

# rrroongaによる 検索サービスの 実装

須藤功平

株式会社クリアコード

モーショロジー2012 #1  
2012/1/26





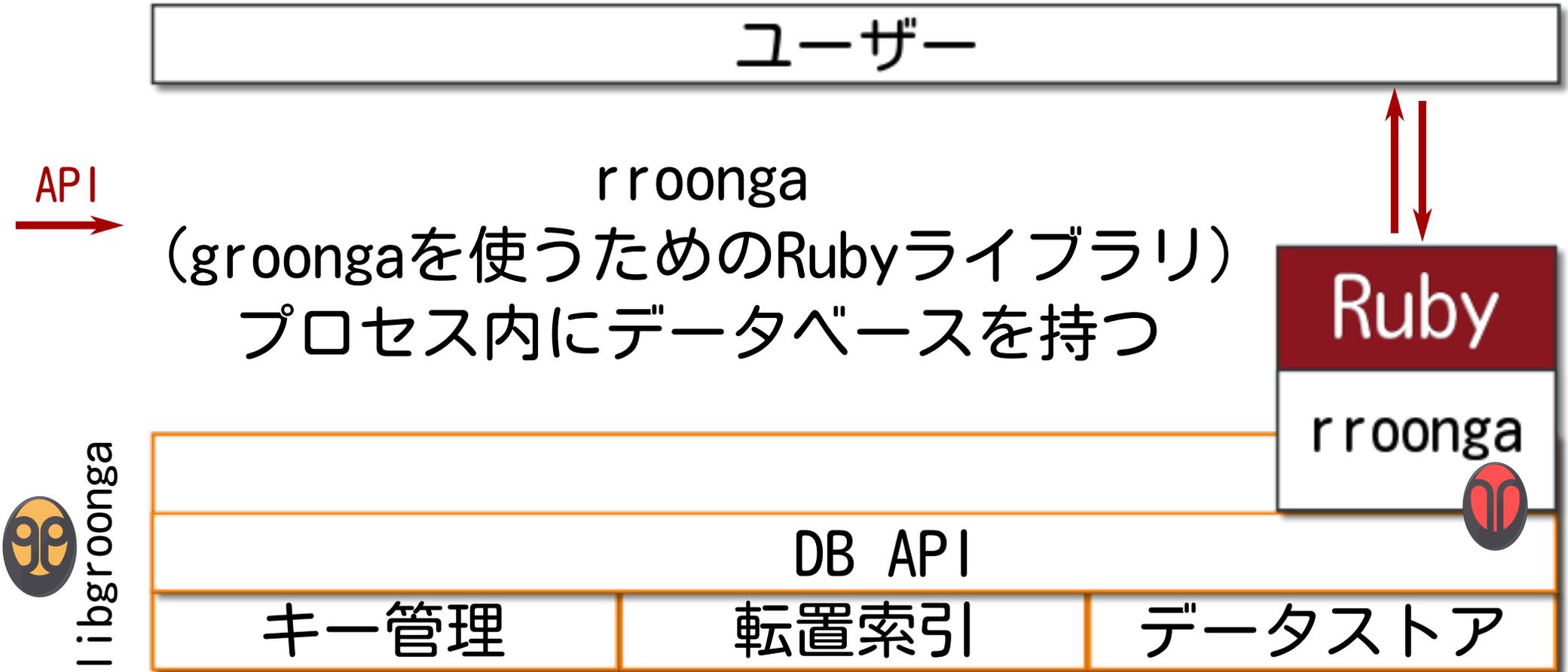
# 話すこと

## rroongaでの検索サービス開発事例

- ★ rroongaを使うメリット
- ★ rroongaを使うデメリット
- ★ 入力補完



# rroonga?





# システム構成



ユーザー



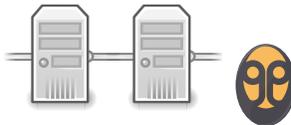
ロード  
 balancer



アプリケーション  
サーバー



(DB込み)



DBサーバー

DBサーバー  
スタイル



インプロセスDB  
スタイル



# メリット

rroongaでの検索サービス開発事例

- ★ rroongaを使う **メリット**
- ★ rroongaを使う **デメリット**
- ★ **入力補完**



# メリット

- ★ 細かい読み書きのコスト： **小**
  - ★ 通信コストがないため
- ★ 検索処理の自由度： **高**
  - ★ 基本演算を柔軟に組み合わせ可能



# 組み合わせ例

## 多段ドリルダウン

(注意: 「多段ドリルダウン」というのは一般的な語ではない)





