



# Ruby



# groonga

須藤功平

株式会社クリアコード

全文検索エンジンgroongaを囲むタベ #1  
2010/11/29



祝！  
1.0.4リリース！





祝！  
1.0.4リリース！



groonga



# 話すこと

---

## Rubyと groongaの 相性のよさ



# ターゲット

## 開発者

- ★ システム構築者・管理者
- ★ groongaを直接利用
  - ★ プログラムを書く
  - ★ APIを叩く



# 自己紹介

★ 須藤功平（すとうこうへい）

★ 趣味

★ プログラミング

★ たいやき屋・和菓子屋をはしご



★ 仕事：趣味が仕事になった感じ



# 自己紹介：趣味

フリー  
ソフトウェア  
開発



# メインで開発

続けることが大事

- ★ RSS Parser: 6年 (Ruby標準添付)
- ★ Rabbit: 6年
- ★ Cutter: 5年
- ★ milter manager: 2年





# 参加プロジェクト

通った後に道ができる

- ★ バグを見つけたらパッチ投稿
  - ★ 広くこまごまと
- ★ GNOME関係/Ruby on Rails
- ★ Mozilla/Postfix/Samba/...



# 自己紹介： 仕事

クリアコード代表取締役



# クリアコード

## フリーソフトウェアでビジネス

- ★ フリーソフトウェアを活用した受託開発
- ★ フリーソフトウェアの導入支援
- ★ フリーソフトウェアのサポート
- ★ ...



# 資本提携

## 協業体制強化

- ★ 有限会社未来検索ブラジル
- ★ ミラクル・リナックス株式会社
- ★ コスモエア株式会社



# 協業

## groonga関連

- ★ 検索アプリ開発
- ★ 単体テストまわり整備
  - ★ Sennaの頃から（Cutterを利用）
- ★ バグ修正
- ★ Rubyまわり

# Ruby



# groonga



# Ruby ♥ groonga

- ★ ラングバプロジェクト
- ★ Rubyとgroongaのこと
- ★ Rubyと全文検索システムのこと



# ラングバプロジェクト

<http://groonga.rubyforge.org/>

 **rroonga** Rubyでgroongaを使って全文検索  
ラングバ

リファレンスマニュアル チュートリアル インストール 開発者向け情報

## ラングバプロジェクト

ラングバプロジェクトは**カスタム機能も備える高速・高機能な全文検索エンジンgroonga**を用いた全文検索システムを提供するプロジェクトです。groongaの機能をRubyから利用するためのライブラリ、様々な文書から検索対象のテキストを抽出するユーティリティ、検索用Webインターフェイスなどを提供します。提供する機能の一覧は以下の通りです。

- **rroonga**: groongaの機能をRubyから利用するためのライブラリ
- **ActiveGroonga**: rroongaをWebアプリケーションから使いやすくなるためのライブラリ
- **racknga**: 検索用Webアプリケーションの構築・運用を支援するユーティリティ
- **ChupaText**: テキスト抽出ユーティリティ
- **ChupaRuby**: ChupaTextの機能をRubyから利用するためのライブラリ
- **文書検索ラングバ**: 全文検索システム

[プロジェクトページ](#)  
[メーリングリスト](#)  
(groongaのメーリングリストをお願いします)

 **rroonga**





# 目標

使いやすい  
検索システムを  
早く、簡単に



# 提供物

- ★ 全文検索システム
  - ★ 全文検索システム利用を支援
- ★ ↑のコンポーネント
  - ★ ライブラリ・ツール
  - ★ 独自の全文検索システム開発を支援

提供物  
名前確認



# 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



検索対象



クローラー  
(文書収集)



インデクサー  
(索引作成)





# groonga

## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga

検索対象



クローラー  
(文書収集)



インデクサー  
(索引作成)





# rroonga

## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga  
rroonga

検索対象



クローラー  
(文書収集)



インデクサー  
(索引作成)



# ChupaText

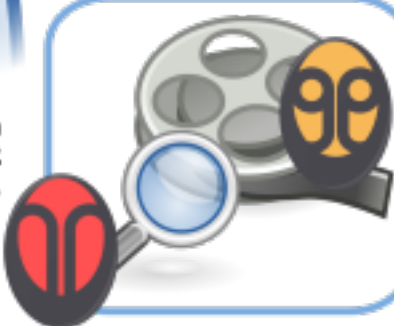
## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga  
rroonga

検索対象



クローラー  
(文書収集)

テキスト抽出  
ChupaText

インデクサー  
(索引作成)





# ChupaRuby

## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga  
rrroonga

検索対象



クローラー  
(文書収集)

