

相互運用性を向上させる サービス構築用テンプレートの開発

片山 俊明

東京大学 医科学研究所 ヒトゲノム解析センター ゲノムデータベース分野

自己紹介



- BioRuby
 - <http://bioruby.org/>
- KEGG API (ウェブサービス)
 - <http://www.genome.jp/kegg/soap/>
- KEGG DAS (ゲノムブラウザ+DAS)
 - <http://das.hgc.jp/>
- オープンバイオ研究会
 - <http://open-bio.jp/>
- クマムシゲノムプロジェクト
 - <http://kumamushi.net/>

The screenshot shows a Mac OS X desktop with a Dock at the bottom containing icons for Finder, Mail, Safari, iCal, Address Book, iPhoto, iMovie, iTunes, iDVD, iSync, iWork, iLife, and iTools. A Dock icon for 'Kumamushi' is also present. The desktop background is black with a large, semi-transparent image of a tardigrade (water bear) in the center. The image has a digital, pixelated texture and is overlaid with Japanese text: 'クマムシゲノムプロジェクト' (Kumamushi Genome Project) at the top right and 'クマムシは、進歩動物門 Tardigrada に属する強靭な生命力を持つたナイスな生き物です。特にトゲクマムシは攻撃機動隊やガンダムに出てきそうな装甲がかっこいいです。' (Kumamushi belongs to the Tardigrada phylum and is a nice creature with strong resilience. Specifically, the spiny tardigrade is cool-looking like an attack mobile suit or Gundam's armor.) Below the image, there is a sidebar with a menu icon and the text 'クマムシ'.

BioRuby

- http://bioruby.org/
- KEGG API
- KEGG DAS
- http://das.hgc.jp/
- KEGG DAS: aae:554596..604595
- オープンバイオ - オープンバイオ研究会
- http://open-bio.jp/
- クマムシゲノムプロジェクト - クマムシゲノムプロジェクト
- http://kumamushi.net/

クマムシゲノムプロジェクト

クマムシは、進歩動物門 Tardigrada に属する強靭な生命力を持つたナイスな生き物です。特にトゲクマムシは攻撃機動隊やガンダムに出てきそうな装甲がかっこいいです。

ニュース

2007/3/13 この日発売の SPA! に、クマムシと鈴木先生の記事が載るそうです。
2007/3/8 ネイチャー＆サイエンスカフェ Vol.5 「クマムシの採集と飼育と観察と研究！」のご案内を頂きましたので転載します。

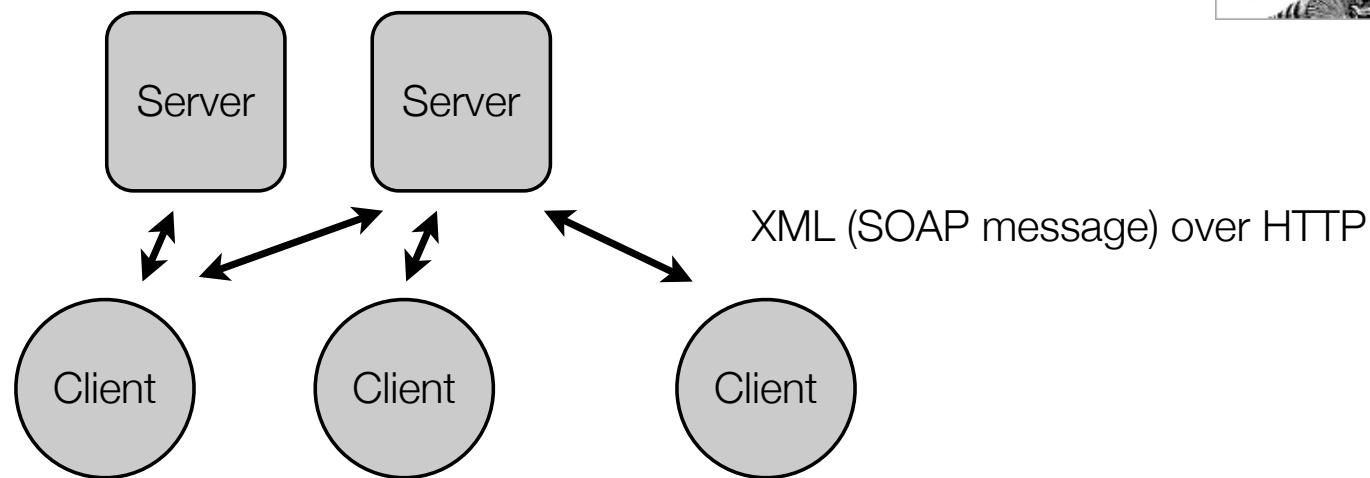
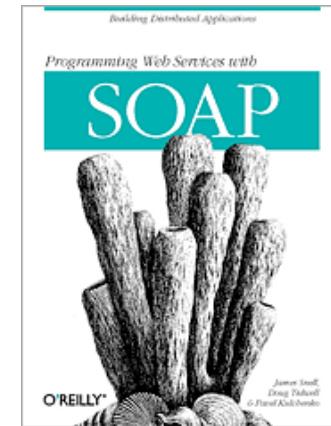
昨年刊行された『クマムシ？！ 小さな怪物』（岩波科学ライブラリー／岩波書店）は、クマムシの魅力やその不思議を明らかにした世界初の一冊向けクマムシ本として、クマムシファンのみならず、広く一般社会にも一ページセッションを巻き起こしました。今回は、その翻案の問題作？！の著者である鈴木忠さんによる登場順って、本には書ききれなかったことや、生物学や生態学など生物研究の面白さ、ご自身の「ヤンスマブワングー」、などについてお話を！ いただき、クマ

アンケート

- ウェブサービス (SOAP/WSDL)
 - 知らない人
 - ___人 (約___%)
 - 使ったことがない人
 - ___人 (約___%)
 - なんて濃いオーディエンスなんでしょう。。。

相互運用性

- ウェブサービス
 - HTTP を利用しプログラム等からサーバの機能を利用
 - 狹義には SOAP/WSDL
 - Service Oriented Architecture Protocol (Simple Object Access Protocol)
 - Web Service Description Language



ウェブサービス - ターゲット分野の例と現実

- E-commerce
 - 応答のスピードが重要 (SOAPのオーバーヘッドが余分?)
 - 実際にはクローズドな少数のサーバ間で最適化?
 - 古いシステムに SOAP の皮をかぶせて相互運用
- バイオインフォマティクス
 - 解析処理などトランザクションに時間がかかるのが普通
 - データベース毎、解析アプリ毎にサーバが多様かつ分散して存在
 - 次々と生まれる新しいデータベースと解析アプリ
 - バイオインフォのオープンさは他分野よりウェブサービス向きかも

