

# 日本発のプログラミング言語Ruby : Webアプリケーションからバイオインフォマティクスまで

Programming Language "Ruby"; From Web Application to Bioinformatics

石井 一夫  
Ishii Kazuo

Rubyは、日本で産まれたプログラミング言語である。非常に習得が容易でありながら強力な文字列データ処理能力を持つ汎用の言語である。この2, 3年の最新のインターネットの技術の進歩に伴って使用が広がり世界的に脚光を浴びている。また、バイオテクノロジーでは、ゲノム解析が進み大量のデータを扱うようになったが、そのデータの解析にもRubyは適しており、注目される。

Ruby is a programming language produced in Japan, which is very easy to learn and has a powerful text-processing capacity. It became a popular language with the latest progress of the Internet worldwide in recent two or three years. In recent biotechnology, huge genomic data analyses are required in clinical and industry field. Ruby is one of the useful and popular programming tools for these analyses. Here, the recent advances of Ruby programming language from web application to bioinformatics will be reported.

**キーワード : Ruby, Webアプリケーション, 国産プログラミング言語, バイオインフォマティクス**

## 1 はじめに

近年、インターネットでのWebアプリケーションの応用の広がりにより、日本生まれのオープンソースのプログラミング言語Rubyが注目されている。Rubyは非常にわかりやすい文法と、読みやすい形式をもつプログラミング言語で、非常に強力なデータ処理能力を持っており、迅速なウェブの構築に向く言語として、世界的に認知されつつある。

特に、2005年ごろから、従来の静的なインターネットのウェブサイトと異なる新しい形態であるブログやWikiなど動的で対話的なウェブサイトの活用が広がってきている。また、

- (1) mixiをはじめとするSNS(ソーシャル・ネットワークワーキング・サービス) や
- (2) Youtube やニコニコ動画などの動画共有サービス,
- (3) Google, Gmail, Amazonなどの洗練されたウェブサービスの利用が普及してきている。

このようなインターネットの新しい形態はWeb 2.0と呼ばれている。Rubyはこのようなインターネットサービスの構築に向く言語のひとつ

```
# Greeter クラスを定義する。
class Greeter
# defによりメソッドを定義する。
# 引数としてnameを受け取るinitializeメソッドを定義する。
def initialize(name)
  @name = name.capitalize
end
# Hello "name" を出力するsaluteメソッドを定義する。
def salute
  puts "Hello #{@name}!"
end
end

# "world"を引数とする新しいオブジェクトgを作成する。
g = Greeter.new("world")

# saluteメソッドにより, "Hello World!"を出力。
g.salute
```

図1 国産プログラミング言語Ruby のプログラム例。Javaなどと比較して直感的でわかりやすい記述である。詳細は本文を参照。

つであり、PerlやPython、PHP、Javaなどのプログラミング言語とともによく用いられる。従来は、一部のコンピュータマニアで使われる言語であったが、ここ数年普及が進み、2007年（第17回）日経BP技術賞を受賞するなど注目を浴びている。そのきっかけは、2005年ごろからRubyを用いたWebアプリケーションのフレームワーク（作成ツール）であるRuby on Railsが、世界的にブレイクしたためである。プログラミング言語の人気ランキングを紹介するウェブサイトであるTIOBE Programming Community Indexでも10位に位置しており（2008年7月現在）、Rubyは、このサイトで2006年のProgramming Language of the Yearに選ばれている。

また、Rubyは、インターネットだけでなく、バイオテクノロジーの分野でも注目されている。バイオのデータ解析専用の関数を集めたライブラリーも整備されている。筆者もこのRubyを使ったデータ解析を行うケースが増えている。

今回、このような国産プログラミング言語のRubyについて、インターネットでの応用と、バイオテクノロジーでの展開について紹介する。

## 2 RubyとWebアプリケーション

### 2.1 Rubyとは

Rubyは、日本人のまつもとゆきひろ氏（（株）ネットワーク応用通信研究所）が開発した国産のプログラミング言語で、1995年に公開された。Rubyの名称は、すでに存在していたプログラミング言語Perlのように宝石の名前に由来する。6月の誕生石がPearl（Perl）であるのに対し、7月の誕生石がRubyでPearlの次に来ることで、Rubyが「Perlに代わる次の言語になる」という思いがこめられている。Rubyは、以下のような特徴をもつ。

- ① 首尾一貫したシンプルな美しい文法と、「驚き最小の原則」による自然な記述
- ② 徹底したオブジェクト指向による言語構造の一貫性
- ③ 簡易にプログラムができて（コンパイルという作業を必要としない）平易に利用できるイ

