

# Ruby 講義

## 第7回 Ruby入門

Kuniaki IGARASHI/igaiga  
2012.5.24 at 一橋大学  
社会科学における情報技術とコンテンツ作成III  
(ニフティ株式会社寄附講義)

○ 剰余金の配当に関するお知らせ

○ ニフティ、「@nifty EMOBILE LTE 定額にねんプラン」の提供を開始

○ 「@nifty温泉」で「母の日 全国一斉！100のありがとう風呂」特設サイト公開

○ 「スマブレ！」のサービス停止について

○ ニフティとサンリオウェーブ、iOS向けアプリ「Hello Kitty Worl...」

○ 平成24年3月期 決算短信

○ 特別損失の計上に関するお知らせ

○ 「シユフモ」登録会員数150万人を突破、「2012年主婦の全国節電調査（冬季...」

ニフティとなら、きっとかなう。  
With Us, You Can.

# ニフティ株式会社

アット・ニフティ  
楽しいサービスがいっぱい

アクセスマップ  
大森から西新宿へ移転いたしました

@nifty Web募金  
東日本大震災復興支援  
募金受付中

- 2012年4月25日 IR [特別損失の計上に関するお知らせ](#)
- 2012年4月25日 IR [剰余金の配当に関するお知らせ](#)
- 2012年4月19日 IR [「@nifty EMOBILE LTE 定額にねんプラン」の提供を開始](#)
- 2012年4月19日 IR [ニフティとサンリオウェーブ、iOS向けアプリ『Hello Kitty World』を台湾で提供開始](#)
- 2012年4月10日 お知らせ [「@nifty温泉」で「母の日 全国一斉！100のありがとう風呂」特設サイト公開](#)

# 提供

講師

# 五十嵐邦明

株式会社万葉

エンジニア



GARASHI

.9.25 at 高専カンファ

いがいが  
⑤

Teaching Assistant 演習 健二  
クックパッド株式会社 エンジニア



先週の  
おさらい

# Wikipediaのアクセス数解析

## コードで書く際の処理の流れ

- ①データファイルを開く
- ②データファイルから1行読み込む
- ③日本語データ以外はパス
- ④データ1行からタイトルとカウントを取得
- ⑤取得データをいれものに詰めてとっておく
- ⑥データファイル全行について繰り返し
- ⑦データファイルを閉じる
- ⑧貯まったデータをカウント順にソート（並べ替え）
- ⑨トップ20件表示

ここまでできました。

# Wikipediaのアクセス数解析

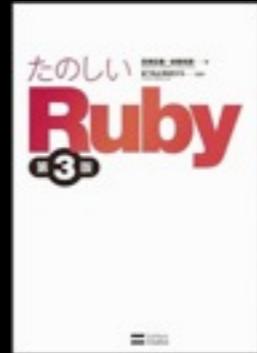
```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close
```

```
# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end

# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

ここまでできました。

# ファイルから1行ずつ読み込み



教科書 p.51~55

## サンプルコード

```
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
while text = file.gets
  puts text
end
file.close
```

# ファイルから各行読み込み

```
filename = "20120301-000000-ja.txt"  
file = File.open(filename, "r", encoding: "UTF-8")  
while text = file.gets  
  puts text  
end
```

**while: 条件成立中は繰り返し**  
この場合、条件は  
**text = file.gets**  
になります。

**text** に1行目のデータが入って**end**まで処理

**while** の行へ戻り、

**text** に2行目のデータが入って**end**まで処理

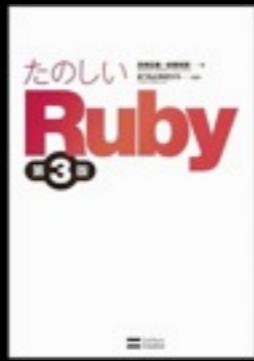
... 全行繰り返し

終端で**file.gets**すると条件不成立になり繰り返し終了

は1行読み込むメソッド

ファイル終端で読み込めない

# next文



教科書 p.102

```
while text = file.gets
```

```
... (A) ...
```

```
next
```

```
... (B) ...
```

```
end
```

繰り返し中にnext文を書くと、それ以降の処理を行わず、  
繰り返し先頭に戻します。

... (A) ... の部分の処理が行われてnextへくると、

... (B) ... の部分の処理は行われず、

繰り返し先頭へ戻ります。

教科書p.102の図6.2が分かりやすいです。

# unless文



教科書 p.76

unless 条件

処理

end

条件が偽の時に処理を実行

if文と対になる文

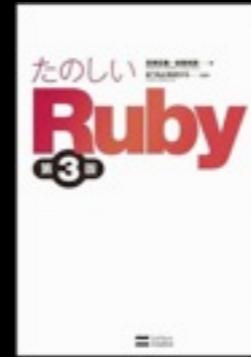
if文とは逆に条件が偽の時に処理実行

以下の2つの文は同じ意味

`puts "hello" if x != 3`

`puts "hello" unless x == 3`

# 正規表現



教科書 p.44~46  
p.267~290

文字列がパターンに一致（マッチ）するか調べる道具

ものすごく便利で、強力で、奥が深いです。  
ここではごく基本的な説明だけを行います。

**文字列 =~ /正規表現パターン/**

`=~` → 正規表現マッチを行う演算子

左辺の文字列中に正規表現パターンが含まれるかを判定します。

パターンが英数字や漢字だけからなる場合、

単純に文字列にパターンが含まれるかどうかを判定します。

`"Ruby" =~ /Ruby/` → 0

# マッチした場合はマッチ位置を返します

`"Ruby" =~ /Diamond/` → nil

# マッチしない場合はnilを返します

# Wikipediaのアクセス数解析

**next unless text =~ /ja/**

**unless** は条件 **text =~ /ja/** が不成立だったらその前の部分(**next**)を実行。

条件 **/ja/** は正規表現でjaから始まる の意味。

**next**は以降の処理を行わずに次の繰り返しへ進む。

→**text** が ja から始まらなかった場合、  
以降の処理は行わず、  
次の繰り返し (次の行) へ進む。

# String#split メソッド

splitは文字列を空白文字で区切ってArrayにします。

