

るりまサーチの作り方

How to create rurema-search

須藤功平

株式会社クリアコード

2010/08/29



話すこと

るりまサーチの使い方

(how to use rurema-search)

- ✓ るりまサーチの紹介
rurema-search introduction

るりまサーチの作り方

(how to create rurema-search)

- ✓ 全文検索システムの作り方
how to create full-text search system
- ✓ るりまサーチの副産物
rurema-search side effects

自己紹介

須藤 功平 (株式会社クリアコード)

株式会社クリアコード代表取締役。
*RSS Parser*と*Rabbit*と*rcairo*と*Ruby-GNOME2*と*ActiveLdap*と
*ActiveSambaLdap*と*test-unit 2.x*と
*rroonga*とるりまサーチと*Ruby/*
*Subversion*と*RWiki*と*Ruby*などの開
発者。

自己紹介: クリアコード

須藤 功平 (株式会社クリアコード)

株式会社クリアコード代表取締役。
*RSS Parser*と*Rabbit*と*rcairo*と*Ruby-GNOME2*と*ActiveLdap*と
*ActiveSambaLdap*と*test-unit 2.x*と
*rroonga*とるりまサーチと*Ruby/*
*Subversion*と*RWiki*と*Ruby*などの開発者。

クリアコード

Gold スポンサーのご紹介



■ ワイクル株式会社

好きなメソッドは「Symbol#to_proc」。



■ 株式会社クリアコード

クリアコードはRubyやMozilla Firefox/Thunderbirdなどのフリーソフトウェアに関する技術力を提供しています。ソースコードがあればなんとかできますので、フリーソフトウェアに関してお困りのことがあればご相談ください。

お問い合わせください。

お問い合わせは、お気軽にメールでお問い合わせください。

お問い合わせは、お気軽にメールでお問い合わせください。

自己紹介: Rabbit

須藤 功平 (株式会社クリアコード)

株式会社クリアコード代表取締役。
*RSS Parser*と*Rabbit*と*rcairo*と*Ruby-GNOME2*と*ActiveLdap*と
*ActiveSambaLdap*と*test-unit 2.x*と
*rroonga*とるりまサーチと*Ruby/*
*Subversion*と*RWiki*と*Ruby*などの開
発者。

Rabbit

nari shugo



Yugui Matz 吳江

znz

okkez

自己紹介: rcairo

須藤 功平 (株式会社クリアコード)

株式会社クリアコード代表取締役。
*RSS Parser*と*Rabbit*と*rcairo*と*Ruby-GNOME2*と*ActiveLdap*と
*ActiveSambaLdap*と*test-unit 2.x*と
*rroonga*とるりまサーチと*Ruby/*
*Subversion*と*RWiki*と*Ruby*などの開
発者。

rcairo



名札には名前を大きく書きましょうジェネレータ

kdmsnr

twitter ID: 生成!



きろ字 - P



青柳衛山フォントT



梅ゴシックS4



美咲明朝



YOzFontEF90

名札には名前を大きく書きましょうジェネレータ

<http://namesplate.clear-code.com/kdmsnr>



Takeo明朝



Serif



M+ 1mn



DejaVu Sans Mono



Takeoゴシック



自己紹介: るりまサーチ

須藤 功平 (株式会社クリアコード)

株式会社クリアコード代表取締役。
*RSS Parser*と*Rabbit*と*rcairo*と*Ruby-GNOME2*と*ActiveLdap*と
*ActiveSambaLdap*と*test-unit 2.x*と
*rroonga*とるりまサーチと*Ruby/*
*Subversion*と*RWiki*と*Ruby*などの開
発者。



るりまサーチの使い方

るりまサーチの使い方

(how to use rurema-search)

✓ るりまサーチの紹介

rurema-search introduction

るりまサーチの作り方

(how to create rurema-search)

✓ 全文検索システムの作り方

how to create full-text search system

✓ るりまサーチの副産物

rurema-search side effects

るりまサーチ

The screenshot shows the homepage of the 'るりまサーチ' (Rurema Search) website. At the top, there is a search bar with a magnifying glass icon and the text 'るりまサーチ'. Below the search bar, there are navigation links for different Ruby versions: 'すべて' (All) with 64429 items, '1.8.7' with 16312 items, '1.8.8' with 16327 items, '1.9.1' with 15812 items, and '1.9.2' with 15978 items. A prompt encourages users to search the Ruby reference manual using 'groonga'. A search input field and a '検索' (Search) button are provided. Below this, a green box lists various Ruby object types and their counts: '種類' (Types) including 'インスタンスメソッド' (33112), '定数' (8845), 'シングルトンメソッド' (8098), 'クラス' (5243), '関数' (2948), 'モジュールファンクション' (2538), 'ライブラリ' (1663), 'モジュール' (964), '変数' (374), and 'マクロ' (364). Another green box highlights '項目数が多い組み込みクラス' (Built-in classes with many items), listing 'String' (539), 'Array' (468), 'IO' (378), 'Encoding' (365), 'Hash' (291), 'File' (283), 'Object' (268), 'Module' (264), and 'Time' (231).

<http://doc.ruby-lang.org/ja/search/>
<http://rurema.clear-code.com/>

るりまサーチとは



- ✓ るりまの成果物を
rurema deliverables
- ✓ Rubyのリファレンスマニュアル
Ruby reference manual
- ✓ 全文検索する
full-text search
- ✓ Webアプリケーション
Web application

デモ



公式サイト (official site)

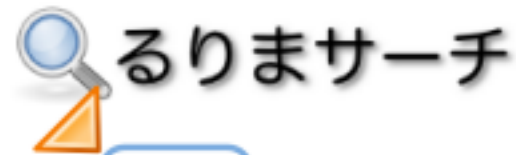
<http://doc.ruby-lang.org/ja/search/>

開発サイト (development site)

<http://rurema.clear-code.com/>

ポイント: ドリルダウン

ドリルダウン:
絞り込み候補と
絞り込み結果を
同時に表示
(ファセットともいう)
・クリックで絞り込み
・0件ヒットを回避



すべて 1.8.7 1.8.8 1.9.1 1.9.2
64429件 16312件 16327件 15812件 15978件

Rubyでgroonga使ってRubyのリファレンスマニュアルを全文検索!