テキスト抽出  
ChupaText



ChupaRuby

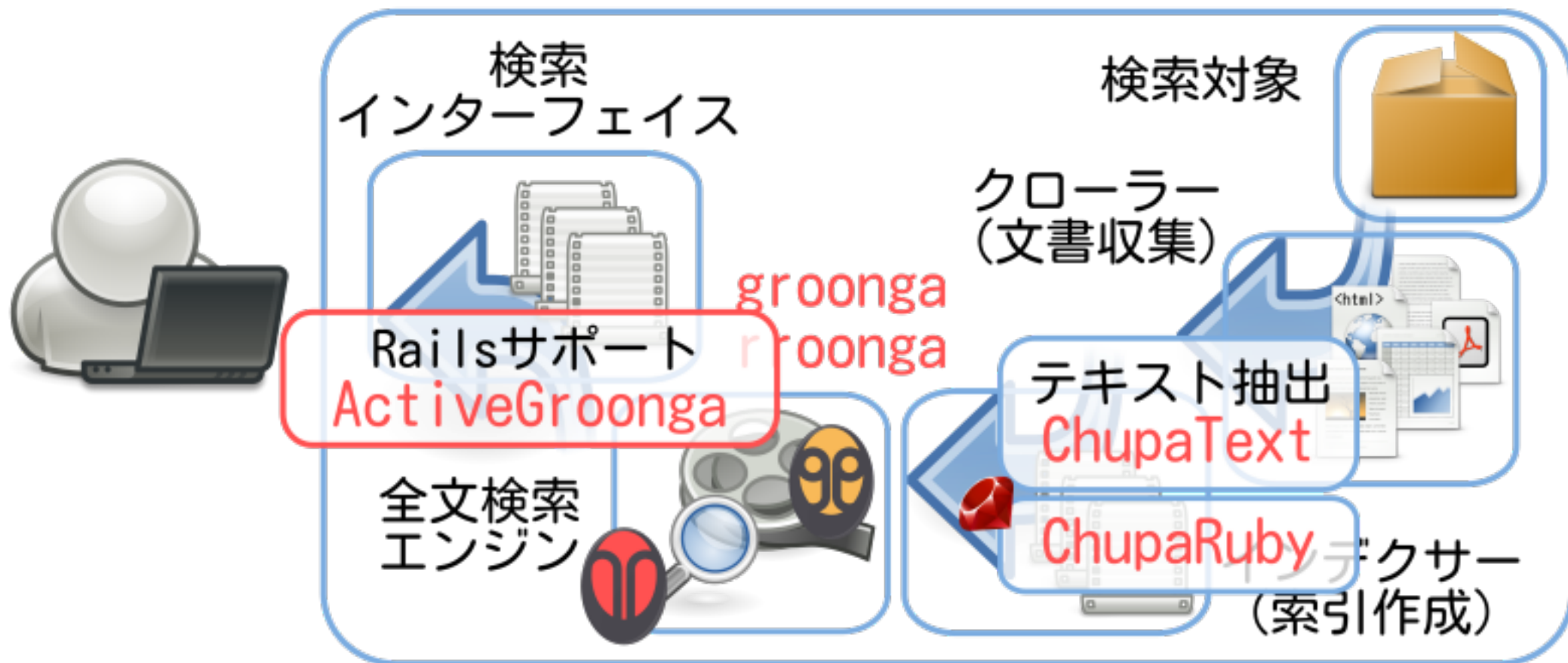
インデクサー  
(索引作成)





