

2012年11月10日

Ruby 初級者向けレッスン@KOF2012

Ruby 関西 ひがきまさる

Ruby とは

- まつもとゆきひろ氏が開発
- オブジェクト指向言語
- 軽量プログラミング言語
- **Ruby on Rails**が有名

試してみよう

- irb
- pry

```
$ irb
irb(main):001:0> puts "Hello World!"
Hello World!
=> nil
```

スクリプトを書いてみよう

```
#!/usr/bin/env ruby
# -*- coding: utf-8; -*-
puts "Hello World!"
```

```
$ chmod +x hello.rb
$ ./hello.rb
Hello World!
$
```

るりま

- **Ruby** リファレンスマニュアル刷新計画

`http://doc.ruby-lang.org/ja/1.9.3/`

`http://doc.ruby-lang.org/ja/search/`

スクレイピングしてみよう

- Webサイトのデータから、必要な部分を抜き出すこと
 - 毎日更新されるデータを自分のPCに保存
 - 複数のサイトから必要なデータを集めて表示
- 禁止されているサイトもある

準備運動

- ファイルから必要な部分を抜き出してみよう
- h2 の文字列を抜き出す例

```
def scrape_file fn
  open(fn) do |file|
    html = file.read
    html.scan(/<h2>(.*?)<\/h2>/)
  end
end
```

省略しないと……

```
begin
```

```
  file = open(fn, "r")
```

```
  html = file.read
```

```
  heads = html.scan(/<h2>(.*?)<\/h2>/)
```

```
  return heads
```

```
ensure
```

```
  file.close
```

```
end
```


実行してみよう

```
scrape_file("rubykansai.html").each do |h2|  
  puts h2  
end
```

「Ruby 関西」とは？

メーリングリスト

イベントカレンダー

勉強会・関西Ruby会議情報

Keyboard Shortcuts (see

Markdown Cheat Sheet

Looking for the GitHub logo?

もっと省略すると……

```
def scrape fn
  open(fn, &:read)
    .scan(%r|<h2>(.*?)</h2>|)
end

puts scrape('rubykansai.html')
```

Web アクセス

- ライブラリ `open-uri` を使用
- `h2` の文字列を抜き出す例

```
require 'open-uri'
def scrape_html url
  open(url) do |http|
    html = http.read
    html.scan(/<h2>(.*?)<\/h2>/)
  end
end
```

Nokogiri を使ってみよう

- XML や HTML のパーサ
- XPath や CSS3 セレクタで検索できる

```
$ gem install nokogiri
```

Nokogiri の例

- h2 の文字列を抜き出す例
- id="wiki-body" 内のみ処理する

```
require 'nokogiri'
def scrape_css url
  doc = Nokogiri::HTML(open(url))
  heads = doc.search("#wiki-body h2")
  heads.map{|head| head.content}
end
```

Mechanize を使ってみよう

- Web アクセスの自動化
 - ログインしないと見えないページ
 - ボタンを押さないと見えないページ

```
$ gem install mechanize
```

Mechanize の例 (Google アクセス)

```
require 'mechanize'
agent = Mechanize.new
agent.get('http://google.com/')
agent.page.form_with(name: 'f') do |form|
  form.q = 'にゃんこ'
end.submit
agent.page.search('h3.r a').each do |link|
  puts link.content
end
```

演習問題 1

スクレイピングしてみよう。

- **Ruby** 関西のページを HTML ファイルに保存しよう。
- そのファイルからリンクを取り出し、文字列と URL を出力しよう。

`https://github.com/rubykansai/`

`workshops/wiki`

演習問題 2

- 自分の好きなページを HTML ファイルに保存しよう。
- そのファイルからリンクを取り出し、文字列と URL を出力しよう。

演習問題 3

Mechanize を使って Google から好きな言葉を検索してみよう。

参考文献

- たのしい **Ruby** 第3版

ISBN 978 4 797357400

- 初めての **Ruby**

ISBN 978 4 873113678

- プログラミング言語 **Ruby**

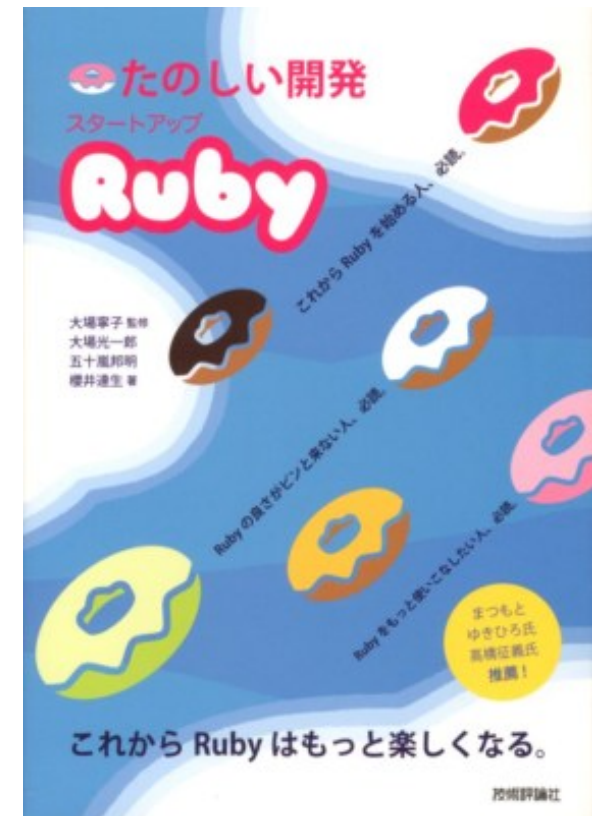
ISBN 978 4 873113944

参考文献

たのしい開発スタートアップ **Ruby**

ジュンク堂書店 KOF 店にて

絶賛 販売中!!!



参考資料

- 公式サイト

`http://www.ruby-lang.org/`

- 日本 **Ruby** の会

`https://github.com/ruby-no-kai/
official/wiki`

- **Ruby** 関西

`https://github.com/rubykansai/
workshops/wiki`

参考資料

- この資料

 - <http://2012.k-of.jp/session/305>

- twitter

 - @rubykansai
 - #rubykansai