# ドリルダウン

全データ

検索結果

ドリルダウン



さらに条件Aで  
絞り込むと？



条件Bでは？



さらに条件Xで  
絞り込んだ結果を  
先行して計算  
有効条件のみを提示



0件ヒットを回避



# 多段ドリルダウン

全データ

途中結果

最終結果



絞り込み1



絞り込み2



検索の  
途中結果でも  
ドリルダウン

ドリルダウン1



ドリルダウン2





# 用途

- ★ 「絞り込み→変更」 したい条件
  - ★ 他の値と比較したい
  - ★ 絞り込み→解除→再絞り込み:面倒
- ★ 例:
  - ★ 価格帯



# デメリット

rroongaでの検索サービス開発事例

- ★ rroongaを使うメリット
- ★ rroongaを使う**デメリット**
- ★ 入力補完



# デメリット

- ★ スケールアウトしない
  - ★ 元データをコピーして頑張る
  - ★ 独自にレプリケーションを実装



# 開発事例

## rroongaでの検索サービス開発事例

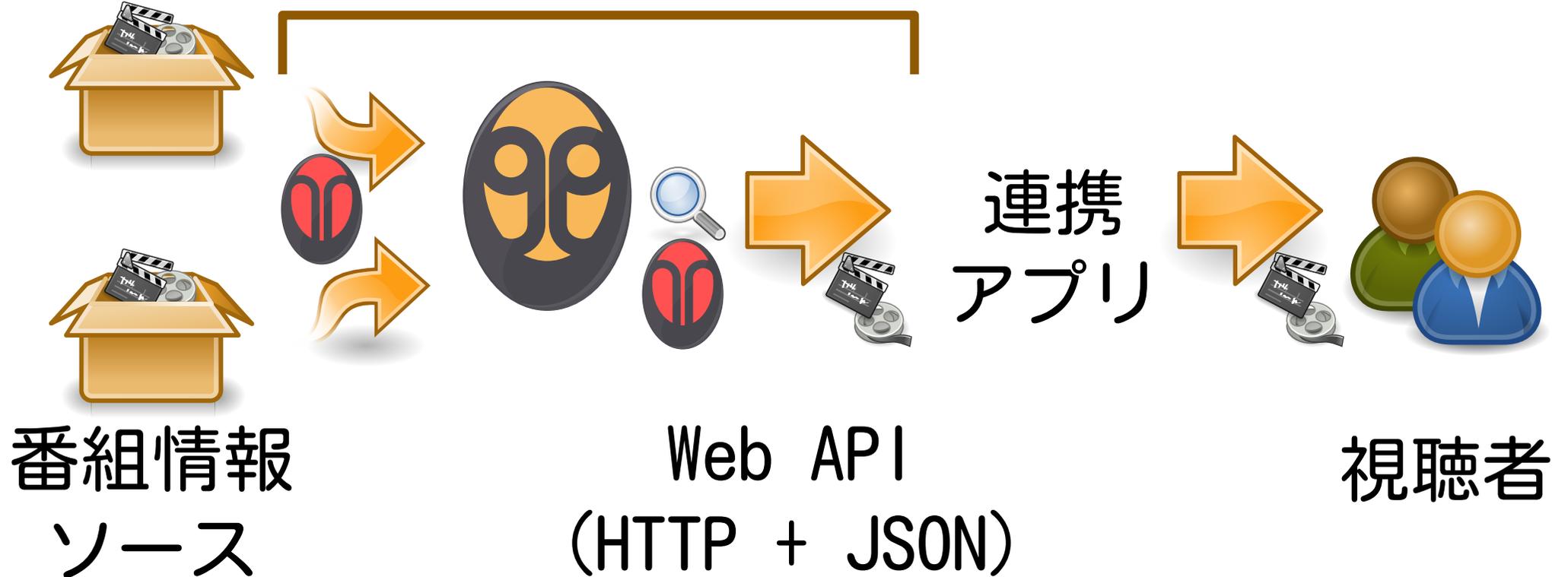
- ★ rroongaを使うメリット
- ★ rroongaを使うデメリット
- ★ 入力補完