# ActiveGroonga

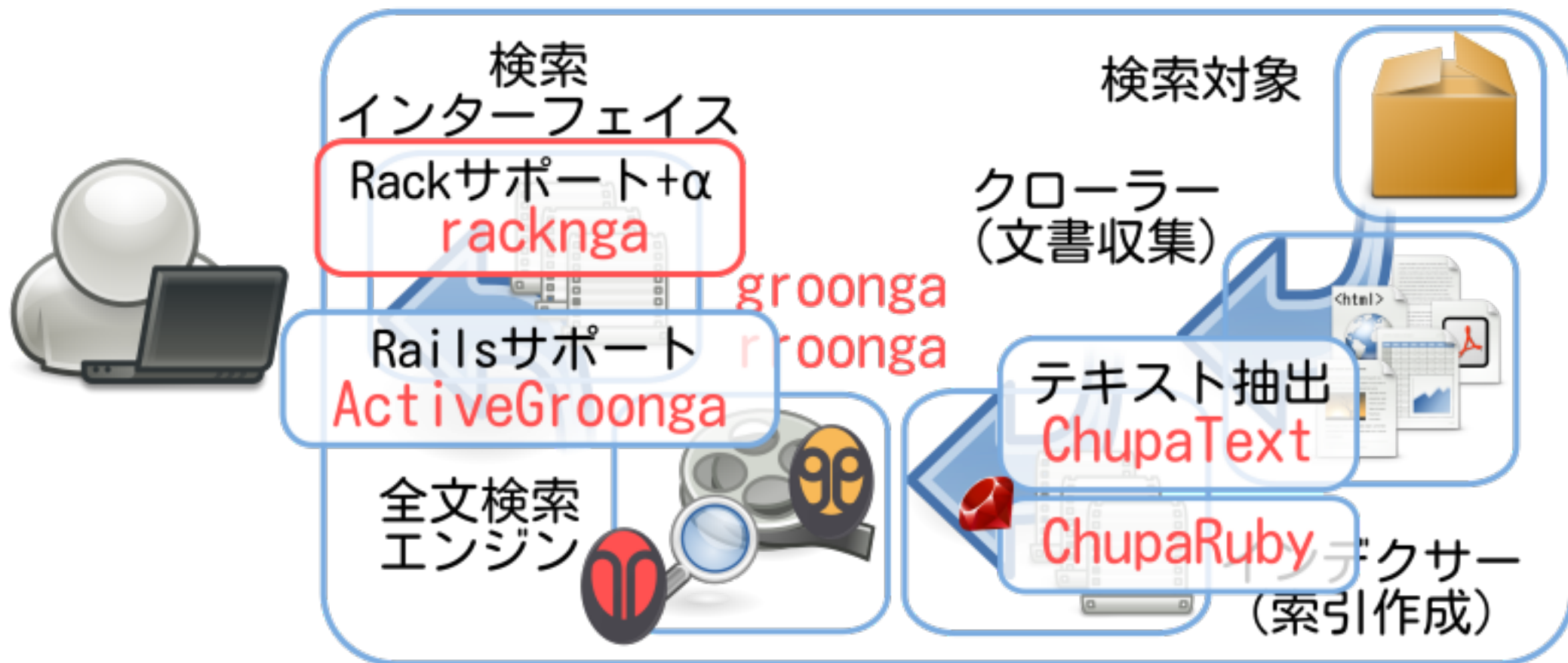
## 全文検索システム





# racknga

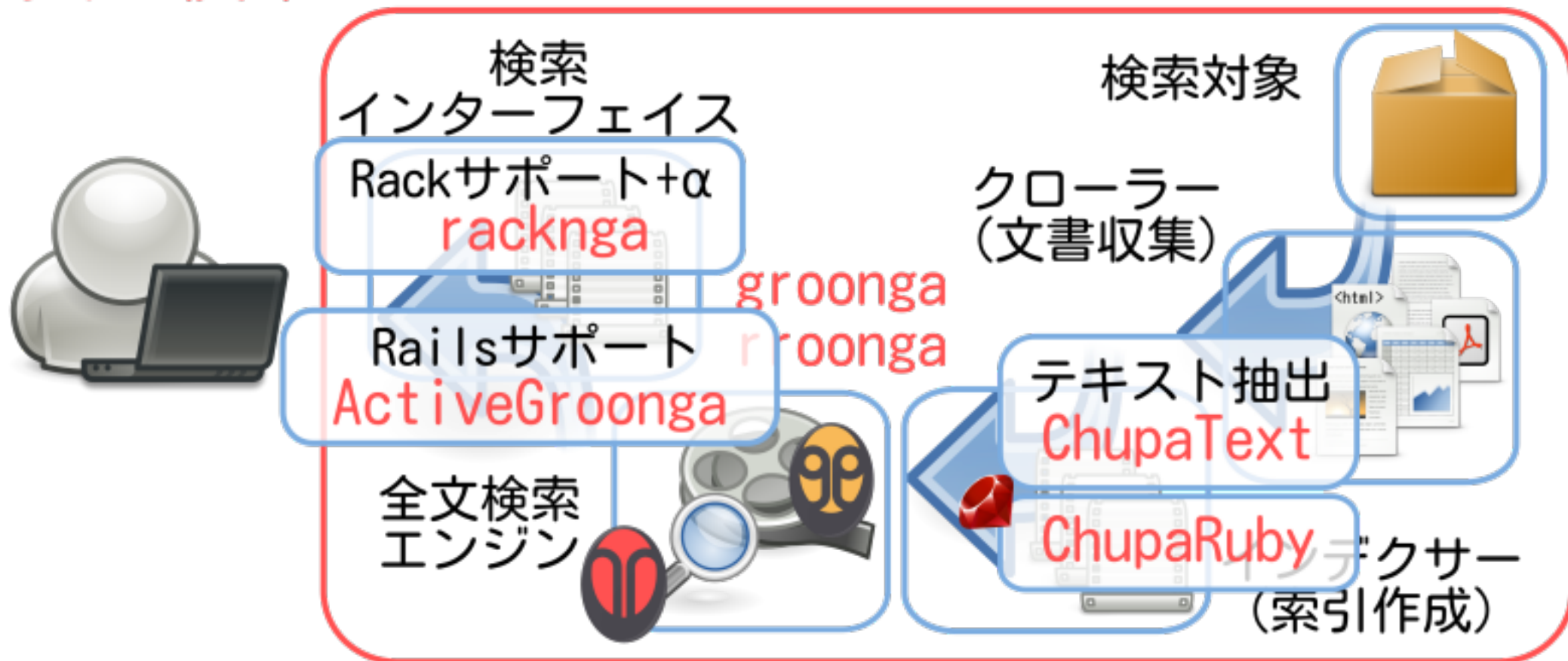
## 全文検索システム





# 文書検索ラングバ

## 文書検索ラングバ全文検索システム



# 復習：全文検索エンジン

groonga (ぐるんが)

全文検索エンジン

rroonga (るるんが)

groongaのRubyバインディング



# 復習：テキスト抽出

ChupaText (ちゅぱてきすと)

汎用テキスト抽出ツール

ChupaRuby (ちゅぱるびー)

ChupaTextのRubyバインディング



# 復習： 検索UI

ActiveGroonga  
(あくていぶぐるんが)

---

- ★ Rails 3用groongaモデル

racknga (らくんが)

---

- ★ Rackミドルウェア集ほか

- ★ Passenger用Muninプラグインなど



# 復習： 検索アプリ

文書検索ラングバ  
(ぶんしょけんさくらんぐば)