検索

種類 インスタンスメソッド (33112) 定数 (8845)
シングルトンメソッド (8098) クラス (5243)
関数 (2948) モジュールファンクション (2538)
ライブラリ (1663) モジュール (964)
変数 (374) マクロ (364)

項目数が多い組み込みクラス
300件~ String (539) Array (468) IO (378) Encoding (365)
~300件 Hash (291) File (283) Object (268) Module (264) Time (231)

ポイント: URL



るりまサーチ Rubyでgroonga使ってRubyのリファレンスマニュアルを全文検索!

検索

すべて 1.8.7 1.8.8 1.9.1 1.9.2

167件ヒット [1-100件を表示] (0.046秒)

トップページ > インスタンスメソッド > 1.9.2 > クラス:String

http://rurema.clear-code.com/type:instance-method/version:1.9.2/class:String/

ビルトイン (143)
kconv (12)
rake (5)
shellwords (2)
scanf (2)
yaml (2)
bigdecimal/util (1)

String#%(args) -> String (3)

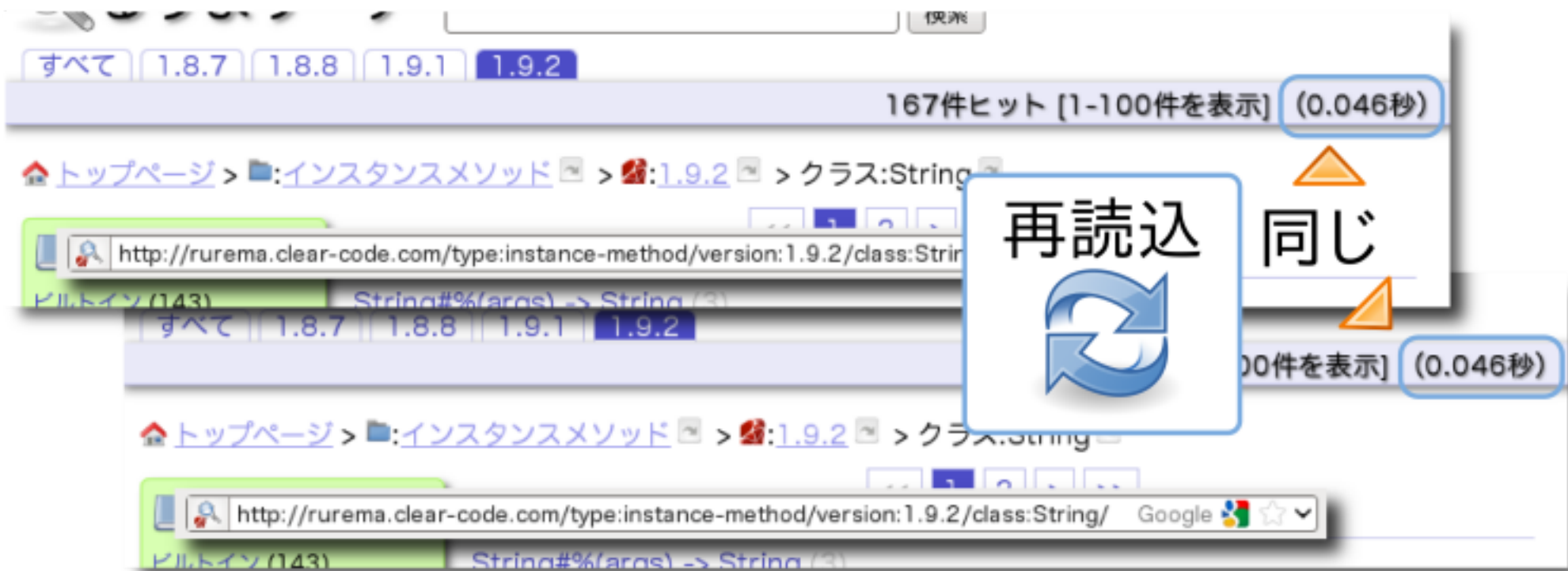
/type:instance-method/ ...フォーマットします。

/version:1.9.2/ ...指定だけが意味を持ちます。:s 文字列を出力します。引数がStringオブジェクトでなければto_sメソッドにより文字列化し...
/class:String/ ...:p Object#inspect の...

URLのパスにクエリを含める(SEO)

String#*(times) -> String (3)

ポイント: キャッシュ



検索結果をキャッシュ(同一クエリ = 同一結果)

- ・レスポンス速度向上
- ・HTML単位
- ・バッチ処理時にクリア

ポイント

- ✓ ドリルダウン
drill down
- ✓ クリックで絞り込み
refine search by clicks
- ✓ URLにクエリ (SEO)
query in URL (SEO)
- ✓ キャッシュ
cache
- ✓ 使いやすさアップ
higher usability





るりまサーチの作り方

るりまサーチの使い方

(how to use rurema-search)