バイオインフォの代表的なウェブサービス

- DDBJ XML <http://www.xml.nig.ac.jp/wsdl/>
- KEGG API <http://www.genome.jp/kegg/soap/>
- EBI WS <http://www.ebi.ac.uk/Tools/webservices/>
- NCBI ESOAP http://www.ncbi.nlm.nih.gov/entrez/query/static/esoap_help.html

The screenshot displays four bioinformatics web services side-by-side:

- NCBI ESOAP**: Shows a grid of buttons for services like EGQuery, EFetch, EInfo, ELink, ESearch, ESpell, and ESummary.
- EBI WS**: Shows a grid of buttons for services like Fasta, WU-Blast2, MPsrch, ScanPS, ClustalW, TCoffee, MUSCLE, WSdbfetch, MSD, Ontology, and Citation.
- DDBJ XML**: Shows a grid of buttons for services like DDBJ, Ensembl, SRS, ARSA, NCBI GA, RefSeq, getEntry, Mapping, OMIM, Gib, GO, Blast, Gtop, TxSearch, Fasta, PML, VecScreen, ClustalW, and SPS.
- KEGG API**: Shows a grid of buttons for services like DBGET, GENES, LinkDB, Motif, PATHWAY, SSDB, GENOME, KO, OC, PC, and LIGAND. A yellow callout box provides statistics for the KEGG API v5.0: 9.5 million hits since Feb 2006, 3735017 Perl, 2818454 Java, 2191769 Ruby, and 287940 Python.

サービス内容は2006年7月時点のもの（なので今は多少増減があるかも）

現在ウェブサービスでできること

| Web services | NCBI | EBI | DDBJ | KEGG |
|--------------|--------|-------------|-------------|------------|
| メソッドの数 | 14 | 56 | 156 | 58 |
| WSDLファイルの数 | 1 | 10 | 21 | 1 |
| エントリ取得 | Entrez | DBfetch | getEntry など | DBGET |
| 配列類似性検索 | | FASTA など | BLAST など | SSDB |
| 配列解析 | | EMBOSS etc. | | |
| モチーフ解析 | | InterPro | | SSDB |
| ゲノム情報 | | | GIB など | (KEGG DAS) |
| 立体構造 | | MSD | | |
| オントロジー | | Ontology | GO | |
| 文献 | PubMed | Citation | OMIM | |
| ケモインフォマティクス | | | | LIGAND |
| パスウェイ | | | | PATHWAY |

その他のウェブサービス

- BRENDA Web service - comprehensive enzyme information system
 - <http://www.brenda.uni-koeln.de/soap/>
 - NAR 2007 DB issue
- OLS - Ontology Lookup Service (OBO)
 - <http://www.ebi.ac.uk/ontology-lookup/>
 - BMC Bioinformatics 2006
- DICOM, HL7 - 医療画像 (CT, MRI, CR)、保健医療情報交換
- PathwayExplorer - 専用クライアントソフトからサーバ機能の呼び出し
 - <https://pathwayexplorer.genome.tugraz.at/>
 - NAR 2005 WS issue
- caCORE (caBIO) - cancer informatics
 - http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/caBIO
 - Bioinformatics 2003
- INCLUSive (2003), GeneCruiser (2005), maxdLoad2 (2005) - マイクロアレイ
... などなど ...



メリットとデメリット

ウェブサービスの利点

- 技術的な面

- クライアントのセットアップが容易 (サーバが増えてくると特に重要)
- 複数のサービスを組み合わせたマッシュアップが容易 (Web2.0)
- 80番ポート (HTTP) なので普通開いている
- プログラミング言語や OS に依存しない

- 実用的な面

- 常にバイオインフォの最新DBを利用可能
(もはや全データベースのコピーをローカルに維持することは困難)
- ローカルにツールをインストールする手間なくすぐに使える
(Solaris のバイナリしかないと、依存ライブラリ地獄とか)

(現状の) ウェブサービスの欠点

- 普及の遅れ
 - バイオインフォマティクスにおいてウェブサービスでできることが少ない
 - ウェブサービスで利用できないツールはまだまだ多い（シカモ増加中）
 - ウェブサービスで利用できないデータベースもまだまだ多い（シカモ増加中）
 - 逆に、独自インターフェイスのウェブアプリを作り込みがち
 - なので、プログラムが使いやすい形で背後のデータベースやツールにアクセスできない
- サーバとネットワークの負荷
 - サーバ側に負荷が集中する（くらい使われれば良いことではある）
 - 転送量が多い場合に口バストではないかも

(現状の) ウェブサービス**最大**の欠点

- ユーザインターフェイスの不備
 - ドキュメントがない（分かりにくい）ことが多い
 - メソッド名のポリシー、引数、戻り値の型がサービスごとにバラバラ
 - ジョブ ID 発行、ポーリングの仕組み等もサーバごとに独自の工夫