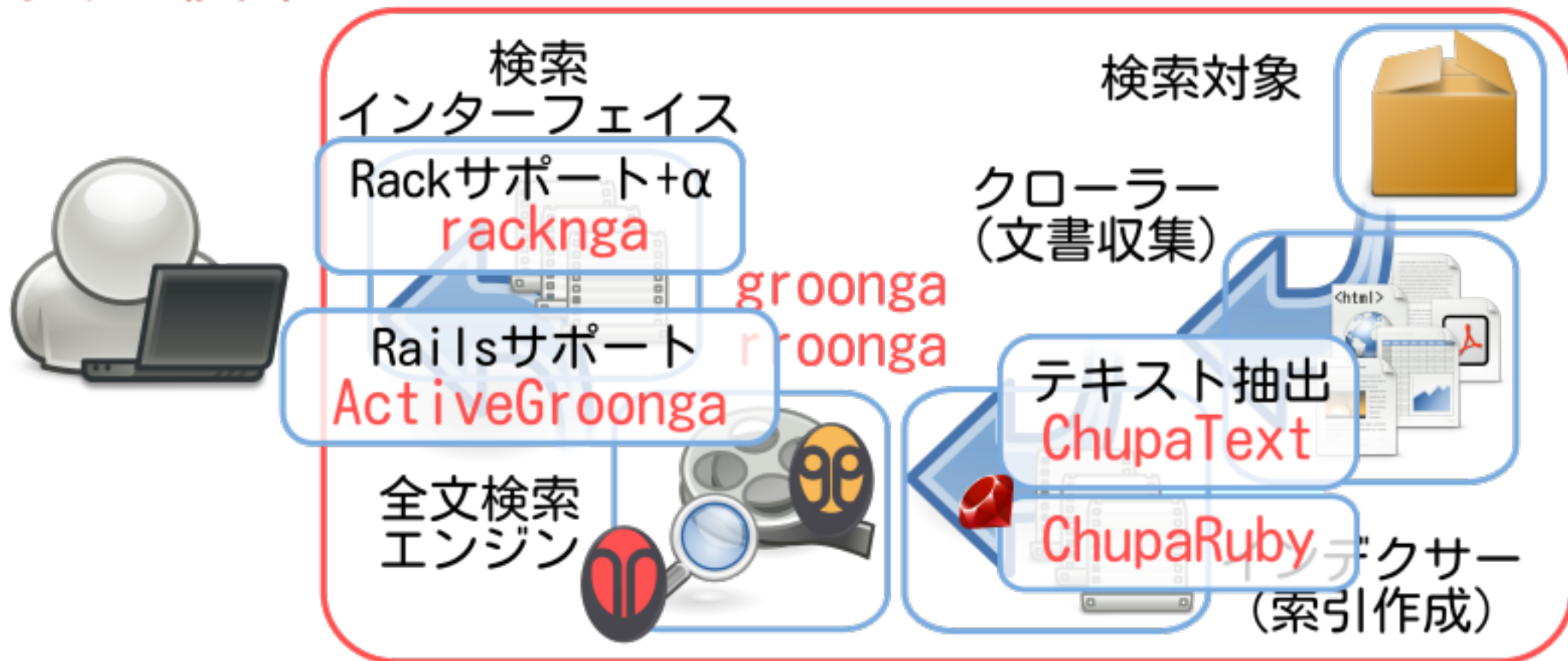
---

groongaベースの文書検索システム



# 復習：全体図

## 文書検索ラングバ全文検索システム





提供物

紹介



# rroongaの紹介

## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga  
rroonga

検索対象



クローラー  
(文書収集)



インデクサー  
(索引作成)



# rroonga

- ★ groongaのRubyバインディング
- ★ groongaをライブラリとして使用
  - ★ groongaクライアント機能は**ほぼ**ない
- ★ Rubyらしく書けることを重視
- ★ 便利な機能は盛り込む  
(例: スキーマ定義DSL)



# groonga API レイヤー



QL-API

DB操作コマンド  
ネットワークアクセス

DB-API

データベース  
テーブル・カラム・型  
手続き

LOW-LEVEL  
API

ハッシュテーブル  
パトリシアトライ  
配列



# サポートしているAPI



QL-API

DB操作コマンド  
ネットワークアクセス

DB-API

データベース  
テーブル・カラム・型  
手続き

LOW-LEVEL  
API

ハッシュテーブル  
パトリシアトライ  
配列



# Rubyらしさ重視

- ★ スキーマ定義: DSL
- ★ カラムアクセス: メソッド
- ★ groonga ↔ Ruby変換: 型を利用
- ★ クエリ: ブロック



# スキーマ定義

## 高レベルスキーマ定義API (QL-API相当)

```
Groonga::Schema.define do |schema|  
  schema.create_table("Users") do |table|  
    table.short_text("name")  
    table.int32("age")  
    table.timestamps  
  end  
end
```



# カラムアクセス

## メソッドとしてアクセス可能

```
users = Groonga["Users"]
morita = users.add(:name => "morita",
                  :age => 18)
morita.name # => "morita"
morita.age = 88
morita.age # => 88
```





# groonga ↔ Ruby変換

groongaの型とRubyのクラスを対応

★ 自動で相互変換

```
morita.updated_at = Time.now  
morita.updated_at.class # => Time  
morita.name.class # => String  
morita.age.class # => Fixnum
```



# クエリ

名前にmoを含むユーザを全文検索

```
users.select do |record|  
  record.name =~ "mo"  
end
```



# クエリ：複雑

```
a_month_ago = Time.parse("2010-10-29")
query = "Ruby 検索"
sites.select do |record|
  condition = record.match(query) do |target|
    (target.title * 100) |
    (target.description * 10) |
    (target.content)
  end
  condition &
    (record.updated_at >= a_month_ago)
end
```



# クエリ：重み付け

```
record.match(query) do |target|  
  # titleにマッチ:      スコア+100  
  # descriptionにマッチ: スコア+10  
  # contentにマッチ:     スコア+1  
  (target.title * 100) |  
    (target.description * 10) |  
    (target.content)  
end
```



# クエリ: AND

```
a_month_ago = Time.parse("2010-10-29")
query = "Ruby 検索"
condition = record.match(query) {...}
# 「Ruby」と「検索」にマッチする
# 1ヶ月以内に更新されたレコードを検索
condition &
  (record.updated_at >= a_month_ago)
```



# インストール

```
% sudo gem install rroonga
```

必要ならgroongaも自動でインストール



# 利用シーン

- ★ インデクサー
  - ★ Rubyで加工しながらデータ追加
- ★ データベースメンテナンス
  - ★ バッチ処理
- ★ groongaの挙動確認 (irb -r groonga)
- ★ テストデータ作成



# 利用例

## るりまサーチ: rroonga + Rack

 るりまサーチ

 **すべて**  
65816件

[1.8.7](#)  
16620件

[1.8.8](#)  
16637件

[1.9.1](#)  
16194件

[1.9.2](#)  
16365件

最速Rubyリファレンスマニュアル検索！

検索

 **種類**

 [インスタンスメソッド](#) (33732)

[定数](#) (8977)

[シングルトンメソッド](#) (8539)

 [クラス](#) (5303)

 [関数](#) (2948)

