10`
のような範囲クエリーを高速実行
 - SQLでいう"`X BETWEEN 1 AND 10`"



2014/1

- DBサイズ増加を抑制
 - レコードの更新をくりかえすとDBサイズが増えることがある
 - 増加しにくくなった
- Groonga Meetup開催
 - 1/29
 - ハンズオン



2014/2

- 2/9に4.0.0リリース！
 - 年に1度の肉の日



- レコード毎の重み付けに対応
 - スコア調整に便利
 - 例：Groongaタグがついていたらスコアを10倍
- 参照
 - selectのadjusterオプション
 - 重み付きベクターカラム



2014/3

- 全文検索エンジンGroonga
勉強会@札幌開催
 - 3/29



2014/4

- Groonga Meetup開催
 - 4/29
 - ハンズオン



2014/5

- Ubuntuパッケージ配布URL変更
 - LaunchpadのPPAを利用
- HerokuでGroongaを利用可能に
 - Rroongaも使える
<http://www.clear-code.com/blog/2014/5/28.html>



2014/5

- Groongaドキュメント読書会1
 - 5/15
 - Groongaの理解を深める会
 - 集まってドキュメントを読む
 - 参加者が気になったところは
随時Groonga開発者（私）が図解



2014/6

- HTTPサーバーがPOST対応
- GQTPサーバーのデフォルトポート番号を変更
 - 10041→10043
- プラグイン用APIを強化



2014/6

- Try Groonga公開
 - Webブラウザ上でGroongaを試せる
 - インストールいらず
 - Heroku上で動作

<http://try-groonga.herokuapp.com/>



2014/6

- Groongaドキュメント読書会2
 - 6/3
 - Groongaの理解を深める会
 - 集まってドキュメントを読む
 - 参加者が気になったところは
随時Groonga開発者（私）が図解



2014/7

- Groongaドキュメント読書会3
 - 7/23
 - Groongaの理解を深める会
 - 集まってドキュメントを読む
 - 参加者が気になったところは
随時Groonga開発者（私）が図解



2014/8

- キーワードハイライト関数追加
 - `highlight_full`と`highlight_html`
- 近傍検索記法を追加
 - 「*N"A Z"」 → Aの近くにZがある
 - 「近く」の閾値はパラメーター



2014/8

- Groongaドキュメント読書会3
 - 8/25
 - Groongaの理解を深める会
 - 集まってドキュメントを読む
 - 参加者が気になったところは
随時Groonga開発者（私）が図解



2014/9

- mruby実装のクエリーオプティマイザーが動くようになった
 - RubyKaigi 2014があった
- ConoHa presents - Groonga "How-To" Talks開催
 - <http://groonga.org/ja/blog/2014/09/03/conoha-presents-groonga-how-to-talks.html>



2014/9

- packages.groonga.orgで「ConoHa支援プログラム」を利用開始
 - <https://www.conoha.jp/community>
 - ソース・パッケージ配布サイト



2014/10

- Windowsでもmruby組込サポート
- トークンフィルター機能追加
 - プラグイン対応
 - TokenFilterStopWord :
ストップワード
 - TokenFilterStem :
ステミング



2014/10

- カラム圧縮サポート
 - zlib/LZ4
- in_values関数を追加
 - SQLでいう"カラム IN(A, B, ...)"
 - だいぶ速くなる



2014/10

- Groongaドキュメント読書会5
 - 10/27
 - Groongaの理解を深める会
 - 集まってドキュメントを読む
 - 参加者が気になったところは
随時Groonga開発者（私）が図解



2014/11

- n個のキーを使ったドリルダウンをサポート
 - 例：ユーザーとタグでドリルダウン
- range_filterコマンドを追加
 - ヒット件数は返せない代わりに高速
 - SQLでいうORDER BY LIMITで速くなるパターン用のコマンド



2014/11

- 管理画面をリニューアル
 - 今のところ検索画面だけ
 - 使いやすさ重視
 - <https://github.com/groonga/groonga-admin/>



Rroonga

- リリースチームを結成
- 最新Groongaをサポート
- メモリ使用量削減



ユーティリティ

- fluent-plugin-groonga
 - ログをGroongaに投入可能
 - スキーマは自動定義
 - →Groongaサーバーを立てたらすぐに使える
 - Fluentdで集めたデータをGroongaに格納する方法
<http://qiita.com/groonga/items/5c674be9daf25054c14c>

Groonga関連情報の見つけ方

- Groonga族の概要
- Groonga族の最新情報
- Groonga関連情報の見つけ方



オンライン (1)

- <http://groonga.org/>
- メールングリスト
 - groonga-dev
- Twitter
 - @groongaをフォロー
 - 拾ってもらおう



オンライン (2)

- Qiita
 - 「groonga」 タグをフォロー
 - Groonga Advent Calendar
 - <http://qiita.com/advent-calendar/2013/groonga>
 - <http://qiita.com/advent-calendar/2014/groonga>



オフライン

- Groongaを囲む夕べ
 - 11/29 (いい肉の日)
- ドキュメント読書会
 - 1,2ヶ月に1回開催
 - 次回は12/11 (木)