ウェブサービスを使ってみる

BioRuby シェル

- BioRuby をインストールすればついてくる
 - Ruby を対話的に利用可能
 - BioRuby の機能を使いやすく
 - ヒストリ、オブジェクトの永続化

```
% sudo gem install bioruby
% bioruby --rails
Loading config (shell/session/config) ... done
Loading object (shell/session/object) ... done
Loading history (shell/session/history) ... done

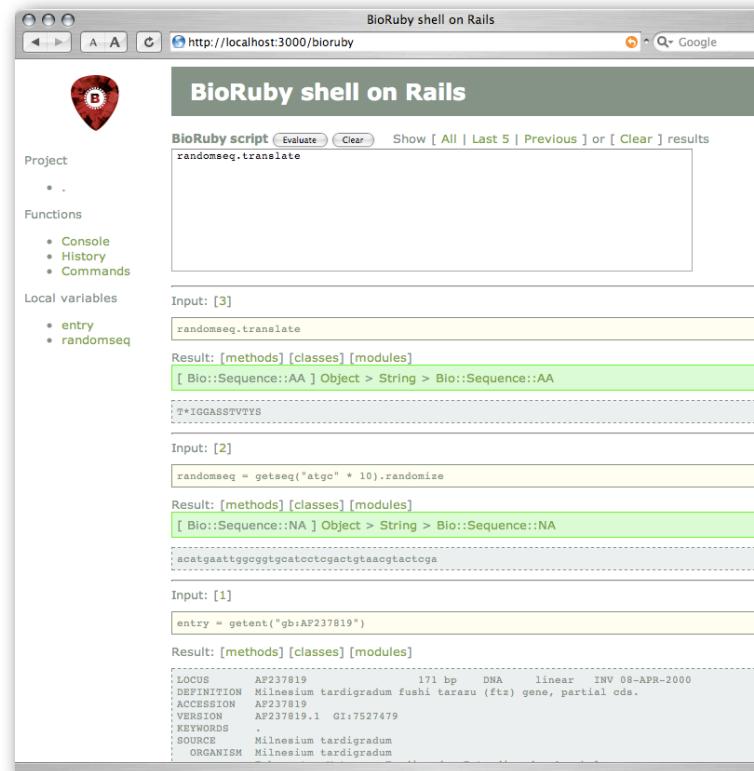
. . . Bi o R u b y i n t h e s h e l l . . .

Version : BioRuby 1.1.0 / Ruby 1.8.6

bioruby> randomseq = getseq("atgc" * 10).randomize
==> "tccctgtaaattcggaaacaaggctgcatcgattcgtcga"

bioruby> randomseq.translate
==> "SL*IREQGCIDSS"

bioruby> entry = getent("gb:AF237819")
LOCUS      AF237819                      171 bp    DNA     linear   INV 08-APR-2000
DEFINITION Milnesium tardigradum fushi tarazu (ftz) gene, partial cds.
ACCESSION  AF237819
VERSION   AF237819.1  GI:7527479
KEYWORDS
SOURCE    Milnesium tardigradum
ORGANISM  Milnesium tardigradum
```



KEGG API の場合

```
bioruby> kegg = Bio::KEGG::API.new
=> #<Bio::KEGG::API:0x3b90f34
  @wsdl="http://soap.genome.jp/KEGG.wsdl",
  @driver=#<SOAP::RPC::Driver:
    #<SOAP::RPC::Proxy:http://soap.genome.jp/keggapi/request_v6.1.cgi>>>

bioruby> kegg.list_methods
=> ["get_glycans_by_pathway", "get_motifs_by_gene", "get_reactions_by_enzyme", "list_organisms", "get_best_best_neighbors_by_gene", "color_pathway_by_elements",
"search_compounds_by_subcomp", "get_pathways_by_kos", "get_linked_pathways", "binfo", "get_enzymes_by_glycan", "search_drugs_by_mass", "list_databases",
"get_elements_by_pathway", "get_linkdb_by_entry", "search_glycans_by_mass", "mark_pathway_by_objects", "get_compounds_by_enzyme", "convert_mol_to_kcf",
"get_pathways_by_compounds", "get_enzymes_by_pathway", "get_html_of_colored_pathway_by_objects", "search_drugs_by_name", "get_enzymes_by_compound", "bconv",
"get_genes_by_ko", "get_genes_by_organism", "search_compounds_by_composition", "get_pathways_by_genes", "get_glycans_by_reaction",
"get_html_of_colored_pathway_by_elements", "search_glycans_by_kcam", "color_pathway_by_objects", "get_reactions_by_compound", "bfind", "list_ko_classes",
"search_compounds_by_name", "get_ko_by_gene", "get_genes_by_enzyme", "get_paralogs_by_gene", "get_pathways_by_glycans", "search_drugs_by_subcomp",
"get_reactions_by_pathway", "get_compounds_by_reaction", "bget", "get_element_relations_by_pathway", "search_glycans_by_composition", "get_pathways_by_reactions",
"get_kos_by_pathway", "search_drugs_by_composition", "get_compounds_by_pathway", "get_reverse_best_neighbors_by_gene", "get_reactions_by_glycan",
"search_compounds_by_mass", "get_enzymes_by_reaction", "get_genes_by_ko_class", "get_genes_by_motifs", "get_glycans_by_enzyme", "get_pathways_by_enzymes", "list_pathways",
"get_best_neighbors_by_gene", "get_html_of_marked_pathway_by_objects", "get_linkdb_between_databases", "get_genes_by_pathway", "get_number_of_genes_by_organism",
"search_glycans_by_name", "get_ko_by_ko_class", "bitit", "get_enzymes_by_gene"]

bioruby> kegg.get_genes_by_pathway("path:hsa00010")
=> ["hsa:10327", "hsa:124", "hsa:125", "hsa:126", "hsa:127", "hsa:128", "hsa:130", "hsa:130589", "hsa:131", "hsa:137872", "hsa:160287", "hsa:1737", "hsa:1738", "hsa:
2023", "hsa:2026", "hsa:2027", "hsa:217", "hsa:218", "hsa:219", "hsa:220", "hsa:2203", "hsa:221", "hsa:222", "hsa:223", "hsa:224", "hsa:226", "hsa:229", "hsa:230", "hsa:
2538", "hsa:2597", "hsa:2645", "hsa:2821", "hsa:3098", "hsa:3099", "hsa:3101", "hsa:3939", "hsa:3945", "hsa:3948", "hsa:441531", "hsa:501", "hsa:5160", "hsa:5161", "hsa:
5162", "hsa:5211", "hsa:5213", "hsa:5214", "hsa:5223", "hsa:5224", "hsa:5230", "hsa:5232", "hsa:5236", "hsa:5238", "hsa:5313", "hsa:5315", "hsa:55902", "hsa:669", "hsa:
7167", "hsa:84532", "hsa:8789", "hsa:92483", "hsa:97", "hsa:98"]

bioruby> puts kegg.bget("hsa:10327")
ENTRY      10327          CDS      H.sapiens
NAME       AKR1A1
DEFINITION aldo-keto reductase family 1, member A1 (aldehyde reductase)
           [EC:1.1.1.2]
ORTHOLOGY  KO: K00002 alcohol dehydrogenase (NADP+)
PATHWAY    PATH: hsa00010 Glycolysis / Gluconeogenesis
           PATH: hsa00561 Glycerolipid metabolism
           PATH: hsa00930 Caprolactam degradation
POSITION   1p33-p32
```