 [モジュールファンクション](#) (2656)





# ChupaTextの紹介

## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga  
rroonga

検索対象



クローラー  
(文書収集)

テキスト抽出  
ChupaText

インデクサー  
(索引作成)





# ChupaText

テキスト抽出  
ツール



# ChupaText: 目標

複数の入力形式  
1つの操作方法



# インターフェイス

- ★ chupatextコマンド
  - ★ MIME形式で出力
- ★ C言語API



# どうしてMIME？

- ★ 広く普及している
- ★ メタデータとテキストを扱える
  - ★ メタデータ：ヘッダー（拡張可能）
  - ★ テキスト：本文
- ★ 複数のファイルを扱える
  - ★ アーカイブ対応



# 拡張性

多くの入力形式に対応したい

- ★ 入力処理部分をプラグイン化
  - ★ 動的に読み込み
- ★ プラグインの追加方法
  - ★ Cで共有ライブラリを作成
  - ★ Rubyスクリプトを作成



# インストール

```
% sudo aptitude install chupatext
```

<http://groonga.rubyforge.org/chupatext/ja/install.html>



# ChupaRubyの紹介

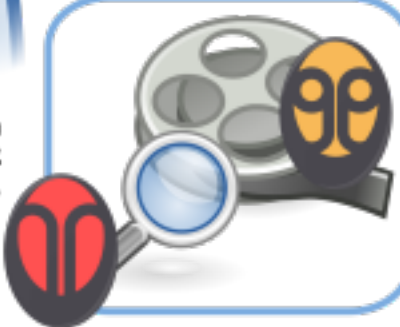
## 全文検索システム



検索  
インターフェイス



全文検索  
エンジン



groonga  
rroonga

検索対象



クローラー  
(文書収集)

テキスト抽出  
ChupaText



ChupaRuby

インデクサー  
(索引作成)





# ChupaRuby

- ★ ChupaTextのRubyバインディング
- ★ テキスト抽出ライブラリ
- ★ HTTPインターフェイス (予定)
  - ★ テキスト抽出Webサービス