# 開発事例

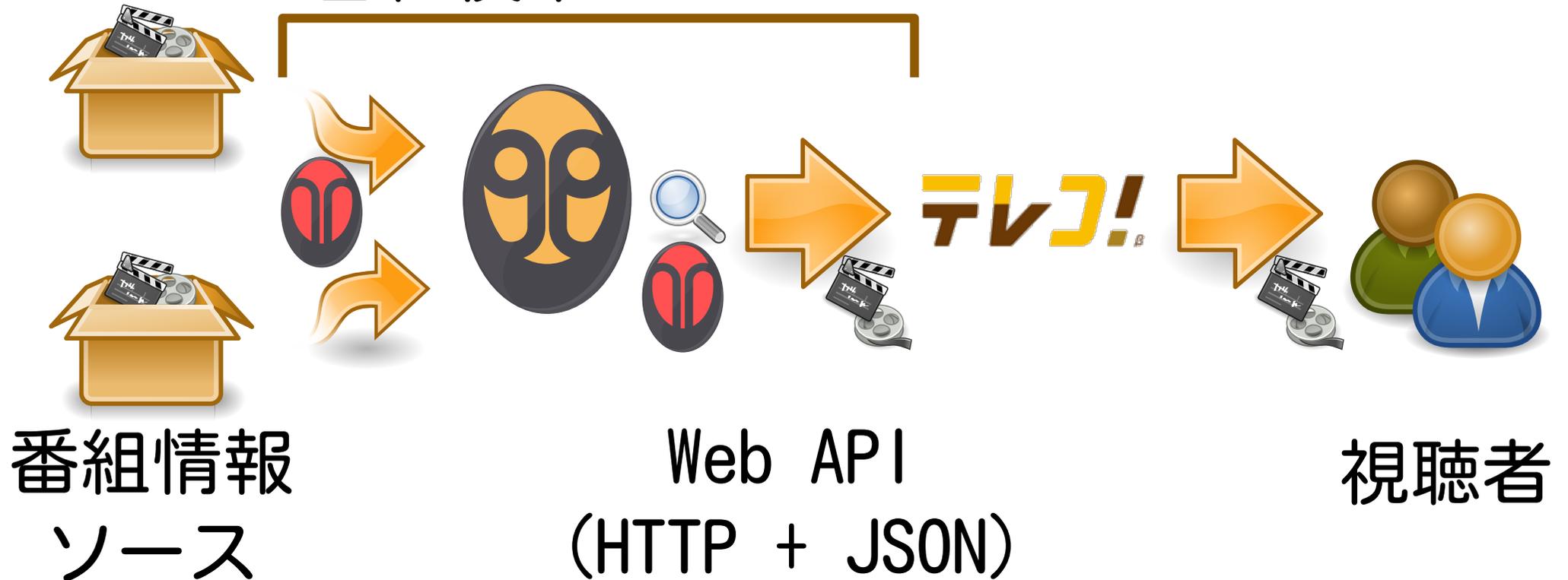
## 番組検索Web API

### 番組検索システム



# 連携サイト例：テレコ！

## 番組検索システム





# テレコ!



地上波・BS・CSの  
多チャンネル放送番組を横断検索!



## テレビをもっと楽しめるサービスが充実

テレコ!では地上波・BS・CSの横断検索をはじめ、  
特集やインタビュー、番組紹介解説など、多彩なサービスが盛りだくさん。  
ユーザーコメントの投稿や好きな番組の登録も可能!  
アナタの好みに沿った!“オリジナル番組リスト”の作成・閲覧ができます!





# 向いているところ

- ★ 細かい読み書きのコスト： 小
  - ★ **メタデータを抽出**できる  
(データロード時)
- ★ 検索処理の自由度： 高
  - ★ **便利な検索API**を作れる  
(groongaサーバーは汎用的なAPIのみ提供)



# メタデータ抽出

- ★ メタデータ
  - ★ 出演者・カテゴリ・…
  - ★ ドリルダウンに有用

番組説明から出演者を抽出



# 出演者抽出

```
description =  
  "出演者: ビートたけし・所ジョージ"  
people = Groonga["People"]  
# ["ビートたけし", "明石家さんま"]  
people.scan(description) do |record, |  
  p record.key # "ビートたけし"  
end
```

multiple string matching



# 入力補完

rroongaでの検索サービス開発事例

- ★ rroongaを使うメリット
- ★ rroongaを使うデメリット
- ★ 入力補完



# サジェスト機能

## groonga本体の機能

- ★ 補完
- ★ 補正
- ★ 提案
- ★ ↑の学習 (ユーザーからの入力を利用)

詳細: <http://groonga.org/ja/docs/suggest.html>



# 補完例

## ローマ字補完

mannyou

万葉仮名  
万葉調  
萬葉集

万葉仮名  
[まんようがな] (n) early  
used phonetically/  
万葉調  
[まんようちょう] (n) the p  
萬葉集  
[まんようしゅう] (n) Many  
of poems/  
万葉集  
[まんようしゅう] (n) Many

## ひらがな・カタカナ補完

まんよう

万葉仮名  
万葉調  
萬葉集

万葉仮名  
[まんようがな] (n) early  
used phonetically/  
万葉調  
[まんようちょう] (n) the p  
萬葉集  
[まんようしゅう] (n) Many  
of poems/  
万葉集  
[まんようしゅう] (n) Many



# 補完方法

- ★ コンテンツベース
  - ★ DB内にある語を補完
  - ★ 番組名・人物名など
- ★ 統計情報ベース
  - ★ ユーザーが検索した語を収集・解析
  - ★ 多くのユーザーが検索した語で補完



# 進め方

## 1. コンテンツベース

- ★ **すでにある情報ではじめられる**
- ★ **誤候補がないが候補が少ない**  
(あり:「名探偵コナン」 なし:「コナン」)

## 2. 統計情報ベースも合わせて

- ★ **ある程度の統計情報が必要**
- ★ **候補が増えるが低精度の可能性アリ**  
(特定の語を恣意的にたくさん検索されるとか…)



# おさらい

- ★ rroonga製検索システム
  - ★ インプロセスDBスタイル
  - ★ メタデータ抽出でドリルダウン支援
  - ★ 高レベルなAPI開発が得意
- ★ 冗長化は頑張らないといけない
- ★ 入力補完もできる



# お知らせ

groonga サポート  
はじめます！

RubyやMySQLなど  
関連ソフトウェアも  
サポート