- ✓ ~~るりまサーチの紹介~~
rurema-search introduction

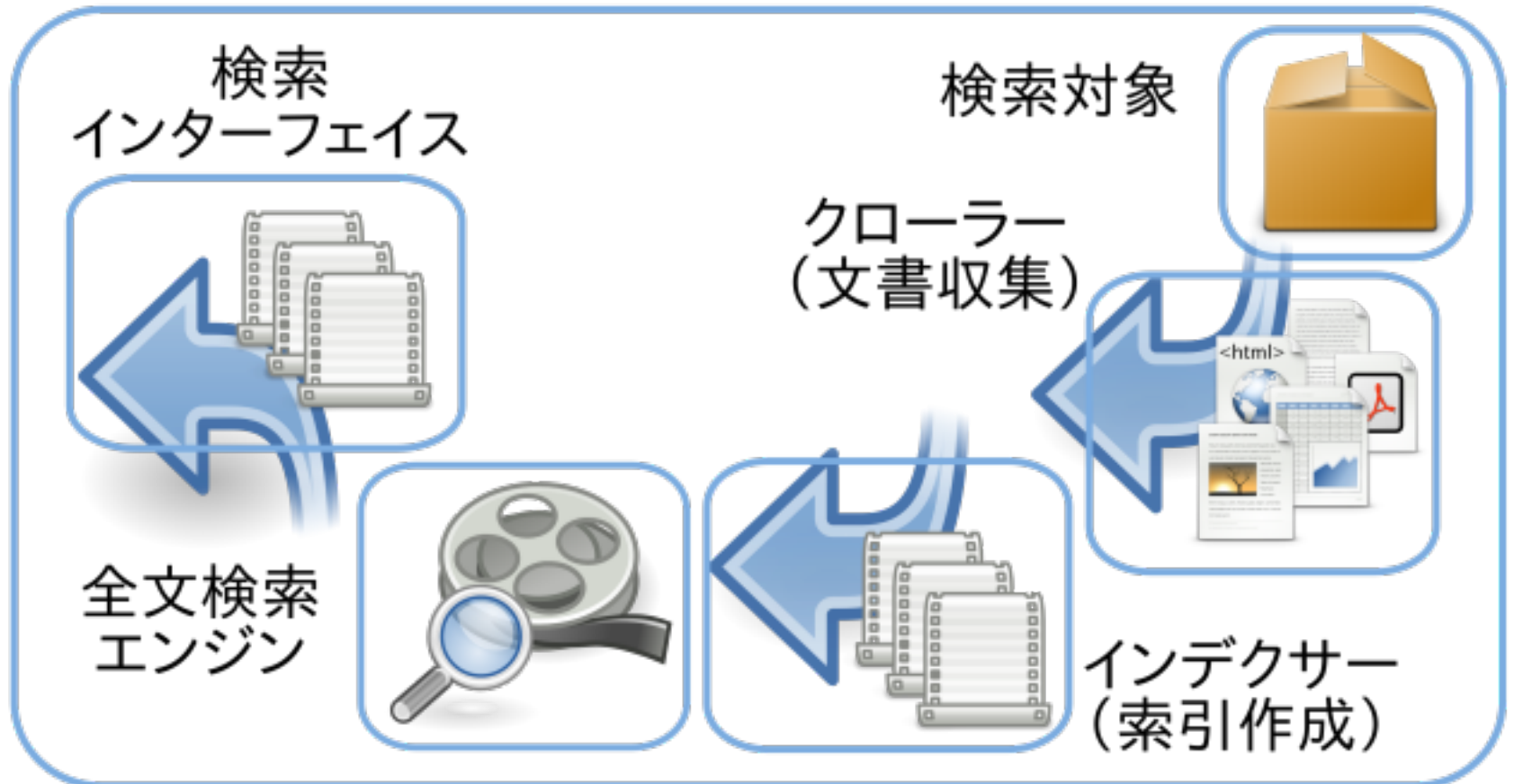
るりまサーチの作り方

(how to create rurema-search)

- ✓ 全文検索システムの作り方
how to create full-text search system
- ✓ るりまサーチの副産物
rurema-search side effects

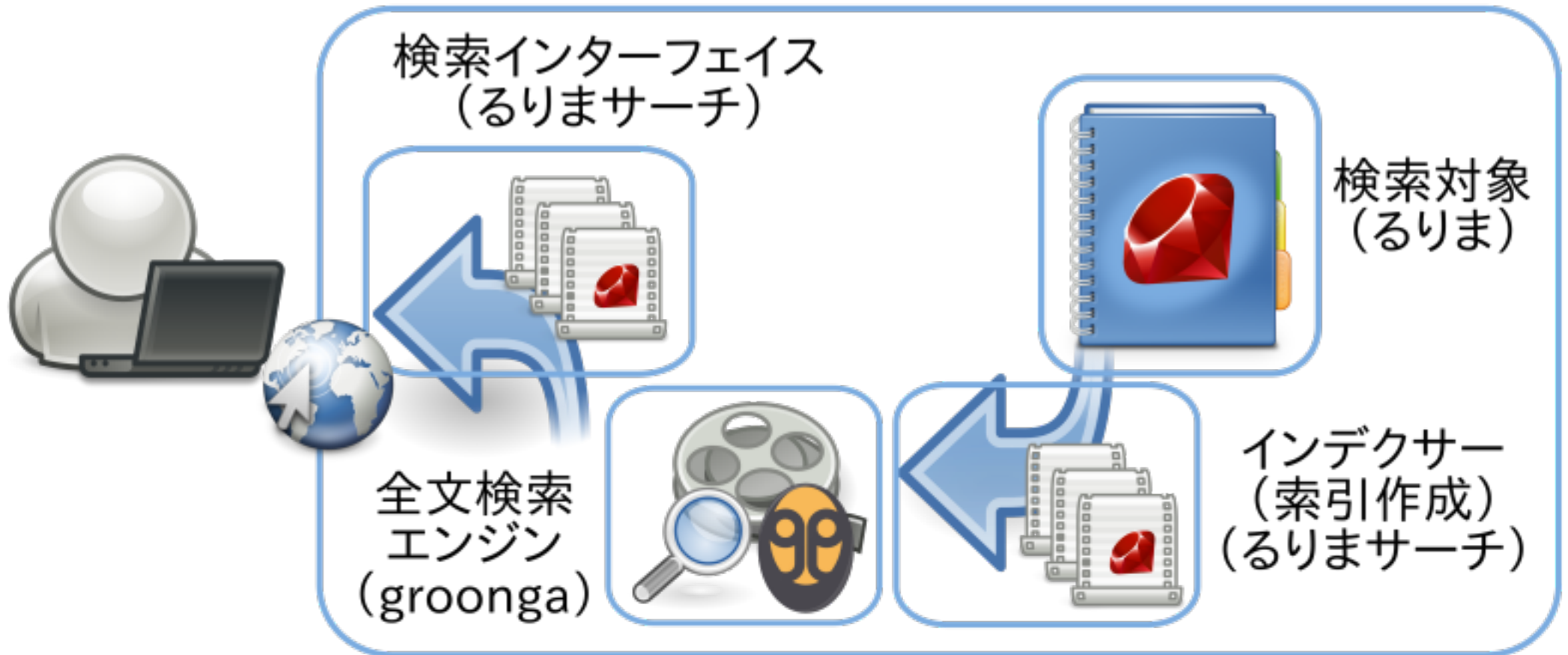
全文検索システム

全文検索システム



るりまサーチの場合

るりまサーチシステム



全文検索エンジン

るりまサーチシステム

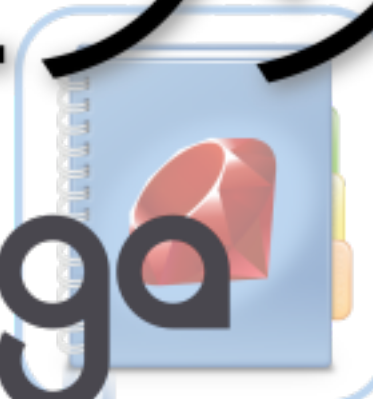
全文検索エンジン



全文検索
エンジン
(groonga)



インデクサー
(索引作成)
(るりまサーチ)



検索対象
(るりま)

groonga



✓ Sennaの後継
The successor to Senna

✓ <http://groonga.org/>

groonga 1.0!!!