KEGG API の場合 - BioRuby との連携

```
biоруби> entry = kegg.bget("hsa:10327")
ENTRY      10327          CDS      H.sapiens
NAME       AKR1A1
DEFINITION aldo-keto reductase family 1, member A1 (aldehyde reductase)
[EC:1.1.1.2]
:
AASEQ      325
MAASCVLLHTGQKMPLIGLGTWKSEPGQVKAALKYALSVGYRHIDCAAIYGNEPEIGEAL
KEDVPGKAVPREELFVTSLWNTKHHPEDVEPALRKTADLQLEYLDLYLMHWPYAFER
GDNPFPKNADGTICYDSTHYKETWKALEALVAKGLVQALGLSNFNSRQIDDILSVASVRP
AVLQVECHPYLAQNELIAHCOARGLEVTAWSPLGSSDRAWRDPEPVLLLEEPVVLAEEK
YGRSPAQILLRWQVQRKVCIPIKSITPSRILQNIKVFDFTFSPEEMQQLNALNKNWRYIV
PMLTVDGKRVPRDAGHPLYPFNDPY
NTSEQ      978
atggcggcttcctgtttactgcacactggcagaagatgcctctgattggctgggt
:
biоруби> gene = Bio::KEGG::GENES.new(entry)

biоруби> gene.name
=> "AKR1A1"

biоруби> gene.definition
=> "aldo-keto reductase family 1, member A1 (aldehyde reductase) [EC:1.1.1.2]"

biоруби> gene.eclinks
=> ["1.1.1.2"]

biоруби> gene.aaseq
=>
"MAASCVLLHTGQKMPLIGLGTWKSEPGQVKAALKYALSVGYRHIDCAAIYGNEPEIGEALKE
DVGPKAVPREELFVTSLWNTKHHPEDVEPALRKTADLQLEYLDLYLMHWPYAFERGDNPFPK
NADGTICYDSTHYKETWKALEALVAKGLVQALGLSNFNSRQIDDILSVASVRP AVLQVECH
PYLAQNELIAHCOARGLEVTAWSPLGSSDRAWRDPEPVLLLEEPVVLAEEK YGRSPAQILL
RWQVQRKVCIPIKSITPSRILQNIKVFDFTFSPEEMQQLNALNKNWRYIV PMLTVDGKRV
PRDAGHPLYPFNDPY"
```

DDBJ XML の場合 - BLAST の場合

```
bioruby> ddbj = Bio::DDBJ::XML.new
==> #<Bio::DDBJ::XML::Blast:0x3dc5bb0
    @wsdl="http://xml.nig.ac.jp/wsdl/Blast.wsdl",
    @driver=#<SOAP::RPC::Driver:
        #<SOAP::RPC::Proxy:http://xml.nig.ac.jp/xddbj/Blast>>>

bioruby> result = ddbj.searchSimple("blastx", "DDBJ", gene.aaseq)
==> "The search and analysis service is very busy now. Please try it again.\n"

... そんな日もあるさ ...

bioruby> result= serv.searchSimple("blastp", "SWISS", gene.aaseq)
/usr/local/lib/ruby/1.8/timeout.rb:54:in `rbuf_fill': execution expired (Timeout::Error)
    from /usr/local/lib/ruby/1.8/timeout.rb:56:in `timeout'
    from /usr/local/lib/ruby/1.8/timeout.rb:76:in `timeout'
    :

... そんなこともあるさ ...
```

EBI の場合 - Dbfetch の例

```
bioruby> ebi = Bio::EBI::SOAP.new
==> #<Bio::EBI::SOAP:0x26ba678
      @wsdl="http://www.ebi.ac.uk/Tools/webservices/wsdl/WSDbfetch.wsdl",
      @driver=#<SOAP::RPC::Driver:
      #<SOAP::RPC::Proxy:http://www.ebi.ac.uk/ws/services/WSDbfetch>>>

bioruby> ebi.list_methods
==> [ "getFormatStyles", "fetchData", "getSupportedDBs", "getDbFormats", "getSupportedFormats",
      "getSupportedStyles", "fetchBatch"]

bioruby> ebi.getSupportedDBs
==> [ "embl", "emblann", "emblcds", "emblcon", "emblsva", "epo_prt", "genomereviews", "hgvbase",
      "interpro", "ipi", "jpo_prt", "medline", "pdb", "refseq", "refseqp", "uniparc", "uniprotkb",
      "uniref100", "uniref50", "uniref90", "unisave", "uspto_prt"]

bioruby> puts ebi.fetchData("embl:BUM", "embl", "raw")
ID  J02231; SV 1; linear; genomic RNA; STD; VRL; 200 BP.
XX
AC  J02231;
XX
DT  29-AUG-2003 (Rel. 77, Created)
DT  29-AUG-2003 (Rel. 77, Last updated, Version 1)
XX
DE  La Crosse virus isolate 174 m-rna 3' sequence.
XX
KW  .
XX
OS  La Crosse virus
OC  Viruses; ssRNA negative-strand viruses; Bunyaviridae; Orthobunyavirus.
XX
RN  [1]
```

EBI の場合 - EMBOSS の例

```
biорубы> emboss = Bio::EBI::SOAP::Emboss.new
==> #<Bio::EBI::SOAP::Emboss:0x3ca7198
    @wsdl="http://www.ebi.ac.uk/Tools/webservices/wsdl/WSEmboss.wsdl",
    @driver=#<SOAP::RPC::Driver:
        #<SOAP::RPC::Proxy:http://www.ebi.ac.uk/cgi-bin/webservices/WSEmboss>>>

biорубы> emboss.list_methods
==> [ "poll", "run", "checkStatus", "getResults", "getTools", "getInfo" ]

biорубы> hash = {
    "tool"      => "water",
    "asequence" => "uniprot:alk1_human",
    "bsequence" => "uniprot:alk1_mouse",
    "email"     => "ebisoap@example.org"
}

biорубы> poll = emboss.run(hash, [])
==> "emboss-20070322-17172499"

biорубы> emboss.checkStatus(poll)
==> "RUNNING"

... 待てど暮らせど終わらなかつたり ...

biорубы> base = emboss.poll(poll, "tooloutput")
biорубы> puts Base64.decode64(base)
```