ンタプリタ型スクリプトプログラミング言語

#### ④ 充実した文字列処理（正規表現）機能

一般的なプログラミング言語のもつ機能（制御構造や、繰返し処理、数値演算、文字列処理機能など）を十分に備えている汎用の言語である。コマンドにより対話的に使用するインタラクティブ（対話）モード（`irb`）を用いてのプログラミングも可能である。ここで、オブジェクト指向とは、ソフトウェアの設計や開発において、操作手順よりも操作対象に着目した方式で、現在主流の方式である。

非常に習得が容易で、読みやすく、書いていて楽しい言語だといわれている。Javaを学習したことがある人ならばわずか「5時間でマスターできる」ともいわれている。

### 2.2 Rubyの文法（図1）

Rubyの文法（シンタックス）はシンプルであり、以下のような構文を基本とする。

```
obj.method(引数)
```

`obj` …オブジェクト：動作の対象、目的物  
`method`…メソッド …動作、命令の内容  
`引数` …引数 …動作の細かい指示  
 通常のプログラミング言語における関数と呼ばれるプログラムの処理過程をまとめたものに相当するものはRubyでは、メソッドと呼ばれる。

また、メソッドの動作処理対象はオブジェクトという名前と呼ばれる。オブジェクトには、数値、文字列、配列、ハッシュ、ファイル、ディレクトリなどが含まれる。これらの動作対象（オブジェクト）ごとにクラスとしてまとめて整理された関数（メソッド）が整備されている。

また、Javaに見られるような、承継、データの隠蔽、関数や変数の多様性（ポリモルフィズム）、ガベージコレクション、マルチスレッド、例外処理などの複雑なオブジェクト指向言語としての機能を十分に備えている。また、関数型言語Lispのラムダ関数の考え方に似た関数ブロックをオブジェクトとして扱うイテレータと呼ばれる機能や、多重承継を形式的に可能にするMixinと

いうユニークで便利な機能が盛り込まれている。

## 2.3 RubyによるWebアプリケーション

Rubyによるインターネットのアプリケーション（Webアプリケーション）としては、Ruby on Rails (RoR, Rails) が有名である。

これは、デンマークのプログラマであるデビット H. ハンソンの開発により2004年に公開され、2005年後半ごろから一気にブレイクし、普及が進み、非常に注目されている。そして、Rubyが世界に飛躍するきっかけとなった。これにより、従来の「Javaの10倍の速さ、コスト、効率」でWebアプリケーションを開発することが可能であるといわれる。アジャイル（迅速）ソフトウェア開発の代表的ソフトウェアとして認知されている。

Ruby on Railsは、モデル・ビュー・コントローラ（MVC）アーキテクチャという考え方に基づき構築されている。MVCとはアプリケーションソフトウェアを、モデル（データおよび手続き処理）・ビュー（出力表示処理）・コントローラ（入力応答処理）に3分割して設計・実装する技法、もしくはそのような構造のことである。

Ruby on Railsでウェブを構築する際の基本的な考え方は「同じ事を繰り返さない」（DRY：Don't Repeat Yourself）、「設定よりも規約」（CoC：Convention over Configuration）である。定義などの作業は一回だけで済ませる（DRY）、標準的な設定は決まったルールに従い、共通でない部分の設定のみを行う（CoC）ように考えられている。

## 3 バイオテクノロジーとRuby

### 3.1 バイオテクノロジーとプログラミング

近年のヒトゲノム解析後のバイオテクノロジーのデータ解析技術の発展により、大量多種のデータ産生を行う技術や機会が普及してきており、解析データをハイスループット（高速）に産生し解析することが日常的になりつつある。このため大量のデータ処理による情報技術活用の促進がなされてきた。

このような生物から得られたデータを扱う分野であるバイオインフォマティクス（生物情報科学）においては、特に、（1）遺伝子・タンパク質の配列データを対象とする「文字列情報の解析」のデータ解析や、（2）マイクロアレイなど「数理解析」によるデータ処理が必要になってきた。このようなデータ解析においては、データベース、プログラミングそしてインターネットなど情報技術の活用が鍵となる。

### 3.2 バイオテクノロジーで用いられるプログラミング言語

バイオインフォマティクスでは、タンパク質の高次構造予測など多くの計算量を必要とする「重い」計算を行う際には、主にC言語などの比較的 low 水準に近い言語が使われてきた。