```
data = list.split
```

```
"ja.b %E3%81%84%E3%82%8D 1 6661".split  
→["ja.b", "%E3%81%84%E3%82%8D","1","6661"]
```

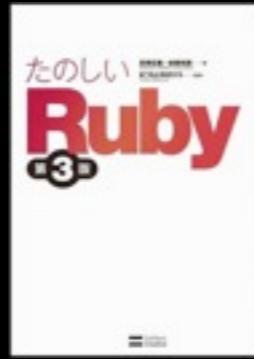
# CGI.unescape メソッド

%XX%XXっていう謎の文字列を、読める形式に変換します。

```
require "cgi"  
CGI.unescape("%E7%99%BD  
%E3%81%A0%E3%81%97")  
→"白だし"
```

この変換をURLデコードと呼びます。  
今はおまじないだと思ってください。

# 例外処理



教科書  
p.153~168

**begin**

例外を発生させる可能性のある処理

**rescue Exception => 变数**

例外が起こった場合の処理

**end**

コード実行中、うまく処理できない場合などに例外を発生させるメソッドがあります。例外が発生した場合、**rescue** 節で例外を捕まえ、その際に実行する特殊処理を書くことができます。

もしも、例外を**rescue** で捕まえない場合は、プログラムはそこでエラー終了します。

今週

ここから

# 目次

## Wikipediaアクセス解析(後編)

Wikipediaのアクセス数の解析の実習2週目です。  
必要なRubyの知識を説明しながら、いよいよ解析します。

ブロック  
リファレンスマニュアル

復習

# Wikipediaのアクセス数解析

wikipediaは1時間ごとのアクセス数データを公開しています。

<http://dumps.wikimedia.org/other/pagecounts-raw/>

## Index of page view statistics for 2012-05

### Pagecount files for 2012-05

Check the [hashes](#) after your download, to make sure your files arrived intact.

- [pagecounts-20120501-000000.gz](#), size 69M
- [pagecounts-20120501-010000.gz](#), size 67M
- [pagecounts-20120501-020000.gz](#), size 67M
- [pagecounts-20120501-030000.gz](#), size 66M
- [pagecounts-20120501-040000.gz](#), size 67M
- [pagecounts-20120501-050000.gz](#), size 77M
- [pagecounts-20120501-060000.gz](#), size 75M
- [pagecounts-20120501-070000.gz](#), size 80M

# Wikipediaアクセス数データ

ja.b %C3%84 1 6499

ja.b %C3%88%C2%B1%C3%AF%C2%BF%C2%BD%C3%A7%C2%AC%C2%AC%C3%AF  
%C2%BD%C2%B3%C3%A6%C3%AF%C2%BF%C2%BD%C3%AF%C2%BF%C2%BD  
%C3%AF%C2%BD%C2%AC%C3%AF%C2%BD%C2%AC973%C3%A8%C2%AD%C3%AF  
%C2%BF%C2%BD%C3%AF%C2%BD%C2%A1 1 6656

ja.b %E3%81%95%E3%81%BE%E3%81%96%E3%81%BE%E3%81%AA%E9%9D  
%A2%E3%81%8B%E3%82%89%E8%A6%8B%E3%81%9F%E6%97%A5%E6%9C%AC\_  
%E5%9C%BO%E7%90%86\_%E6%B0%97%E5%80%99 1 18210

ja.b %E3%82%A6%E3%82%A3%E3%82%AD%E3%83%9A  
%E3%83%87%E3%82%A3%E3%82%A2%E3%81%AE%E6%9B%88%E3%81%8D  
%E6%96%B9\_%E3%83%9D%E3%83%BC%E3%82%BF%E3%83%AB%E3%83%BB  
%E3%83%97%E3%83%AD%E3%82%B8%E3%82%A7%E3%82%AF  
%E3%83%88%E6%A1%88%E5%86%85 1 12093

ja.b %E3%82%A6%E3%82%A3%E3%82%AD%E3%83%9A  
%E3%83%87%E3%82%A3%E3%82%A2%E3%81%AE%E6%9B%88%E3%81%8D  
%E6%96%B9\_%E5%85%A5%E9%96%80%E7%B7%A8-  
%E3%82%A6%E3%82%A3%E3%82%AD%E3%83%9A  
%E3%83%87%E3%82%A3%E3%82%A2%E3%81%A8%E3%81%AF%EF%BC%9F 1  
15837

...

っていうデータが数十万行

# Wikipediaアクセス数データ

データ構造を見るためにちょっと読みやすく変えたもの

ja.b アーティキュレーションを表す記号 1 7642

ja.b カテゴリ:stab 1 68732

ja.b カテゴリ:大学入試 3 45061

ja.b カテゴリ:民法 1 47619

ja.b カテゴリ:社会学 1 6661

ja.b カテゴリ:User\_bg 1 7931

ja.b カテゴリ:User\_uk-3 1 6599

ja.b ガス事業法第2条 1 8541

ja.b ガリア戦記 1 12936

ja.b ガリア戦記/参照画像一覧 1 54089

ja.b コントラクトブリッジ/ルール 2 7957

ja.b コントラクトブリッジ/ルール/スコアリング 1 14903

...

っていうデータが数十万行

# Wikipediaアクセス数解析

言語種別 ページタイトル アクセス数 容量

ja.b %E3%81%84%E3%82%8D 1 6661

スペース区切りで以下の4項目が書かれている

言語種別： ja から始まるのが日本のデータ

ページタイトル： アクセスされたページ名

ただし、プログラムで扱い易い形式(%XX)になっている

アクセス数： アクセスされた回数

容量： そのページのデータサイズ

# Wikipediaアクセス数データ

言語種別 ページタイトル アクセス数 容量

ja.b %C3%84 1 6499
ja.b %C3%88%C2%B1%C3%AF%C2%BF%C2%BD%C3%A7%C2%AC%C2%AC %C3%AF%C2%BD%C2%B3%C3%A6%C3%AF%C2%BF%C2%BD%C3%AF%C2%BF %C2%BD%C3%AF%C2%BD%C2%AC%C3%AF%C2%BD %C2%AC973%C3%A8%C2%AD%C3%AF%C2%BF%C2%BD%C3%AF%C2%BD %C2%A1 1 6656
ja.b %E3%81%95%E3%81%BE%E3%81%96%E3%81%BE%E3%81%AA%E9%9D %A2%E3%81%8B%E3%82%89%E8%A6%8B%E3%81%9F%E6%97%A5%E6%9C %AC_%E5%9C%BO%E7%90%86_%E6%BO%97%E5%80%99 1 18210
...

このデータを解析して、ある1時間のアクセス数トップ20をコードを書いて調べてみます。  
簡単に言うと、「アクセス数」欄の数が大きいものから20個、その「ページタイトル」を表示させる

# Wikipediaのアクセス数解析

## コードで書く際の処理の流れ

- ① データファイルを開く
- ② データファイルから1行読み込む
- ③ 日本語データ以外はパス
- ④ データ1行からタイトルとカウントを取得
- ⑤ 取得データをいれものに詰めてとておく
- ⑥ データファイル全行について繰り返し
- ⑦ データファイルを閉じる
- ⑧ 貯まったデータをカウント順にソート（並べ替え）
- ⑨ トップ20件表示

# Wikipediaのアクセス数解析