2010/08/29

groonga 1.0

リリース!

Senna: 特徴1

- ✓ 転置索引ライブラリ

an inverted index library

- ✓ 完全転置索引

full inverted index

- ✓ データストアなし

no data store

- ✓ 外部データストアと連携 (例: RDB)

work with an external data store (e.g.: RDB)



Senna: 特徴2

✓ MySQL連携
work with MySQL

✓ Tritonn
Tritonn

✓ リアルタイム検索
real-time search

✓ 高速でロックフリーな更新
fast and read lock free update



Senna: 問題点



- ✓ RDB側の制限が性能に影響
impedance mismatch
- ✓ 更新時に参照ロック
read block while update
- ✓ 互換性の維持
keeping compatibility
- ✓ 性能改善の足かせ
block performance improvements



Senna → groonga: 維持

よいところはそのまま

keep good points

✓ 完全転置索引

full inverted index

✓ MySQL連携

work with MySQL (mroonga)

✓ リアルタイム検索

real-time search

✓ 高速でロックフリーな更新

fast and read lock free update



Senna → groonga: 発展

Sennaでの限界を突破

break through limitation for Senna

✓ 独自データストア

original data store

✓ カラム指向

column oriented

✓ 性能改善

higher performance

✓ 互換性なし

incompatible



groonga: 特徴



- ✓ 完全転置索引+データストア
full inverted index and data store
- ✓ カラム指向: 高速なドリルダウン
column oriented: fast drill down
- ✓ 大規模文書対応
large scale documents support
- ✓ ライブラリ+サーバ
library and server



groonga: 追加機能

✓ 位置情報対応

geographical location support

✓ 索引を用いた高速な検索・ソート

fast search and sort by index

✓ ネットワーク対応

network support



✓ HTTP, memcached, 独自プロトコル

HTTP, memcached and original protocol

✓ Web管理画面

Web administration interface

Sennaとgroongaの違い

項目	Senna	groonga
データ ストア	×	○
MySQL 連携	Tritonn 組み込み	mroonga プラグイ ン
Ruby 連携	○	○ rroonga

groongaの使い方



1. スキーマ定義

schema definition

2. データ登録

data registration

3. 検索！

search!



スキーマ



✓ テーブル
table

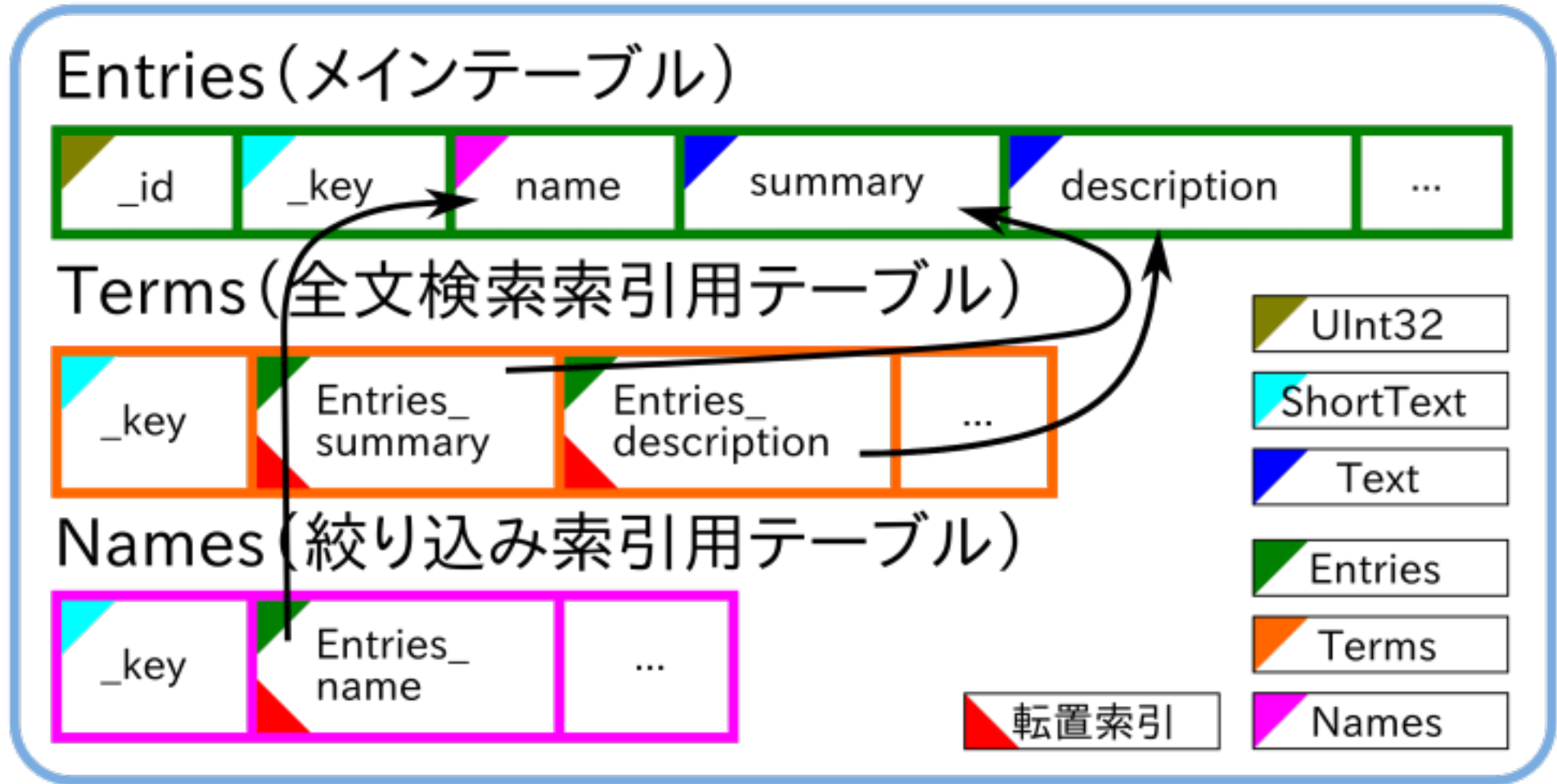
✓ カラム
column

✓ 型
type



定義例: るりまサーチ

スキーマ



テーブル



✓ キーあり: ハッシュ・ツリー
with key: hash and tree

✓ キーには型あり
key has a type



✓ キーなし: 配列
without key: array

カラム

✓ データカラム

data column

✓ 型あり

column has a type

✓ 索引カラム

index column

✓ 検索・ソート用

for search and sort



型

基本型 (basic type)