# インストール

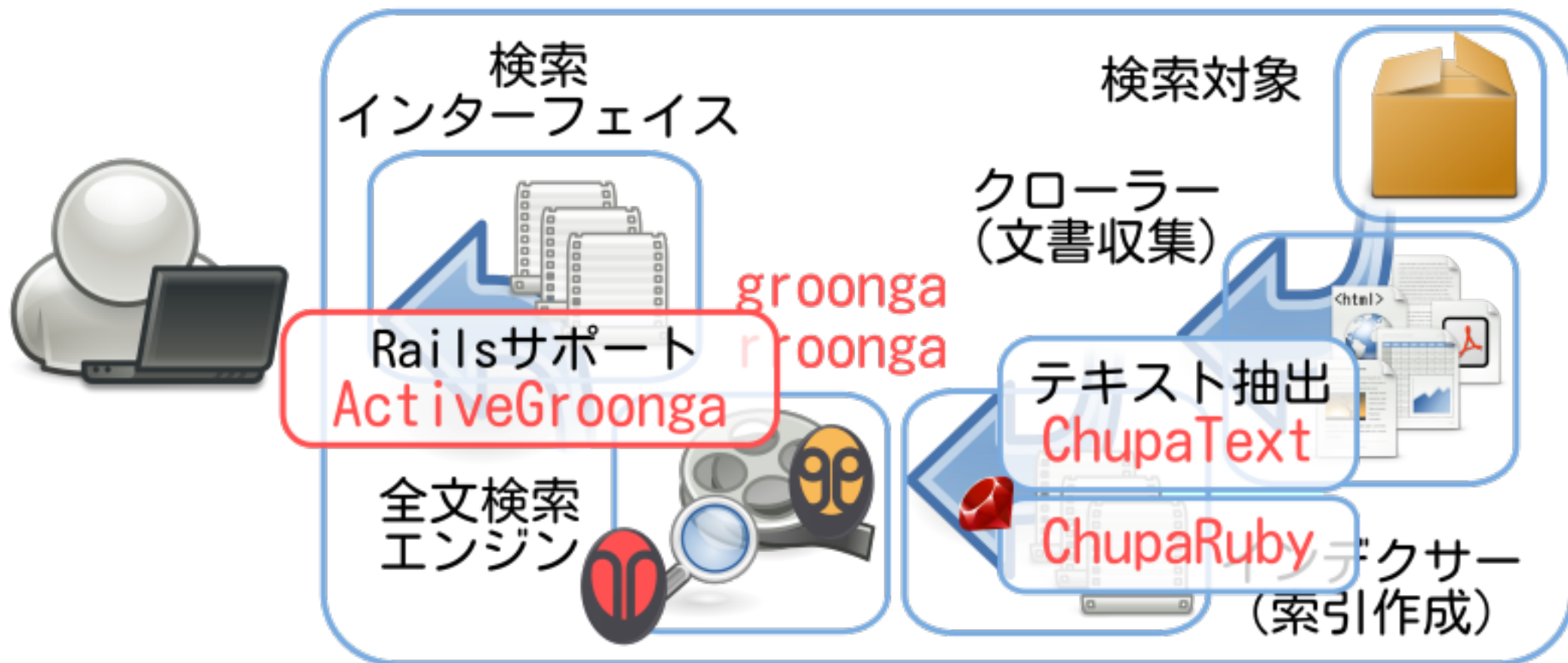
```
% sudo gem install chuparuby
```

ChupaTextは自動インストールされない



# ActiveGroongaの紹介

## 全文検索システム





# ActiveGroonga

- ★ Rails 3用groongaモデル
- ★ 実装: rroonga + ActiveRecord



# 機能

- ★ `ActiveGroonga::Base`  
(`ActiveRecord::Base`相当)
- ★ `rake groonga:*` (`db:*`相当)
- ★ マイグレーション
- ★ バリデーション
- ★ リレーション



# インストール

```
% sudo gem install activegroonga
```



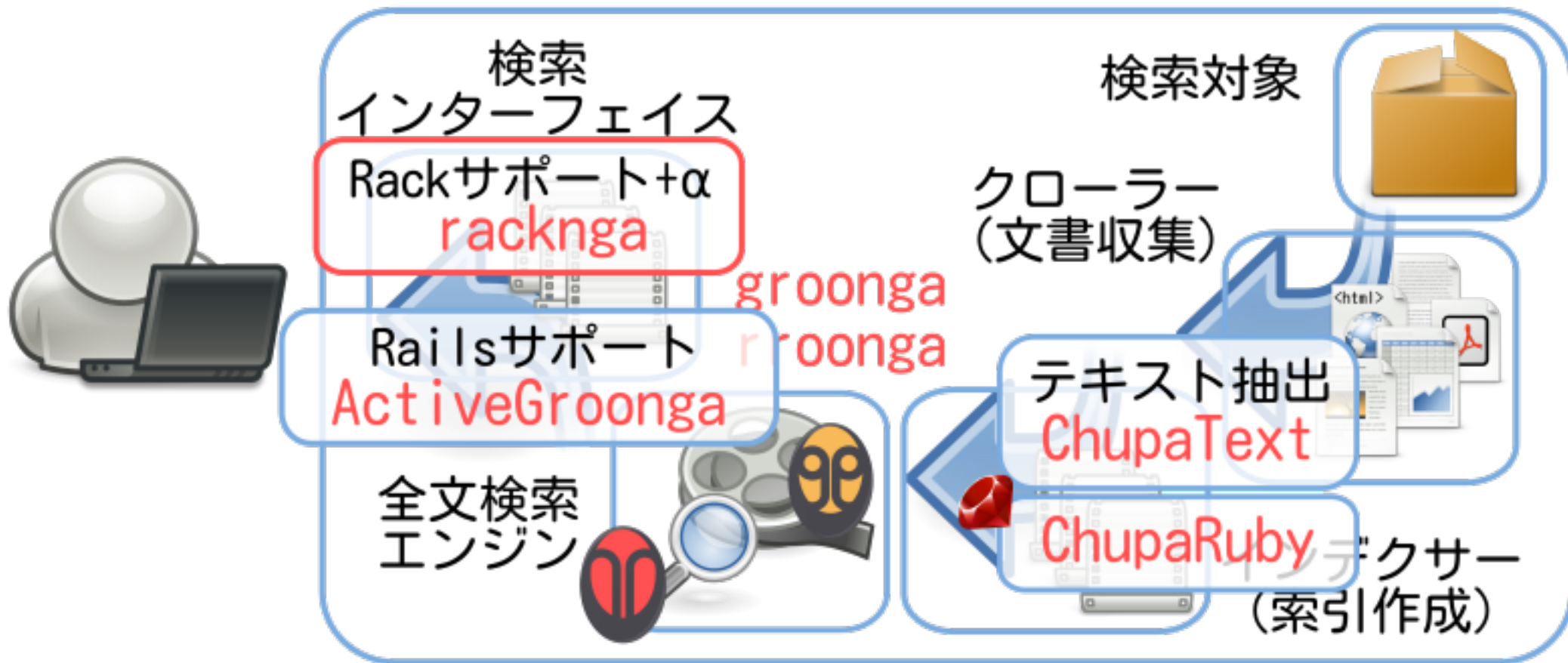
# デモ

# マイクロブログ



# rackngaの紹介

## 全文検索システム







# racknga

## ユーティリティパッケージ

(入れ場所に困ったらここ)

- ★ Rackミドルウェア

  - ★ キャッシュ・JSONP・...