```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close

# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end

# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

30行くらいで書けます

# Wikipediaのアクセス数解析

## コードで書く際の処理の流れ

- ①データファイルを開く
- ②データファイルから1行読み込む
- ③日本語データ以外はパス
- ④データ1行からタイトルとカウントを取得
- ⑤取得データをいれものに詰めてとっておく
- ⑥データファイル全行について繰り返し
- ⑦データファイルを閉じる
- ⑧貯まったデータをカウント順にソート（並べ替え）
- ⑨トップ20件表示

今週はここを説明

# Wikipediaのアクセス数解析

```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close

# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end

# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

今週はここを説明

# Wikipediaのアクセス数解析

```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close
```

ここまで、できしたこと

# count順にソート
データファイルの各行を解析して、  
結果を変数list(Arrayオブジェクト)へ詰める

```
# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

# データファイル各行の解析結果

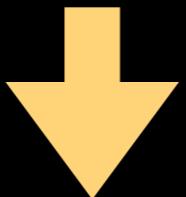
言語種別 ページタイトル アクセス数 容量

ja.b %E3%81%84%E3%82%8D 1 6661

ja.b %E3%81%95%E3%81%BE 3 2342

ja.b %E4%BC%81%E6%A5%AD 2 1656

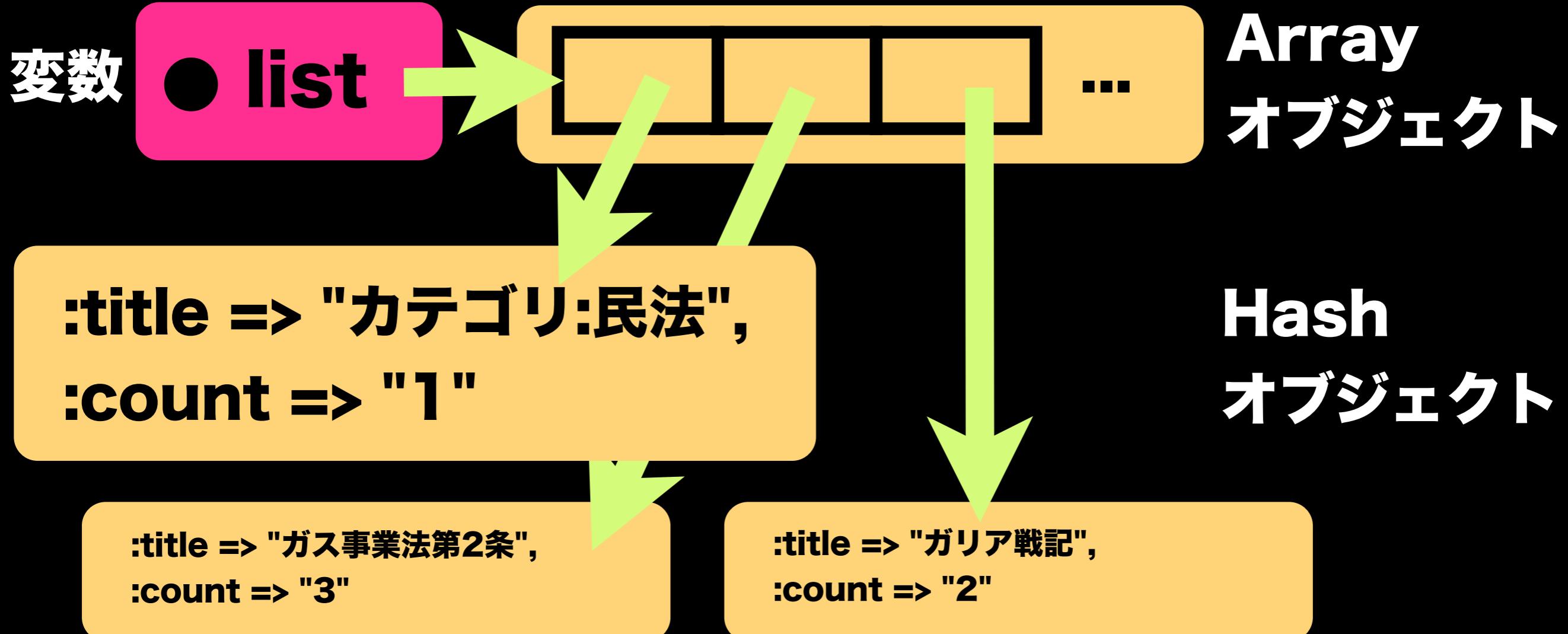
変数listの中身



```
[{:title=>"カテゴリ:民法", :count=>"1"},  
{:title=>"ガス事業法第2条", :count=>"3"},  
{:title=>"ガリア戦記", :count=>"2"},  
...]
```

# 変数listの中身の概念図

```
[{:title=>"カテゴリ:民法", :count=>"1"},  
 {:title=>"ガス事業法第2条", :count=>"3"},  
 {:title=>"ガリア戦記", :count=>"2"},  
 ...]
```