Text, Int32, Float, Time, GeoPoint, ...

参照型 (reference type)

テーブル、カラム
table, column

✓ ドリルダウン・索引用
for drill down and index



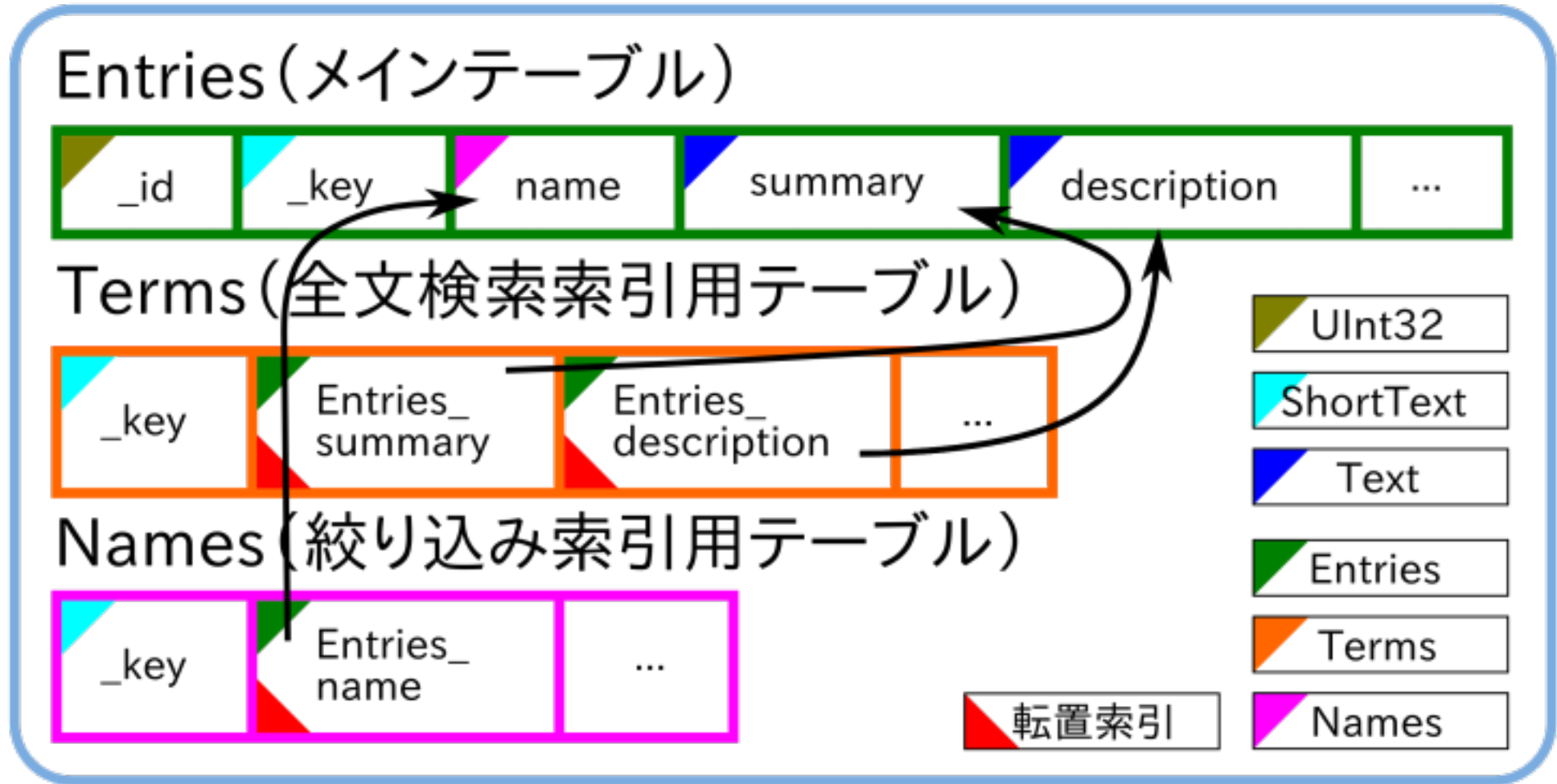
索引カラム

- ✓ 転置索引
inverted index
- ✓ 自動更新
auto update
- ✓ 定義するだけでOK
just need to be defined
- ✓ 参照型を使用
use reference type



定義例: るりまサーチ

スキーマ



スキーマ定義



1. 組み込みのDDL
built-in DDL



2. rroongaのDSL
rroonga's DSL



組み込みDDL



```
table_create Entries ...  
column_create Entries name ...  
...
```

rroongaのDSL



```
Groonga::Schema.define do |schema|
  options = {type: :hash,
             key_type: "ShortText"}
  schema.create_table("Entries", options) do |table|
    table.short_text("name")
    # ...
  end
  # ...
end
```

データ登録



1. ~~スキーマ定義~~

schema definition

2. データ登録

data registration



3. 検索！

search!

登録処理: るりまサーチ

るりまサーチシステム



登録例: るりまサーチ

Entries (メインテーブル)

_id	_key	name	summary	description	...
1	String#sub	sub	置換	1つ置換	
2	String#gsub	gsub	置換	全部置換	

Names (絞り込み索引用)

_key	Entries_name	...
sub	1	
gsub	2	

Terms (全文検索索引用)

_key	Entries_summary	Entries_description	...
置換	1, 2	1, 2	
1つ		1	
つ置		1	
...			