NCBI の場合 - EUtils の例

```
bioruby> ncbi = Bio::NCBI::SOAP.new
ignored element: {http://www.w3.org/2001/XMLSchema}sequence of WSDL::XMLSchema::Sequence
ignored element: {http://www.w3.org/2001/XMLSchema}choice of WSDL::XMLSchema::Sequence
:
==> #<Bio::NCBI::SOAP:0x2458100
  @wsdl="http://www.ncbi.nlm.nih.gov/entrez/eutils/soap/eutils.wsdl",
  @driver=#<SOAP::RPC::Driver:
    #<SOAP::Proxy:http://www.ncbi.nlm.nih.gov/entrez/eutils/soap/soap_adapter_1_5.cgi>>>

bioruby> ncbi.list_methods
==> ["run_eSpell", "run_eFetch", "run_eSummary", "run_eInfo", "run_eSearch", "run_eGquery", "run_eLink"]

bioruby> result = ncbi.run_eSearch("db" => "protein", "term" => "pparg human")
==> #<SOAP::Mapping::Object:0x1220fdc {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Count="94" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}
RetMax="20" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}RetStart="0" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}IdList=#<SOAP::Mapping::Object:
0x121fc72 {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Id=[ "13432234", "116284373", "116284370", "116284368", "20336229", "13905056", "50871915",
"47678891", "119584529", "119584528", "86451955", "66471678", "17978516", "66471844", "66471842", "48762805", "109157724", "109157723",
"99031685"]> {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}TranslationSet=#<SOAP::Mapping::Object:0x121c158 {http://www.ncbi.nlm.nih.gov/soap/eutils/
esearch}Translation=<SOAP::Mapping::Object:0x121bf78 {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}From="human" {http://www.ncbi.nlm.nih.gov/soap/
eutils/esearch}To="( \"Homo sapiens\"[Organism] OR human)">> {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}
TranslationStack=#<SOAP::Mapping::Object:0x121aa7e {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}TermSet=[#<SOAP::Mapping::Object:0x121a808 {http://
www.ncbi.nlm.nih.gov/soap/eutils/esearch}Term="pparg[All Fields]" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Field="All Fields" {http://
www.ncbi.nlm.nih.gov/soap/eutils/esearch}Count="165" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Explode="Y">, #<SOAP::Mapping::Object:0x1218634
{http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Term="\\"Homo sapiens\"[Organism]" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}
Field="Organism" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Count="393224" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Explode="Y">,
#<SOAP::Mapping::Object:0x1216ad2 {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Term="human[All Fields]" {http://www.ncbi.nlm.nih.gov/soap/eutils/
esearch}Field="All Fields" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Count="1190298" {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}Explode="Y">]
{http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}OP=[ "OR", "GROUP", "AND", "GROUP" ]> {http://www.ncbi.nlm.nih.gov/soap/eutils/esearch}
QueryTranslation="pparg[All Fields] AND (\\"Homo sapiens\"[Organism] OR human[All Fields])>

...
なんじやこりや ...

bioruby> result.__xmlele[3][1].__xmlele[0][1].first
==> "13432234"
```

caCORE の場合

```
biорубы> wsdl = "http://cabio.nci.nih.gov/cacore32/ws/caCOREService?wsdl"
==> "http://cabio.nci.nih.gov/cacore32/ws/caCOREService?wsdl"

biорубы> cabio = Bio::SOAPWSSDL.new(wsdl)
==> #<Bio::SOAPWSSDL:0x2407fd4
     @wsdl="http://cabio.nci.nih.gov/cacore32/ws/caCOREService?wsdl",
     @driver=#<SOAP::RPC::Driver:
      #<SOAP::RPC::Proxy:http://cabio.nci.nih.gov/cacore32/ws/caCOREService>>>

biорубы> cabio.list_methods
==> [ "query", "getProcessOntology", "getDataObject", "getMaximumRecordsPerQuery", "exist",
      "getRecordsPerQuery", "getVersion", "getTotalNumberOfRecords", "queryObject"]

... ここでマニュアルを読む ...

biорубы> cabio.queryObject("IL*")
SOAP::Mapping::MappingError: nil not allowed: targetClassName
  from /usr/local/lib/ruby/1.8/soap/mapping/wsdlliteralregistry.rb:133:in `complexobj2soap'
  from /usr/local/lib/ruby/1.8/wsdl/xmlSchema/complexType.rb:56:in `each_element'
  :
```

挫折

BRENDA の場合

```
bioruby> wsdl = "http://www.brenda.uni-koeln.de/soap/brenda.wsdl"
=> "http://www.brenda.uni-koeln.de/soap/brenda.wsdl"

bioruby> brenda = Bio::SOAPWSSDL.new(wsdl)
ignored element: {http://www.w3.org/2001/XMLSchema}complexType
RuntimeError: part: id cannot be resolved
  from /usr/local/lib/ruby/1.8/wsdl/soap/methodDefCreator.rb:144:in `rpcdefinedtype'
  from /usr/local/lib/ruby/1.8/wsdl/soap/methodDefCreator.rb:49:in `collect_rpcparameter'
  from /usr/local/lib/ruby/1.8/wsdl/soap/methodDefCreator.rb:47:in `collect'
  from /usr/local/lib/ruby/1.8/wsdl/soap/methodDefCreator.rb:47:in `collect_rpcparameter'
  from /usr/local/lib/ruby/1.8/soap/wsdlDriver.rb:130:in `create_param_def'
  from /usr/local/lib/ruby/1.8/soap/wsdlDriver.rb:101:in `add_operation'
  from /usr/local/lib/ruby/1.8/xsd/namedelements.rb:58:in `each'
  from /usr/local/lib/ruby/1.8/xsd/namedelements.rb:57:in `each'
  from /usr/local/lib/ruby/1.8/soap/wsdlDriver.rb:96:in `add_operation'
  from /usr/local/lib/ruby/1.8/soap/wsdlDriver.rb:40:in `create_rpc_driver'
  from /Users/k/lib/ruby/bio/io/soapwsdl.rb:63:in `create_driver'
  from /Users/k/lib/ruby/bio/io/soapwsdl.rb:57:in `initialize'
  from (irb):22:in `new'
  from (irb):22
  :
...
.... WSDL が読み込めない ...
```