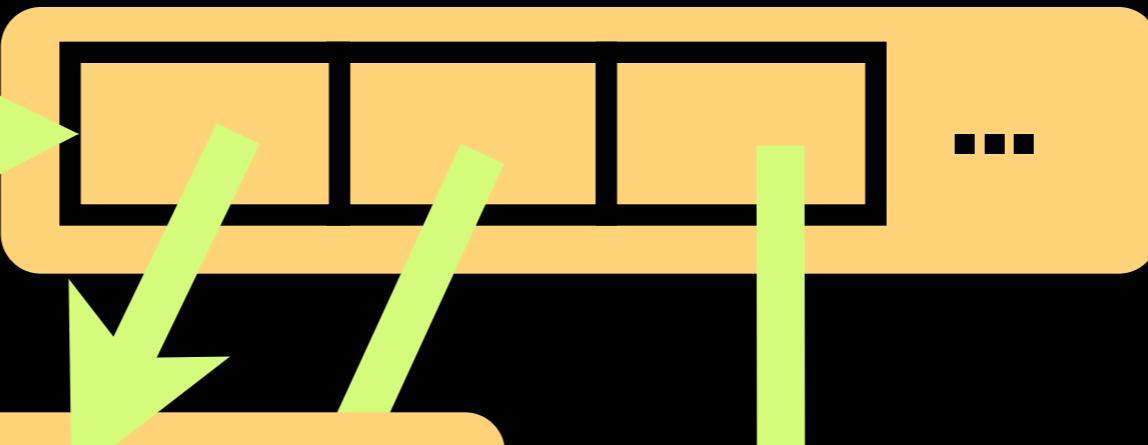
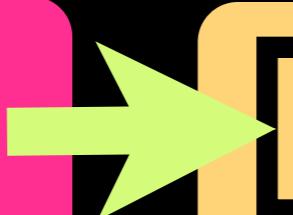


# count順に並べ変えたい =アクセスが多いランキングを見たい

```
[{:title=>"カテゴリ:民法", :count=>"1"},  
{:title=>"ガス事業法第2条", :count=>"3"},  
{:title=>"ガリア戦記", :count=>"2"},  
...]
```

変数

● list



Array  
オブジェクト

:title => "カテゴリ:民法",

:count => "1"

Hash  
オブジェクト

:title => "ガス事業法第2条",  
:count => "3"

:title => "ガリア戦記",  
:count => "2"

# count順に並べ変えたい

```
[{:title=>"カテゴリ:民法", :count=>"1"},  
{:title=>"ガス事業法第2条", :count=>"3"},  
{:title=>"ガリア戦記", :count=>"2"},  
...]
```

変数

● list



Array  
オブジェクト

```
[{:title=>"?", :count=>"XXXXXX"},  
{:title=>"?", :count=>"XXXXXX"},  
{:title=>"?", :count=>"XXX"},  
...]
```

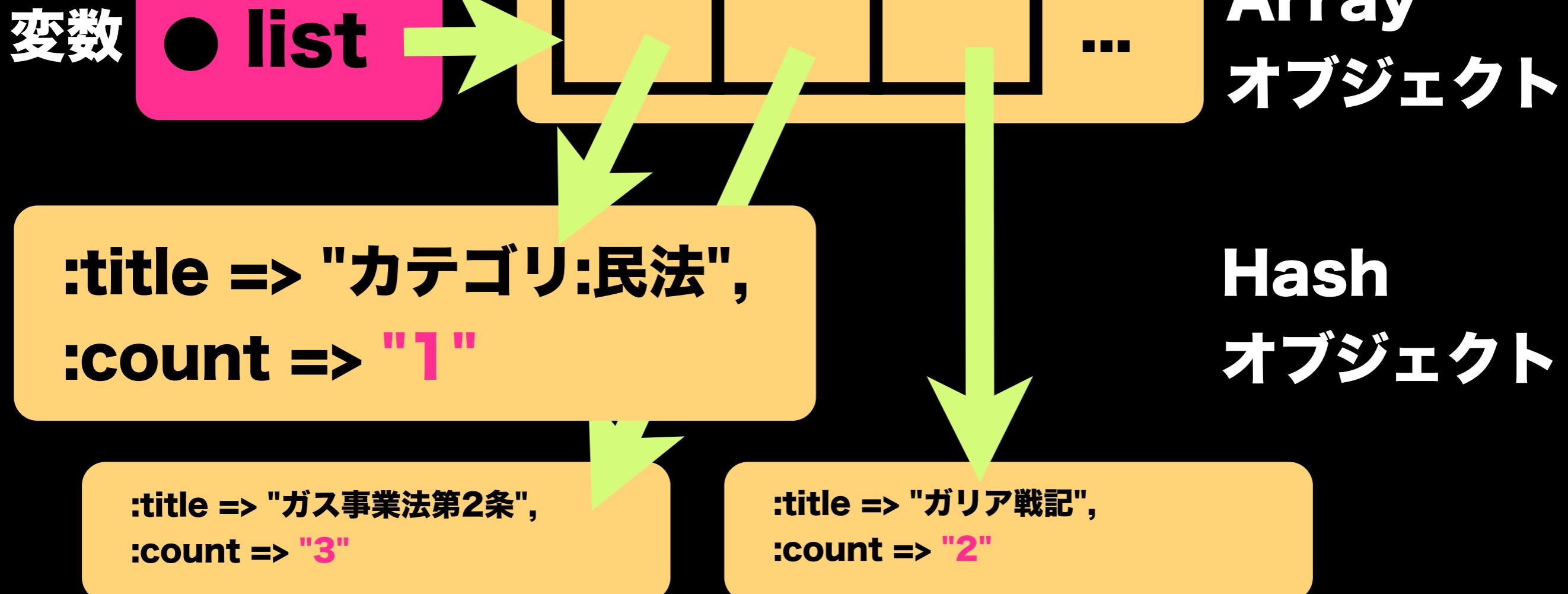
:title => "ガス事業法第2条",  
:count => "3"

:title => "ガリア戦記",  
:count => "2"

Hash  
オブジェクト

# count順に並べ変えるには

[{:title=>"カテゴリ:民法", :count=>"1"},  
list(Array)内の中身について  
{:count}を調べて並びかえる  
...]