- ★ Passenger用Muninプラグイン

- ★ will\_paginateサポート

- ★ ...

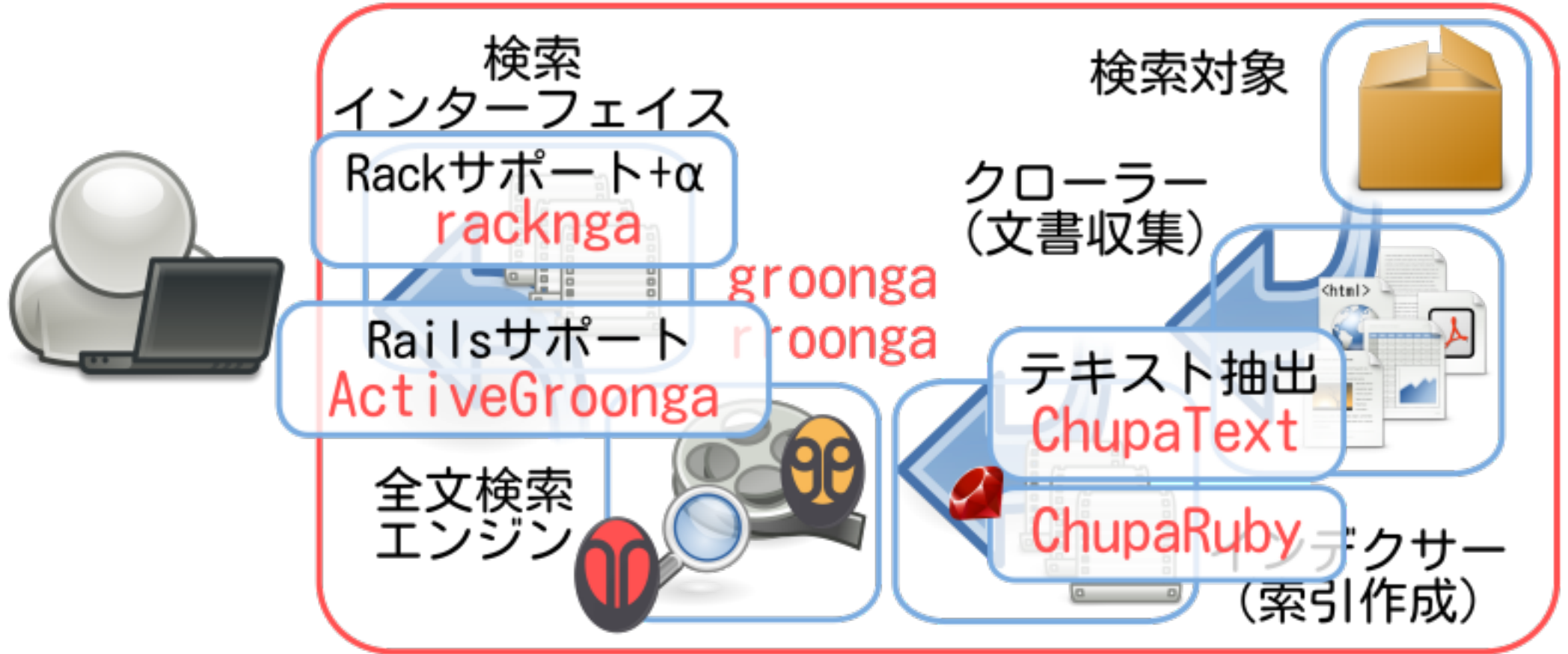


# インストール

```
% sudo gem install racknga
```

# 文書検索ラングバの紹介

## 文書検索ラングバ全文検索システム





# 文書検索ラングバ

- ★ 全文検索システム一式
  - ★ ホスト1台で運用できる規模向け
  - ★ groongaサーバ構成ではないため
- ★ groongaの特性を活かすUI
  - ★ ドリルダウン
- ★ リリース日: coming soon...



# まとめ

- ★ Rubyとgroongaは相性がいい
- ★ 関連ライブラリ・ツール色々
  - ★ rroonga・ActiveGroonga・...
- ★ 使ってみよう！



# お知らせ

# クリアコード

# 開発者募集中

応募条件  
プログラミングが好きなこと