一方、生物の基本的構造を示す遺伝子の塩基配列など、巨大な文字列を処理する際には、スクリプト型のプログラミング言語がよく用いられる。なかでもPerlと呼ばれる言語は、文字列処理を得意とし、習得が容易で、バイオ解析のためのライブラリー（専用プログラムの集合体）が豊富であるため、ゲノムやたんぱく質の配列解析では、盛んに用いられてきた。

しかし、Perlは、（1）可読性が悪く、（2）バイオでもよく使われるようになりつつあるオブジェクト指向という新しい形式のプログラミングを行うのには向いていない。このため、近年は言語のシンプルさやオブジェクト指向への適合性、そしてPerlと同様にバイオ解析用ライブラリーが充実しつつあることなどの理由から、Perlに替わってRubyやPythonも使われるようになってきている。PythonとRubyは、よく似ているが、欧米ではPythonが、日本ではRubyがポピュラーである。

### 3.3 さらなるブレイクスルー—個別化医療の実現化とプログラミング言語Ruby

また、最近になり454 FLX (Roche社)、SOLiDシステム (Applied Biosystem社)、Genome Analyzer (Illumina社、写真1) など、

一度に数キガ（数十億）塩基の解析も可能となる次世代のDNA自動解析装置（次世代DNA自動シーケンサー）が発売された。



写真1 個別化医療の実現のための次世代DNA自動解析装置の例 (Illumina社 Genome Analyzer)

これにより、ヒトゲノムの解析が1週間程度で可能になる。今後、個人レベルのゲノム解析をもとにした個人別に最適化された医療（個別化医療）の実現化が一気に現実味を帯びてきており、バイオインフォマティクスの需要が急速に伸びることが予想される。これらの機器により産生される大量の遺伝子配列データの解析には、文字列処理（正規表現）機能の充実したRubyは特に威力を発揮すると思われる。また、Rubyは、アジャイル（急速な）ソフトウェア開発に向く言語であり、今後の個別化医療の実現の一躍を担う言語としても注目される。

## 4 Rubyによるバイオインフォマティクス

### 4.1 Rubyによるバイオインフォマティクスアプリケーション

バイオテクノロジーでRubyを使用する利点に、関連のツールが充実しているということがあがる。Rubyによる代表的なバイオ関連ツールとしては以下のようなものがある。

- ① BioRuby (Rubyのバイオ解析用ライブラリ)
- ② ChemRuby (化学解析用ライブラリ),
- ③ RSRuby (統計解析Rアクセスパッケージ)

BioRuby (<http://bioruby.org/>) は、バイオインフォマティクスのためのメソッドを集めたクラスライブラリである。配列解析、パスウェイ解析、文献解析、データベースのほか、各種アプリケーション、ウェブサービスなどを包含している。また、化学物質を扱う際にはChemRuby (<http://chemruby.org/>) が用いられる。Ruby

を用いてDNAマイクロアレイ解析などの数値解析を行う統計解析ソフトRへアクセスしデータ解析を行うにはRSRuby (<http://web.kuicr.kyoto-u.ac.jp/~alexg/rsruby/>) をインストールして使うことができる。

この3つのツールは、いずれも日本国内で開発されており、使用説明書などのドキュメントを日本語で入手でき開発者とも日本語でやりとりできるという点は他の言語にない利点である。

## 5 おわりに

筆者は主に、遺伝子の塩基配列データの検索や文字列処理をRubyを用いて行い、数値計算をRというソフトを用いて行っている。また、日本Rubyの会に所属し、Ruby関連のドキュメント作成や、関連イベントのサポートも行っている。

国産のプログラミング言語であるため開発の拠点が国内にあり情報がいち早く日本語で入手できる点は他の言語に比べ有利である。海外の普及も盛んであり、国際的な開発協力体制もしっかりしている。最近では、Javaや.NETなど他の言語から使用できるJRubyやIronRubyなどの環境が整備されてきた。日本発の国際標準となりうるプログラミング言語として今後非常に有望であり、今後の普及に努めたいと考えている。

### <参考文献>

- 1) 高橋征義, 後藤裕蔵: たのしいRuby 第2版, ソフトバンククリエイティブ, 2006
- 2) Dave Thomas et al.: プログラミングRuby第2版 言語編・ライブラリ編, オーム社, 2006
- 3) Dave Thomas et al.: RailsによるアジャイルWebアプリケーション開発 第2版, オーム社, 2007
- 4) 樋口千洋, 石井一夫: 統計解析環境Rによるバイオインフォマティクスデータ解析, 共立出版, 2007

石井 一夫 (いしい かずお)  
技術士 (生物工学部門)

東京理科大学DDS研究センター  
博士 (医学)