# count順並び替えソースコード

```
# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end
```

詳細を説明していきます。

# Array#sort\_by メソッド

Arrayの中身をブロック(後述)の評価結果で並び替え

サンプルコード

```
array = [5,1,3]
result = array.sort_by do |i|
  i
end
p result
```

[1,3,5]

# ブロック

do ~ end で書かれる処理のかたまり

桃色の部分がブロック

```
array.sort_by do |i|
```

```
i
```

```
end
```

以前出てきた each メソッドについてたのも実はブロックです。

```
array.each do |i|
```

```
  puts i
```

```
end
```

# メソッド + ブロック

`Array#sort_by` や `Array#each` は

ブロックを添えて呼び出します。

ブロックを添えて呼び出す `Array` のメソッドは

大抵 `Array` の全要素について繰り返し実行します。

```
array.sort_by do |i|
```

```
i
```

```
end
```

`sort_by` : 全要素をブロックの評価結果で並び替え

```
array.each do |i|
```

```
  puts i
```

```
end
```

`each` : 全要素についてブロックを実行

# ブロックの評価結果

`sort_by`は「Arrayをブロックの評価結果で並び替え」するメソッドです。ブロックの評価結果はブロックで最後に実行された文になります。

```
array.sort_by do |i|
  i
end
```

←ブロックで最後に実行された文が  
ブロックの評価結果

# Array#sort\_by メソッド

Arrayの中身をブロックの評価結果で並び替え

サンプルコード

```
array = [5,1,3]
result = array.sort_by do |i|
  i
end
p result
```

1回目 i = 5 評価結果 5  
2回目 i = 1 評価結果 1  
3回目 i = 3 評価結果 3

[1,3,5]

実行結果

# count順並び替えソースコード

もしも変数listが以下の3つしか入っていないとすると

```
[{:title=>"カテゴリ:民法", :count=>"1"},  
 {:title=>"ガス事業法第2条", :count=>"3"},  
 {:title=>"ガリア戦記", :count=>"2"}]
```

```
result = list.sort_by do |i|  
 i[:count].to_i  
end
```

1回目の i {:title=>"カテゴリ:民法", :count=>"1"}

2回目の i {:title=>"ガス事業法第2条", :count=>"3"}

3回目の i {:title=>"ガリア戦記", :count=>"2"}

# count順並び替えソースコード

```
result = list.sort_by do |i|
  i[:count].to_i
end
```

i[:count]はHash要素読み出し

i[:count]には文字列オブジェクト "1" が入っているので、  
to\_i で整数オブジェクト 1 にする

変数 i

評価結果

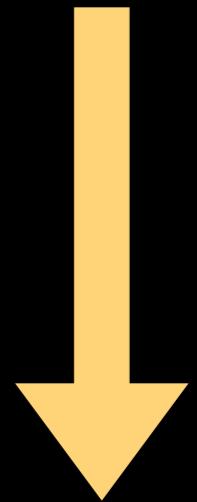
1回目	{:title=>"カテゴリ:民法", :count=>"1"}	1
2回目	{:title=>"ガス事業法第2条", :count=>"3"}	3
3回目	{:title=>"ガリア戦記", :count=>"2"}	2

# count順並び替えソースコード

```
[{:title=>"カテゴリ:民法", :count=>"1"},  
 {:title=>"ガス事業法第2条", :count=>"3"},  
 {:title=>"ガリア戦記", :count=>"2"}]
```

変数list

```
result = list.sort_by do |i|  
 i[:count].to_i  
end
```



```
[{:title=>"カテゴリ:民法", :count=>"1"},  
 {:title=>"ガリア戦記", :count=>"2"},  
 {:title=>"ガス事業法第2条", :count=>"3"}]
```

変数result

ただ、これだとカウントが少ない順なので、後で逆順にします。

# sort\_byの演習

## 1. Arrayオブジェクト

```
[{:name => "apple", :price => 50},  
 {:name => "grape", :price => 100},  
 {:name => "orange", :price => 30} ]
```

を:price安い順に並び替えて表示するコードを書いてください。

## 2. Arrayオブジェクト [3,1,-5] を絶対値が小さい順に並び替えて表示するコードを書いてください。

ヒント：絶対値の取得は 1.abs メソッド

## 3. 【上級】 あるArrayオブジェクト(例えば[3,1,5]とか[1,2,3]とか) が小さい順に並び替え済みかどうか判定するコードを書いてください。

# Wikipediaのアクセス数解析

## コードで書く際の処理の流れ

- ①データファイルを開く
- ②データファイルから1行読み込む
- ③日本語データ以外はパス
- ④データ1行からタイトルとカウントを取得
- ⑤取得データをいれものに詰めてとっておく
- ⑥データファイル全行について繰り返し
- ⑦データファイルを閉じる
- ⑧貯まったデータをカウント順にソート（並べ替え）
- ⑨トップ20件表示

ここができた

# Wikipediaのアクセス数解析

## コードで書く際の処理の流れ

- ①データファイルを開く
- ②データファイルから1行読み込む
- ③日本語データ以外はパス
- ④データ1行からタイトルとカウントを取得
- ⑤取得データをいれものに詰めてとっておく
- ⑥データファイル全行について繰り返し
- ⑦データファイルを閉じる
- ⑧貯まったデータをカウント順にソート（並べ替え）
- ⑨トップ20件表示

最後の部分です

# Wikipediaのアクセス数解析

```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close
```

```
# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end
```

```
# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

ここができた

# Wikipediaのアクセス数解析

```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close

# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end

# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

いよいよ最後の部分

# トップ20表示コード

```
# トップ20表示  
result.reverse.first(20).each do |i|  
  puts i  
end
```

詳細を説明していきます。

# ランキングを出す戦略1

```
[{:title=>"?", :count=>"1"},  
{:title=>"?", :count=>"1"},  
...  
{:title=>"?", :count=>"XXX"},  
{:title=>"?", :count=>"XXXX"},  
{:title=>"?", :count=>"XXXXX"}]
```

現在の変数**result**の中身  
**count**の少ない順にソート  
された**Array**オブジェクト

↓ **count**の多い順にしたい  
= 全要素を逆順にする

```
[{:title=>"?", :count=>"XXXXX"},  
{:title=>"?", :count=>"XXXX"},  
{:title=>"?", :count=>"XXX"},  
...  
{:title=>"?", :count=>"1"},  
{:title=>"?", :count=>"1"}]
```