流しましょう

まあ、Rubyだけの問題かもしれないし

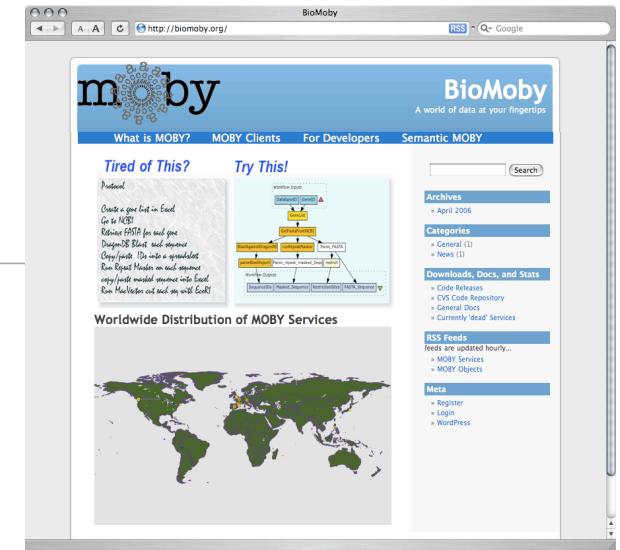


標準化の必要性

- Web Services Interoperability Organization (WS-I)
 - <http://www.ws-i.org/>
 - WS-I is an open industry organization chartered to promote Web services interoperability across platforms, operating systems and programming languages. The organization's diverse community of Web services leaders helps customers to develop interoperable Web services **by providing guidance, recommended practices and supporting resources**. All companies interested in promoting Web services interoperability are encouraged to join the effort.
- バイオインフォマティクスにおいても国際コンソーシアムが必要かもしれない
 - ガイドラインの制定（メソッド名、オブジェクト型、ポーリング等）
 - BioMOBY : サービスのレポジトリとオブジェクトの種類の標準化 The logo for BioMOBY, featuring the word "BioMOBY" in a stylized, blocky font where the letters are composed of small circles and dots.
 - SoapLab : コマンドラインプログラムを SOAP 化するラッパー The logo for SoapLab, showing the word "SoapLab" in a stylized font surrounded by a circular pattern of green and grey dots.

BioMOBY plan

- BioMOBY (Model Organism Bring Your own)
 - 2001年～
 - <http://biomoby.org/>
- バイオインフォで使われるデータ型のオントロジーを構築
 - オントロジーのシリアルライズ方法を規定
 - オントロジーを利用したAPIを作成
 - ウェブサービスのI/Oとオントロジーの対応
 - オントロジーに応じてサービスをレジストリに登録
- これにより：
 - コンピュータが適切なサービスを見つけることができる
 - Machines can execute that service anattended
 - オントロジーはコミュニティによって拡張可能