データの登録方法



1. 組み込みのloadコマンド

built-in load command

✓ フォーマット: JSON

format: JSON



2. Rubyスクリプト

Ruby script

データ登録: load



```
load --table Entries
[
  ["_key", "name", "description"],
  ["String#sub", "sub", "1つ置換"],
  ...
]
```


データ登録: rroonga



```
entries = Groonga["Entries"]
entries.add("String#sub",
           name: "sub",
           description: "1つ置換")
```

...

登録例: るりまサーチ

Entries (メインテーブル)

_id	_key	name	summary	description	...
1	String#sub	sub	置換	1つ置換	
2	String#gsub	gsub	置換	全部置換	

Names (絞り込み索引用)

key	Entries name	...
sub	1	
gsub	2	

Terms (全文検索索引用)

key	Entries summary	Entries_ description	...
置換	1, 2	1, 2	
1つ		1	
つ置		1	
...			

検索



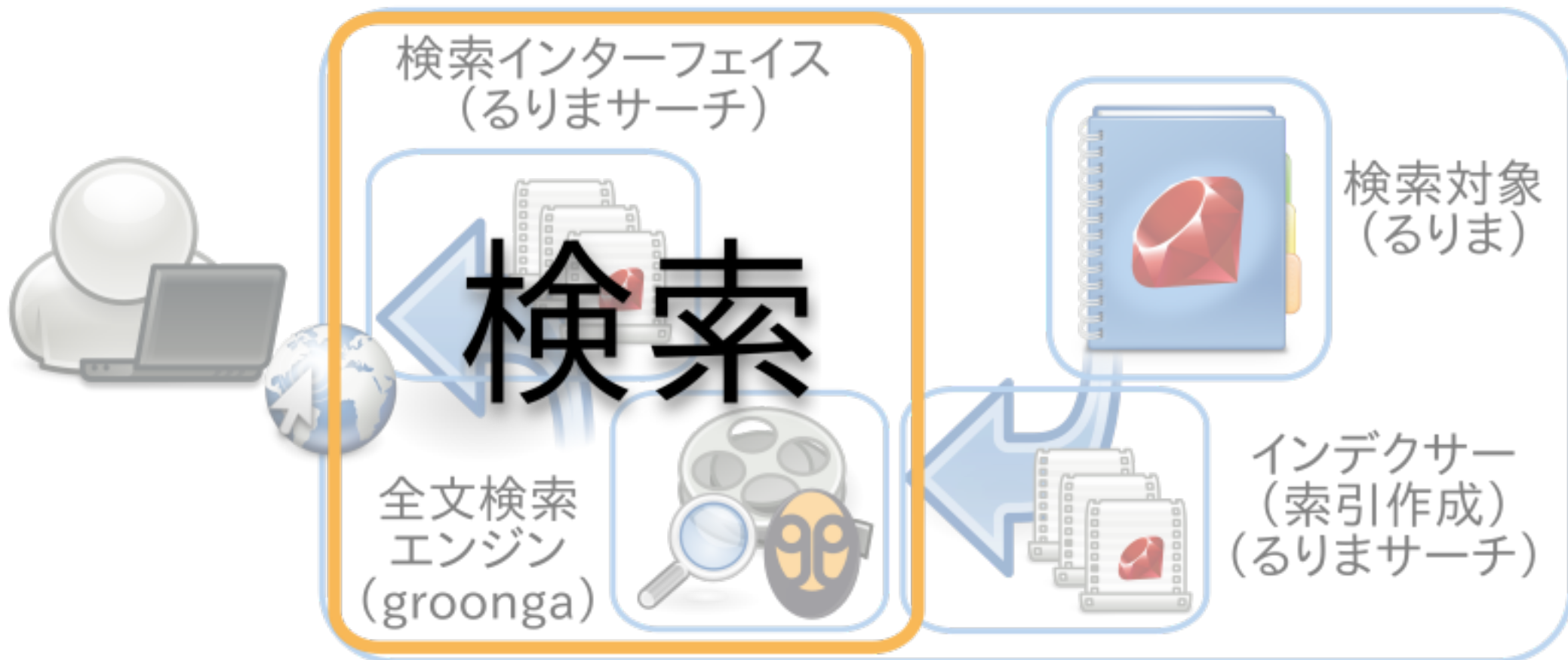
1. ~~スキーマ定義~~
schema definition
2. ~~データ登録~~
data registration
3. **検索!**
search!



検索処理: るりまサーチ



るりまサーチシステム



検索例: るりまサーチ

Entries (メインテーブル)

_id	_key	name	summary	description	...
1	String#sub	sub	置換	1つ置換	
2	String#gsub	gsub	置換	全部置換	

Names (絞り込み索引用)

_key	Entries_name	...
sub	1	
gsub		

Terms (全文検索索引用)

_key	Entries_summary	Entries_description	...
置換	1, 2	1, 2	
1つ		1	
つ置		1	
...			



検索方法

1. 組み込みのselectコマンド

built-in select command

✓ フォーマット: JSON

format: JSON



2. Rubyスクリプト

Ruby script

検索: select: gqtp



```
select Entries description "1つ"  
[[...],  
  [[[24], [...]],  
  ["String#sub", "sub", "1つ置換", ...],  
  ...]]
```

検索: select: HTTP



```
# 本当はURLは全部つなげてエスケープもする
% wget -O - 'http://localhost:10041/d/select?
           'table=Entries&
           'match_columns=description&
           'query=1つ'

[[[...],
  [[[24], [...]],
   ["String#sub", "sub", "1つ置換", ...],
   ...]]
```


検索: rroonga



```
entries = Groonga["Entries"]
result = entries.select do |record|
  record.description =~ "1つ"
end
result.each do |record|
  p [record.name, record.description]
  # => ["sub", "1つ置換"]
end
```

検索例: るりまサーチ

Entries (メインテーブル)

_id	_key	name	summary	description	...
1	String#sub	sub	置換	1つ置換	
2	String#gsub	gsub	置換	全部置換	

Names (絞り込み索引用)

_key	Entries_name	...
sub	1	
gsub		

Terms (全文検索索引用)