# 調べてみよう

Arrayオブジェクトの全要素を逆順にするには？

Rubyのリファレンスマニュアルでメソッドを探せます。

例えば以下のページでリファレンスマニュアルを検索できます。

<http://miyamae.github.com/rubydoc-ja/1.9.3/>  
or 「サクサク Ruby リファレンス」でgoogle検索

The screenshot shows the Ruby 1.9.3 Reference Manual for the Array class. The search bar at the top contains the word "Array". The main content area displays the Array class with its methods listed:

- Array**  
<Enumerable  
配列クラスです。配列は任意の Ruby オブジェクトを要素として
- Array#&**  
self & other -> Array  
集合の積演算です。両方の配列に含まれる要素からなる新しい配列
- Array#\***  
self \* times -> Array  
配列の内容を times 回 繰り返した新しい配列を作成し返します。
- Array#+**  
self + other -> Array  
自身と other の内容を繋げた配列を生成して返します。
- Array#-**  
self - other -> Array  
自身から other の要素を取り除いた配列を生成して返します。

On the right side of the page, there is a sidebar with the title "Ruby 1.9.3 リファレンスマニュアル" and the heading "オブジェクト指向スクリプト言語 Ruby リファレンスマニュアル". Below the heading, there is a list of links:

- Ruby オフィシャルサイト <http://www.ruby-lang.org/ja/>
- version 1.9 対応リファレンス
- 原著：まつもとゆきひろ
- 最新版URL: <http://www.ruby-lang.org/ja/documentation/>

# リファレンスマニュアル検索



1. 画面左上の検索窓に「Array」と入力します。
2. 表示候補の最初の行「Array」をクリックします。

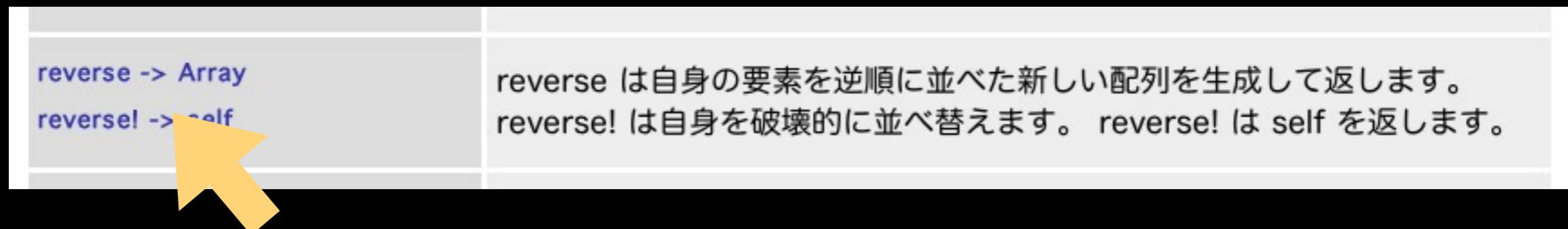
3. 右側にArrayのメソッド一覧などの説明が表示されます。

A screenshot of the detailed documentation for the 'Array' class. The title 'class Array' is prominently displayed. Below it, the inheritance chain is listed: 'クラスの継承リスト: Array < Enumerable < Object < Kernel < BasicObject'. A section titled '要約' (Summary) contains the text: '配列クラスです。配列は任意の Ruby オブジェクトを要素として持つことができます。一般的には配列は配列式を使って' (An array class. An array can hold any Ruby object as an element. Generally, arrays are used as array literals). On the left side, there is a sidebar with the same search results for 'Array' as shown in the previous screenshot.

下の方へ探していくと・・・

# リファレンスマニュアル検索

メソッド名と、そのメソッドの説明が並んでます。



4. 使えそうなメソッドのあたりをつけたらメソッド名をクリック

A screenshot of the Ruby 1.9.3 Reference Manual. The URL in the address bar is 'Ruby 1.9.3 リファレンスマニュアル > ライブラリ一覧 > 組み込みライブラリ > Arrayクラス > reverse'. The title of the page is 'instance method Array#reverse'. Below the title, it says 'reverse -> Array' and 'reverse! -> self'. The explanatory text reads: 'reverse は自身の要素を逆順に並べた新しい配列を生成して返します。 reverse! は自身を破壊的に並べ替えます。 reverse! は self を返します。'. At the bottom, there are two code snippets:

```
a = ["a", 2, true]
p a.reverse      #=> [true, 2, "a"]
p a              #=> ["a", 2, true] (変化なし)

a = ["a", 2, true]
p a.reverse!
p a              #=> [true, 2, "a"]
```

メソッド名

説明

サンプルコード

reverseを使えばできそうです。

TODO

迷うべき？

# リファレンスマニュアル検索

## 5. 分かり易いパターンでirbで試してみる

```
$ irb
```

```
[3,2,1].reverse
```

```
#=> [1, 2, 3]
```

reverseを使えばできそうです。

# リファレンスマニュアル検索演習

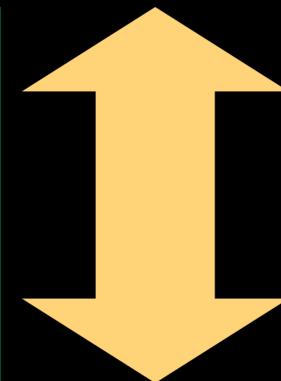
1. `["080","1234","5678"]` から  
`"080-1234-5678"` という文字列を作るメソッドをリファレンスマニュアルから探して、コードを作って実行してください。
2. 【上級】 `[1,2,3]` を、各要素を3倍した値で置き換えるメソッドをリファレンスマニュアルから探し、コードを作って実行してください。

# ランキングを出す戦略2

reverseを使って逆順にできました。

```
[{:title=>"?", :count=>"XXXXX"},  
 {:title=>"?", :count=>"XXXX"},  
 {:title=>"?", :count=>"XXX"},  
 ...  
 {:title=>"?", :count=>"1"},  
 {:title=>"?", :count=>"1"}]
```

全部で  
数万行



先頭から20個

全部表示するのは大変なので、先頭20件だけ取得したい

リファレンスマニュアルで検索してみましょう。

`first(n) -> Array`

先頭の n 要素を配列で返します。n は 0 以上でなければなりません。

Ruby 1.9.3 リファレンスマニュアル > ライブラリー一覧 > 組み込みライブラリ > `Array` クラス > `first`

## instance method `Array#first`

`first -> object | nil`

配列の先頭の要素を返します。要素がなければ `nil` を返します。

```
p [0, 1, 2].first #=> 0  
p [].first #=> nil
```

