

Rroongaの全文 検索性能

Naoya(@naoa_y)

61st Ruby/Rails勉強会@関西 LT 2014/05/31

自己紹介

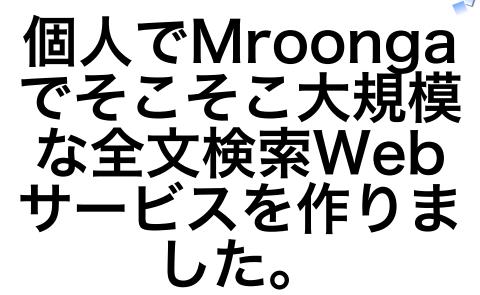
- ✓ Naoya (@naoa_y)
 - √ 大学は情報系
 - ✓ 新卒で3年半ほど金融系のユーザSI でインフラSE
 - ✓ 現在は3年半ほどITと無縁の仕事
 - ✓ Ruby歴は実働数日ほど



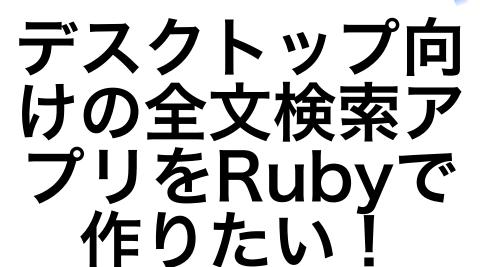
Groongaって?

1

- ✓ C言語で書かれた高速な全文検 索ライブラリ/サーバ
 - ✓ MySQLのストレージエンジンとして実装されたMroonga
 - ✓ Rubyらしい記法でRubyから全文検 索できるRroonga



Groongaは少 ないハードウェ ア資産でもかな り高速な全文権 索ができる



Rroongaって?

1

- ✓ Rroongaのほかにサーバがな くともローカルのみで高速な全 文検索機能が実現可能
- ✓ アプリに組み込みやすい。イメ ージとしてはSQLiteみたいな もの

Rroongaの利用例

- √ るりまサーチ
 - ✓ Rubyのリファレンスマニュアルを 検索できるWebアプリ http:// docs.ruby-lang.org/ja/search/
- Milkode
 - ✓ ソースコード検索エンジン、検索ア プリ http:// milkode.ongaeshi.me/ (るびま 0045号に記事掲載)

実際のところ Rroongao てどのくら 速いの?





検証環境

- ✓ MacBook Air
 - ✓ CPU 1.4GHzデュアルコアIntel Core i5
 - √ Memory 4GB
 - √ SSD 128GB

検証方法

- ✓ 日本語Wikipediaのカテゴリ上位100件を検索
- ✓ カテゴリ上位100件で日本語 Wikipediaのタイトルと本文を 全文検索
- ✓ 検証プログラム http:// github.com/naoa/rroongawikipedia-bench

日本語Wikipedia 1万件 XML 168MB

	SQLite3	Rroonga
カテゴリ	2314	47 msec
	msec	
本文	103	6 msec
	msec	

日本語Wikipedia 5万件 XML 512MB

	SQLite3	Rroonga
カテゴリ	9339	88 msec
	msec	
本文	341	30 msec
	msec	

日本語Wikipedia 183万件(全件) XML 17GB

	Rroonga	
カテゴリ	1601 msec	
本文	56 msec	

日本語Wikipedia 366万件(全件 *2) XML 33GB

	Rroonga
カテゴリ	3087 msec
本文	113 msec







Mroongaでの経験 則から言うとこの数 倍ぐらいなら高速に 全文検索できると思 います。(日本語で Bigramの場合。

Mroongaを使った全 文検索のWebサービ スでは、最大で 400GB超、1000万 レコード超でもサーバ 1台でそこそこ高速に 検索できています。

Rroongaで は、高速な全文 検索のほかに位 置情報検索機能 もあります。

先日、Heroku でRroongaが 使えるビルドパ ックがリリース されています。

Rroonga くなりまし たよね?

おわりに

- ✓ 2014/6/27(金)19:00から神戸でGroongaに関する勉強会をします。
- ✓ Groongaに興味が沸いた方は 是非参加してください!

おわりに

- ✓ Rroonga,Rabbit等の作者、 Ruby/Groongaコミッタの須 藤さん(@ktou)が東京から来て くれます!
- ✓ 2014/6/28(土)に大阪で Groongaに関するイベントが できればと考えています。
- ✓ 会場提供等に協力していだける 方がいたら嬉しいです。