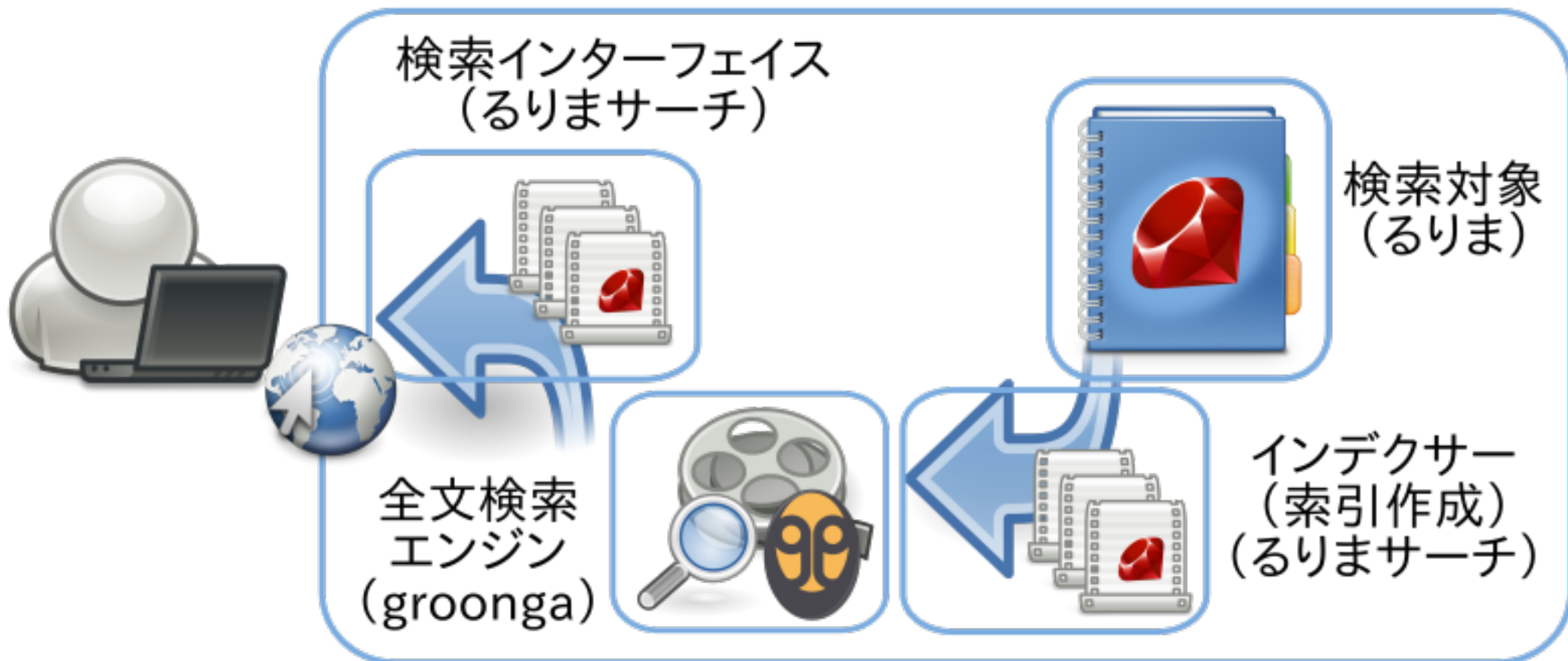
_key	Entries_summary	Entries_description	...
置換	1, 2	1, 2	
1つ		1	
つ置		1	
...			



description

るりまサーチ

るりまサーチシステム





るりまサーチの作り方

るりまサーチの使い方

(how to use rurema-search)

- ✓ ~~るりまサーチの紹介~~
rurema-search introduction

るりまサーチの作り方

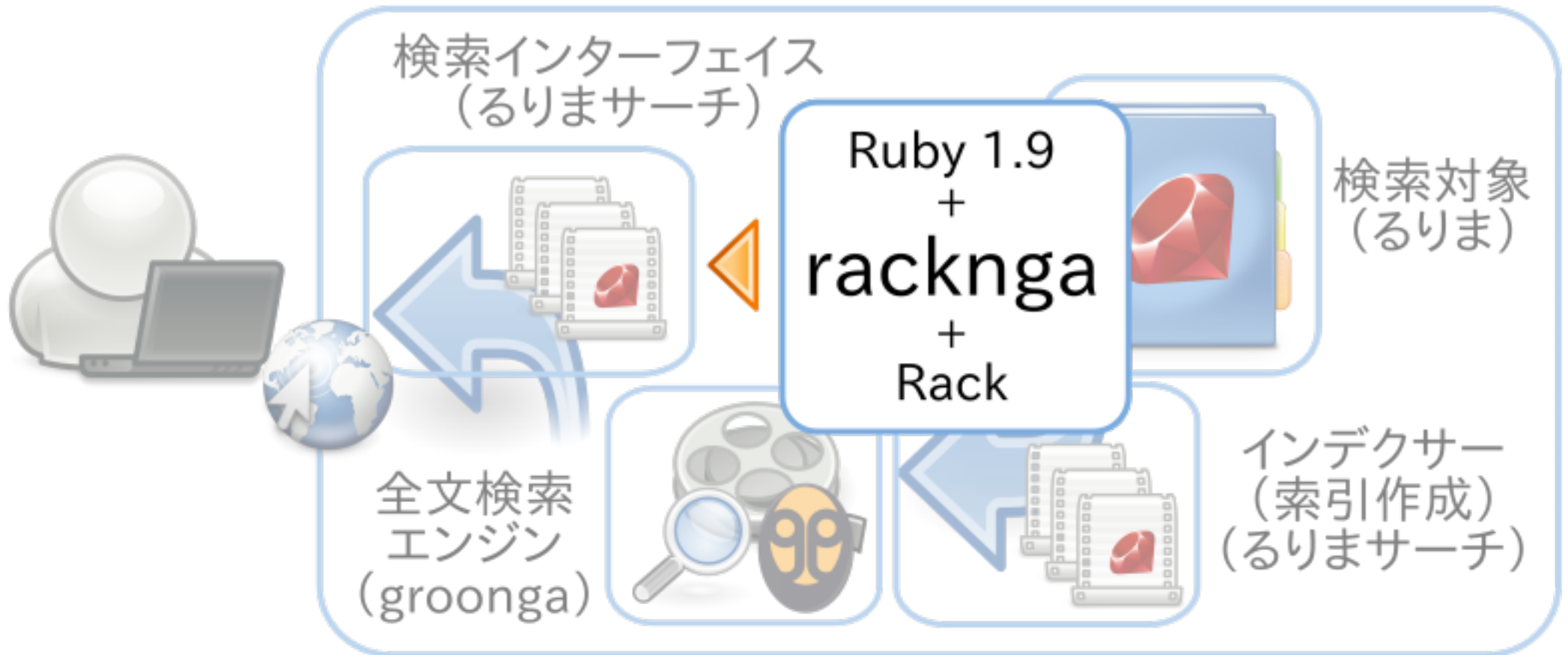
(how to create rurema-search)