[SEE ALSO] `Array#last`

`first(n) -> Array`

先頭の n 要素を配列で返します。n は 0 以上でなければなりません。

[PARAM] n:

取得したい要素の個数を整数で指定します。

[EXCEPTION] `ArgumentError`:

n が負値の場合発生します。

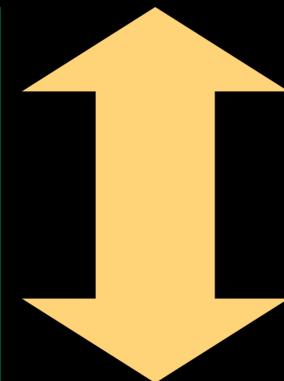
```
ary = [0, 1, 2]  
p ary.first(0)  
p ary.first(1)  
p ary.first(2)  
p ary.first(3)  
p ary.first(4)  
# => []  
[0]  
[0, 1]  
[0, 1, 2]  
[0, 1, 2]
```

これが  
使えそう

# ランキングを出す戦略2

firstを使って最初の20個を取得できました。

```
[{:title=>"?", :count=>"XXXXX"},  
 {:title=>"?", :count=>"XXXX"},  
 {:title=>"?", :count=>"XXX"},  
 ...]
```



先頭から20個

20個の要素を持つArrayオブジェクトができたので、  
前にやったeach文で繰り返し処理を行い、  
各要素を表示すればランキングを表示できそうです。

では、ここまで処理をつなげてみましょう。

# トップ20表示コード

```
result.reverse.first(20).each do |i|  
  puts i  
end
```

このようにつなげて書くことができます。

reverseメソッドも、

firstメソッドも、

Arrayオブジェクトを返すからです。

このような書き方をメソッドチェインと言います。

# Wikipediaのアクセス数解析

## コードで書く際の処理の流れ

- ①データファイルを開く
- ②データファイルから1行読み込む
- ③日本語データ以外はパス
- ④データ1行からタイトルとカウントを取得
- ⑤取得データをいれものに詰めてとっておく
- ⑥データファイル全行について繰り返し
- ⑦データファイルを閉じる
- ⑧貯まったデータをカウント順にソート（並べ替え）
- ⑨トップ20件表示

全部できました！

# Wikipediaのアクセス数解析

```
# encoding: utf-8
require "cgi"
filename = "20120301-000000-ja.txt"
file = File.open(filename, "r:UTF-8")
list = []
while text = file.gets
begin
  next unless text =~ /^ja/
  data = text.split
  h = { :title => CGI.unescape(data[1]), :count => data[-2] }
  list.push h
rescue Exception => e
  p e
end
end
file.close
```

```
# count順にソート
result = list.sort_by do |i|
  i[:count].to_i
end
```

```
# トップ20表示
result.reverse.first(20).each do |i|
  puts i
end
```

では、実行してみましょう！

# Wikipediaのアクセス数解析演習

1. 前述のコードを実行してください。.rbファイルと同じフォルダに  
20120301-000000-ja.txtを置いて実行してください。VMの人は  
~/src/wikipedia-analysis/20120301-000000-ja.txt  
に置いてあります。RailsInstallerやMacで実行する人はUSBディスク  
の中に入っているのでコピーしてください。（実行時間が長いので、ダウ  
ンロードしたデータファイルから日本語部分だけ抜き出して小さくしてあ  
ります。）

※Winで出力が "\u4FDD\u5143\u306E\u4E71" となる人は10  
行目付近を以下のコードで差し替えてください。

cp932は文字コードで、shift\_jisのwin版です。

旧 : h = { :title => CGI.unescape(data[1]), :count => data[-2]}

新 : h = { :title => CGI.unescape(data[1]).encode("cp932"), :count => data[-2]}

# Wikipediaのアクセス数解析演習

2. 【上級】データを配布しているサイトから任意のデータをダウンロードして解析してください。

<http://dumps.wikimedia.org/other/pagecounts-raw/>

ヒント：.gz形式で圧縮してあるので、解凍が必要です。

VM, Mac の場合は端末から \$ gunzip ファイル名 で解凍できます。

Windows の場合は例えばLhaplusを使って解凍できます。

<http://www.forest.impress.co.jp/lib/arc/archive/archiver/lhaplus.html>

また、コード内のfilenameを解凍したファイル名にする必要があります。

まとめ

# ブロック

do ~ end で書かれる処理のかたまり

桃色の部分がブロック

```
array.sort_by do |i|
```

```
i
```

```
end
```

以前出てきた each メソッドについてたのも実はブロックです。

```
array.each do |i|
```

```
  puts i
```

```
end
```

# メソッド+ブロック

`Array#sort_by` や `Array#each` は

ブロックを添えて呼び出します。

ブロックを添えて呼び出す `Array` のメソッドは

大抵 `Array` の全要素について繰り返し実行します。

```
array.sort_by do |i|
```

```
i
```

```
end
```

`sort_by` : 全要素をブロックの評価結果で並び替え

```
array.each do |i|
```

```
  puts i
```

```
end
```

`each` : 全要素についてブロックを実行

# ブロックの評価結果

`sort_by`は「Arrayをブロックの評価結果で並び替え」するメソッドです。ブロックの評価結果はブロックで最後に実行された文になります。

```
array.sort_by do |i|
  i
end
```

←ブロックで最後に実行された文が  
ブロックの評価結果

# Array#sort\_by メソッド

Arrayの中身をブロックの評価結果で並び替え

サンプルコード

```
array = [5,1,3]
result = array.sort_by do |i|
  i
end
p result
```

1回目 i = 5 評価結果 5  
2回目 i = 1 評価結果 1  
3回目 i = 3 評価結果 3

[1,3,5]