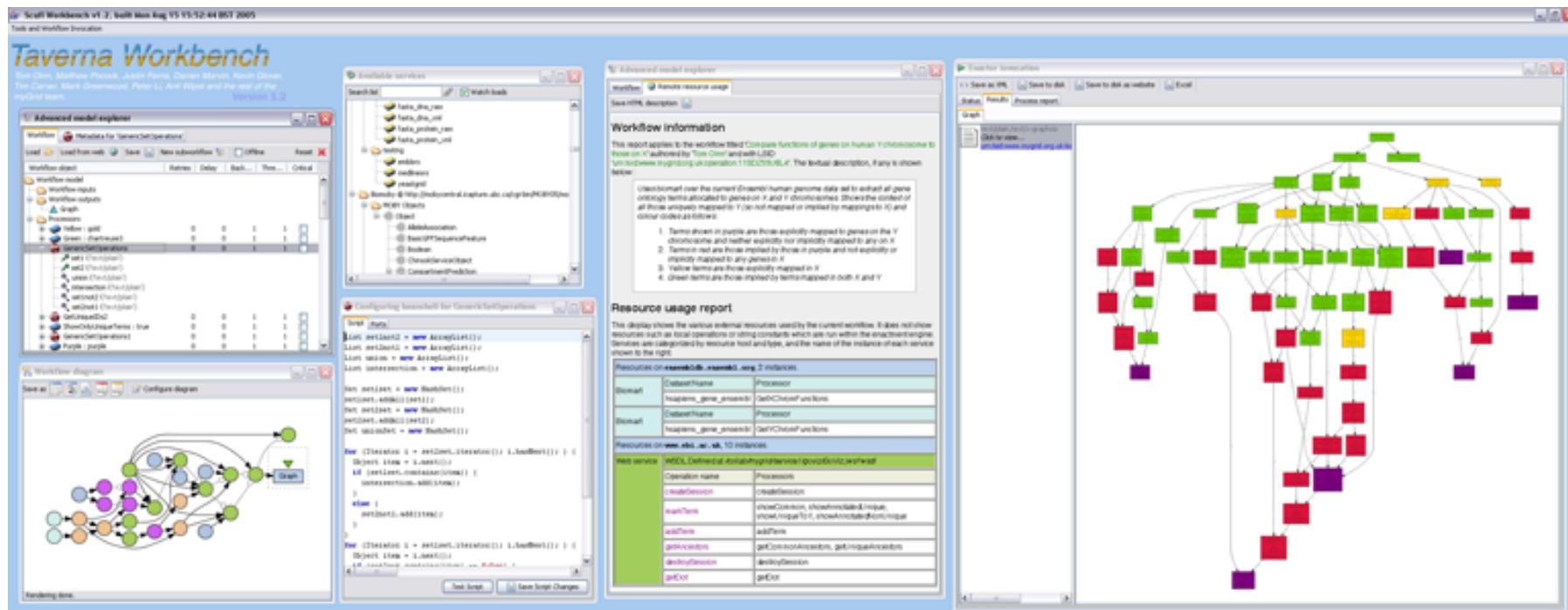
バイオインフォ 2.0

ウェブサービスが増えると

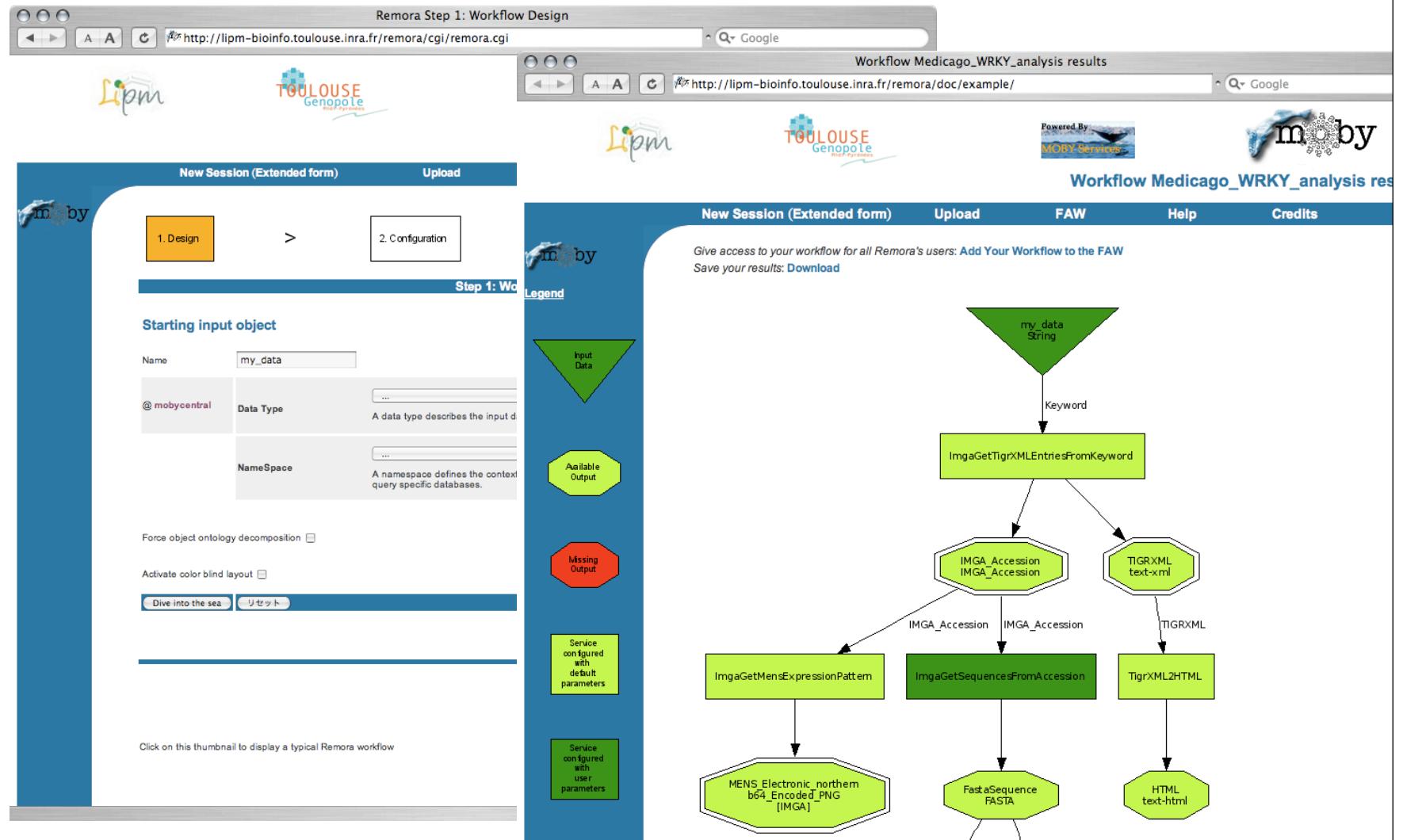
- いつもの解析手順の全てのステップをワークフロー化できる
 - DB も解析アプリも常に最新
 - 他の研究者との共有、検証が容易に
 - 論文のセマンティクスを表現
- マッシュアップされた便利なサイトやツールが続々登場する（かも）
 - Taverna
 - Remora
 - Yahoo! Pipes

Taverna

- ワークフロー構築の GUI
 - <http://taverna.sf.net>



Remora - ブラウザ上でワークフローを構築



Yahoo! Pipes - <canvas> タグがすごい

Pipes: editing 'IT News Search'
http://pipes.yahoo.com/pipes/PoSJlvi72xGDHCpijwtvUw/edit

Wikipedia (日本語)

Search for pipes and feeds

Back to My Pipes New Save a copy

IT News Search

Layout Expand All Collapse All

Sources

- Yahoo! Search
- Yahoo! Local
- Fetch
- Google Base
- Flickr

User inputs

Operators

Url

String

Date

My pipes

Fetch

- URL
- http://rss.rssad.jp/rss/tm/rss.xml
- http://rss.itmedia.co.jp/rss/1.0/
- http://rss.itmedia.co.jp/rss/1.0/

Fetch

- URL
- http://itpro.nikkeibp.co.jp/rss/1/
- http://itpro.nikkeibp.co.jp/rss/0/

Fetch

- URL
- http://feed.japan.cnet.com/rss/
- http://japan.zdnet.com/news/

Union

Union

Filter

Permit items that match all of the following

Rules

title Contains text

キーワードを入力 (text)

Name: textinput1
Prompt: キーワードを入力
Position: number
Default:
Debug:

Pipe Output

Time taken: 4.908649s Refresh

▶ Apple、アップデートでiChatなどの脆弱性に対処

▶ EMCの買収攻勢はひと休み？ トウチチceoが示唆

▶ Firefoxに脆弱性、他者サイト改ざんの恐れも

Debugger: Filter (152 items)

Clear Inspect Options

Console Debugger Inspector

```
> <div class="mod" id="mod15" ismodule="true" style="width: 174px; left: 272px; top: 174px; visibility: visible;" hasobj="true">
>   <div class="mod" id="mod16" ismodule="true" style="width: 174px; left: 340px; top: 237px; visibility: visible;" hasobj="true">
>     <div class="mod moduleselected" id="mod17" ismodule="true" style="width: 418px; left: 174px; top: 302px; visibility: visible;" hasobj="true">
>       <div class="mod" id="mod18" ismodule="true" style="width: 188px; left: 655px; top: 205px; visibility: visible;" hasobj="true">
>         <canvas class="wire" width="196" height="177" style="left: 137px; top: 70px; width: 196px; height: 177px;"/>
>         <canvas class="wire" width="12" height="61" style="left: 385px; top: 117px; width: 12px; height: 61px;"/>
>         <canvas class="wire" width="219" height="124" style="left: 453px; top: 117px; width: 219px; height: 124px;"/>
>         <canvas class="wire" width="46" height="41" style="left: 355px; top: 200px; width: 46px; height: 41px;"/>
>         <canvas class="wire" width="52" height="43" style="left: 379px; top: 263px; width: 52px; height: 43px;"/>
```

/html/body/div[3]/div[2]/div

完了

Source Style Layout Events DOM

ノートブックを開く

普及を妨げているものは何か？

パラダイム

- バイオインフォのワークフローに必須といえるほど普及が進んでいない
(クライアントがウェブサービスを知らないため要求仕様に含まれないという面も)
- バイオインフォでウェブサービスを構築できる技術者も不足気味?
(これまでDBの情報公開といえば小綺麗なウェブサイト構築だった)