- ✓ ~~全文検索システムの作り方~~
how to create full-text search system
- ✓ るりまサーチの副産物
rurema-search side effects

るりまサーチの副産物



るりまサーチシステム



racknga

全文検索システム用 ユーティリティ集

- ✓ Rackミドルウェア

Rack middlewares

- ✓ Muninプラグイン

Munin plugins

エラー通知

Webアプリケーション
(Ruby 1.9 + Rack) + racknga



キヤツシユ

Webアプリケーション
(Ruby 1.9 + Rack) + racknga



条件付き圧縮

Webアプリケーション
(Ruby 1.9 + Rack)

+ racknga



条件付き
圧縮



キャッシュ



JSONP

Webアプリケーション
(Ruby 1.9 + Rack)

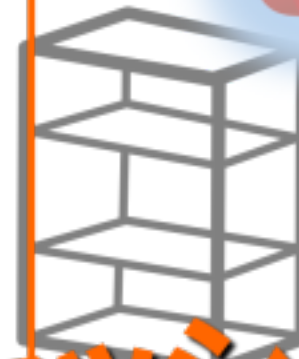
+ racknga



JSONP



キャッシュ



rack
powers web applications

Range

Webアプリケーション
(Ruby 1.9 + Rack) + racknga



**Range
If-Range
(for HTML5 video)**



Muninプラグイン

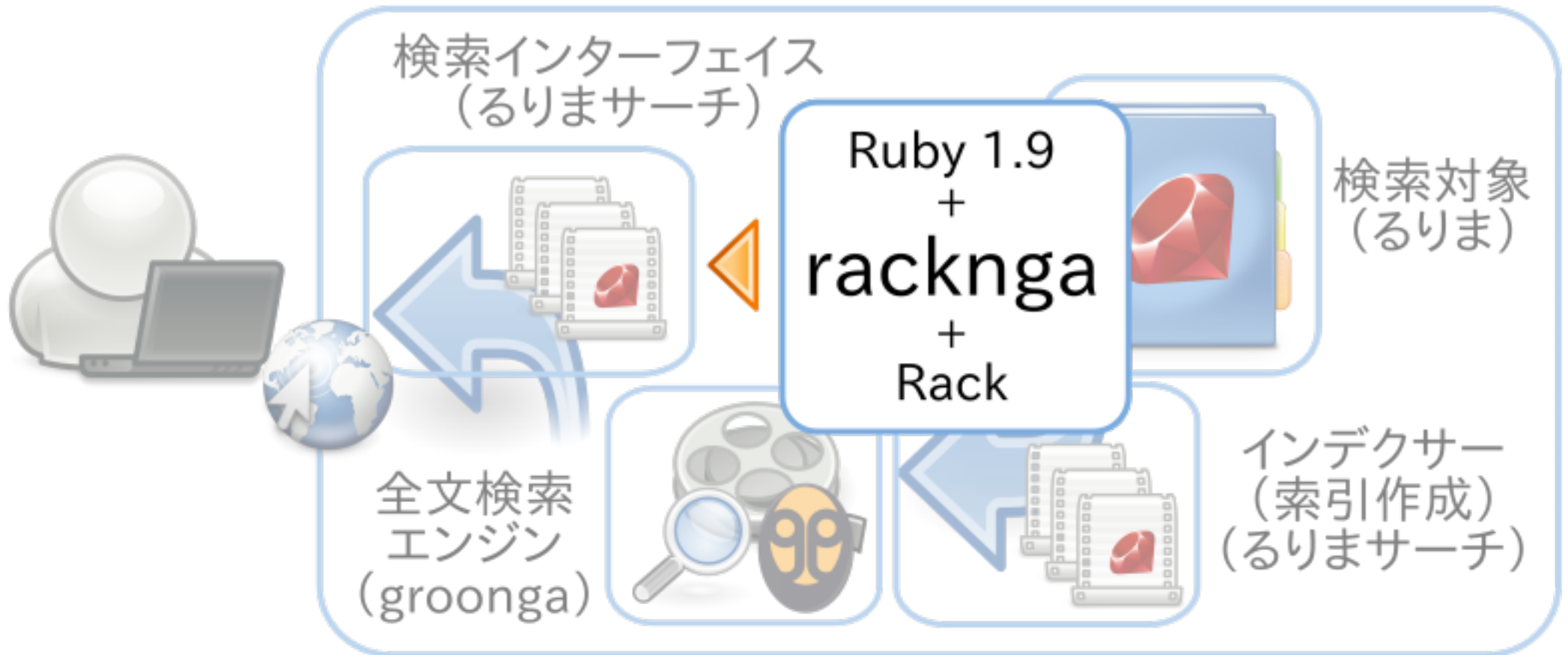


- ✓ Passenger用
for Passenger
- ✓ 処理したリクエスト数
number of processed requests
- ✓ 処理中のリクエスト
number of processing requests
- ✓ プロセスの状態
process statuses
- ✓ プロセスの起動時間
process uptimes

るりまサーチの副産物



るりまサーチシステム



まとめとお知らせ



✓ まとめ

- ✓ るりまサーチが提供すること
- ✓ るりまサーチが示していること
- ✓ るりまサーチの副産物

✓ お知らせ

- ✓ クリアコードのこと

まとめ

- ✓ るりまがもっと便利に
 - ✓ ストレスなく目的のエントリへ
- ✓ 全文検索エンジンはgroonga!
 - ✓ Rubyとの高い親和性: rroonga
- ✓ Ruby 1.9 + Rackをもっと便利に
 - ✓ 汎用ユーティリティ集: racknga

お知らせ



開発者募集中

応募条件

プログラミングが好きなこと
受付にパンフレットあり