実行結果

# リファレンスマニュアル検索



1. 画面左上の検索窓に「Array」と入力します。
2. 表示候補の最初の行「Array」をクリックします。

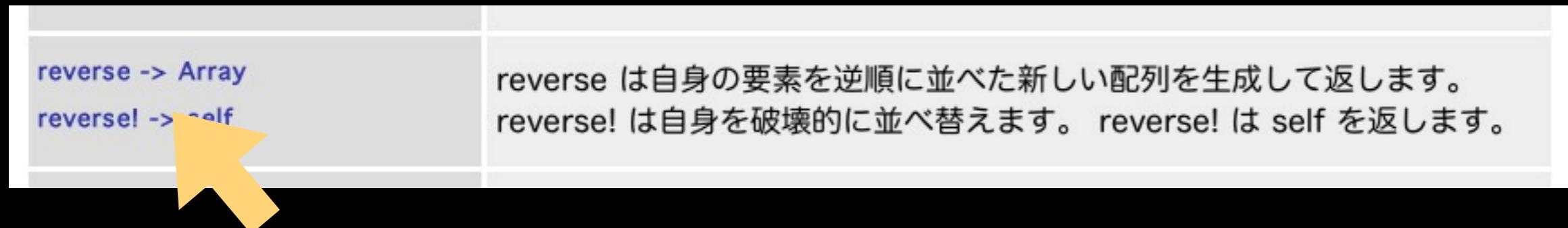
3. 右側にArrayのメソッド一覧などの説明が表示されます。

A screenshot of the detailed documentation for the 'Array' class. The title 'class Array' is prominently displayed. Below it, the inheritance chain is listed: 'クラスの継承リスト: Array < Enumerable < Object < Kernel < BasicObject'. A section titled '要約' (Summary) contains the text: '配列クラスです。配列は任意の Ruby オブジェクトを要素として持つことができます。一般的には配列は配列式を使って' (An array class. An array can hold any Ruby object as an element. Generally, arrays are used as array literals). On the left side, there is a sidebar with the same search results for 'Array' as shown in the first screenshot.

下の方へ探していくと・・・

# リファレンスマニュアル検索

メソッド名と、そのメソッドの説明が並んでます。



4. 使えそうなメソッドのあたりをつけたらメソッド名をクリック

A screenshot of the Ruby 1.9.3 Reference Manual. The URL in the address bar is 'Ruby 1.9.3 リファレンスマニュアル > ライブラリ一覧 > 組み込みライブラリ > Arrayクラス > reverse'. The main title is 'instance method Array#reverse'. Below it, there are two entries: 'reverse -> Array' and 'reverse! -> self'. The 'reverse!' entry is highlighted with a blue background. To the right of the entries is a explanatory text: 'reverse は自身の要素を逆順に並べた新しい配列を生成して返します。 reverse! は自身を破壊的に並べ替えます。 reverse! は self を返します。'. At the bottom, there are two code snippets:

```
a = ["a", 2, true]
p a.reverse      #=> [true, 2, "a"]
p a              #=> ["a", 2, true] (変化なし)

a = ["a", 2, true]
p a.reverse!
p a              #=> [true, 2, "a"]
```

メソッド名

説明

サンプルコード

reverseを使えばできそうです。

# リファレンスマニュアル検索

## 5. 分かり易いパターンでirbで試してみる

```
$ irb
```

```
[3,2,1].reverse
```

```
#=> [1, 2, 3]
```

reverseを使えばできそうです。

演習回答

# sort\_byの演習

## 1. Arrayオブジェクト

```
[{:name => "apple", :price => 50},  
 {:name => "grape", :price => 100},  
 {:name => "orange", :price => 30} ]
```

を:price安い順に並び替えて表示するコードを書いてください。

```
array = [{:name => "apple", :price => 50},{:name =>  
 "grape", :price => 100},{:name => "orange", :price =>  
 30}]  
  
result = array.sort_by do |i|  
   i[:price]  
end  
  
p result
```

# sort\_byの演習

2. Arrayオブジェクト [3,1,-5] を絶対値が小さい順に並び替えて表示するコードを書いてください。

ヒント：絶対値の取得は `i.abs` メソッド

```
array = [3,1,-5]
result = array.sort_by do |i|
  i.abs
end
p result
```

# sort\_byの演習

3. 【上級】 あるArrayオブジェクト(例えば[3,1,5]とか[1,2,3]とか)が小さい順に並び替え済みかどうか判定するコードを書いてください。

```
array = [3,1,5]
```

```
result = array.sort_by do |i|  
  i
```

```
end
```

```
p result == array
```

Arrayオブジェクトは == 演算子で同じかどうか判定できます。

sort\_by は対象のオブジェクトそのものを並び替えるのではなく、対象のオブジェクトを並び替えた新しいオブジェクトを返します。

なので、array の中身はsort\_by 後も変わらず [3,1,5] です。

対象のオブジェクトを並び替える sort\_by! メソッドもあります。

対象のオブジェクトを書き換えるメソッドは破壊的メソッドと呼び、メソッド名末尾に ! がつくものが多いです。

# リファレンスマニュアル検索演習

1. ["080","1234","5678"] から

"080-1234-5678" という文字列を作るメソッドをリファレンスマニュアルから探して、コードを作って実行してください。

```
["080","1234","5678"].join("-")
```

2. 【上級】 [1,2,3] を、各要素を3倍した値で置き換えるメソッドをリファレンスマニュアルから探し、コードを作って実行してください。

```
[1,2,3].map! do |i|
  i * 3
end
```

# 参考文献

ウィキペディア記事閲覧回数の特徴分析  
人工知能学会研究会資料 **SIG-SWO-A901-03**  
曾根 広哲、山名 早人  
早稲田大学大学院基幹理工学研究科 早稲田大学理工学術院 国立情報学研究所  
<http://sigsw.org/papers/SIG-SWO-A901/SIG-SWO-A901-03.pdf>

# 講義資料置き場

講義資料置き場をつくりました。  
過去の資料がDLできます。

<https://github.com/hitotsubashi-ruby/lecture2012>

or

<http://bit.ly/ruby-lecture>

# 雑談・質問用facebookグループ

facebookグループを作りました

<https://www.facebook.com/groups/hitotsubashi.rb>

- ・加入/非加入は自由です
- ・加入/非加入は成績に関係しません
- ・参加者一覧は公開されます
- ・書き込みは参加者のみ見えます
- ・希望者はアクセスして参加申請してください
- ・雑談、質問、議論など何でも気にせずどうぞ～
- ・質問に答えられる人は答えてあげてください
- ・講師陣もお答えします
- ・入ったら軽く自己紹介おねがいします