テクノロジー

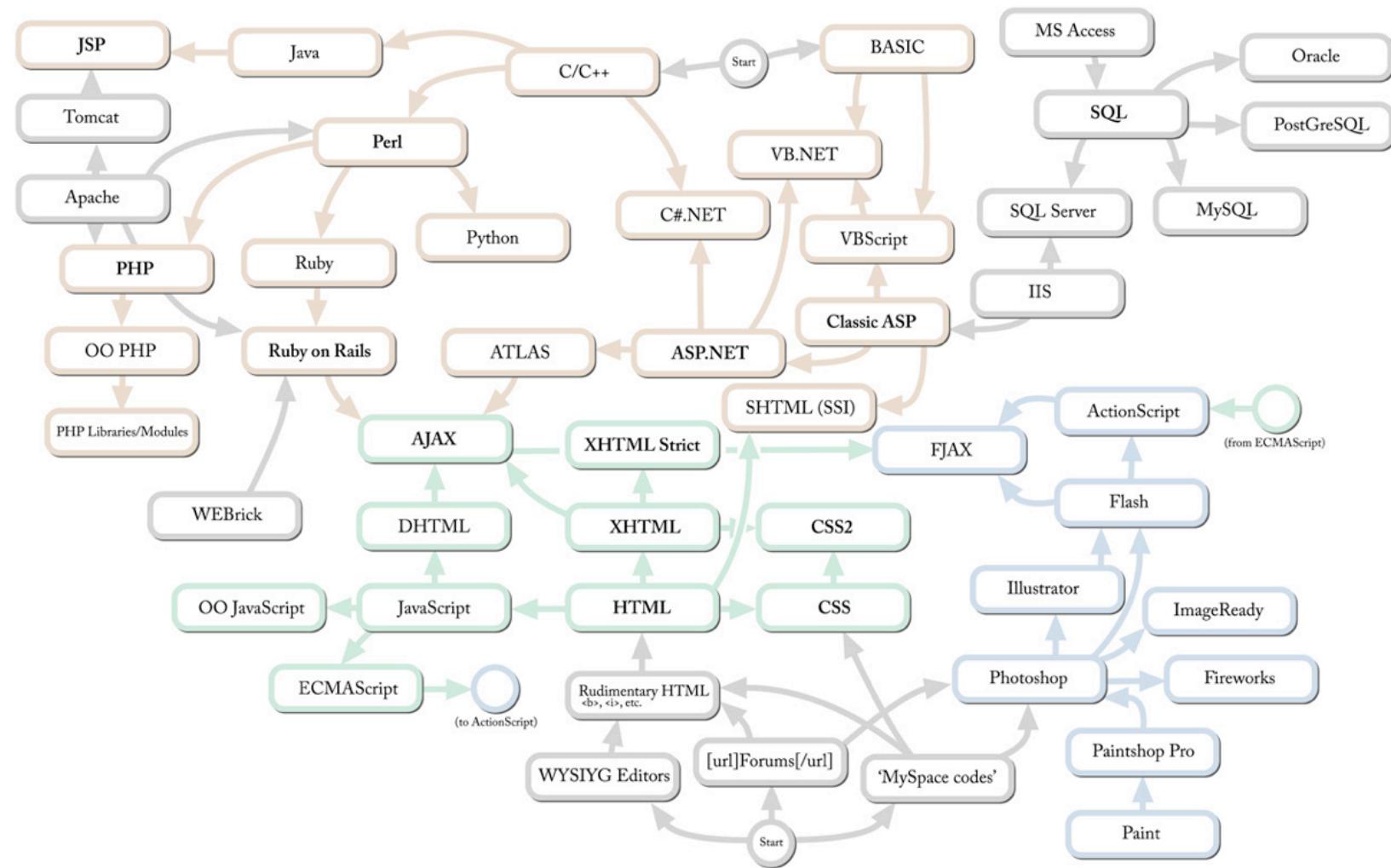
- ウェブサービスを提供するための手順が不明
 - 利用可能なフレームワークが言語ごとにさまざま
 - Ruby - Ruby on Rails + Action Web Service (SOAP4R)
 - Perl - SOAP::Lite
 - Java - Apache Axis
 - (使えない WSDL があったりすることも)
- 移行のための情報不足
 - どれくらい大変な作業になるか見積もれない
 - 名前や型など決めなければいけないことが多い (しデファクトもない)

打開策としてのテンプレートとドキュメント整備

- すぐに使える、真似してわかる
 - 実際に動く実例を提供する
 - 典型的な具体例をドキュメント化する
- テンプレートをカスタマイズして利用するサイトが増加
 - ウェブサービスで利用できるサイトが充実
- 同じフレームワークを使う
 - ノウハウや情報を共有できる
 - 全てを自作するより開発自体も効率化される

小規模サイトこそ凝った独自インターフェイスよりウェブサービスで使いたい

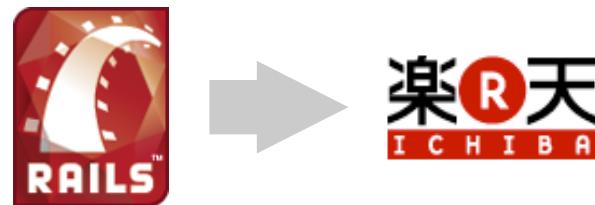
Web Technology Family Tree



Ruby on Rails

- ウェブアプリ開発のフレームワーク
 - <http://www.rubyonrails.org/>

- Ruby on Rails ならサーバ構築はカンタン
 - RDBとのO/Rマッピング内蔵
 - ウェブサービス内蔵
- 他のフレームワークがいい、という人
 - すでに技術力がある → ご自由に



The screenshot displays a news article from CNET Japan. The header reads "cnet CNET Japan IDEAS FOR INNOVATION". The main headline is "楽天、ネットサービス開発で「Ruby on Rails」を採用--NaClが全面協力". Below the headline, the author is listed as "吉澤亨史" and the date as "2007/03/15 22:42". At the bottom, there are social sharing links for Trackback, Comments, Bookmarks, and Print.

楽天は3月15日、自社が提供するインターネットサービスの開発に、ウェブアプリケーションの開発フレームワーク「Ruby on Rails」を採用、利用を開始したと発表した。

テンプレートの仕様

- Ruby on Rails + Action Web Service で構築
- バイオ系の主要 DB やアプリは最初からサポート
- 作業工程をドキュメントにして公開

RoR による WS 構築のデモ

どれくらい簡単か

- MyクマムシDBのサービスを構築
 - 実行例と wiredump

クマムシ - 乾燥すると究極の極限環境耐性



- 高温 : +151°C (ラーム, 1921)
- 低温 : -273°C (ベクレル, 1950)
- 高圧 : 7.5GPa (750,000mの水圧)
- 低圧 : 5×10^{-4} Pa (ほぼ真空)
- 放射線 : X線 100,000 Gy
(ヒト致死量の1,000倍)
ガンマ線 7,000Gy
Heイオン 8,000Gy
- 化学物質 : アルコール (消毒剤)
臭化メチル (防虫剤)

RoR による WS 構築のデモムービー

テンプレートに求められるもの

- Bio 系オブジェクトのモデル化
 - 充実させること
 - おしつけ！（雨後のタケノコのように林立する類似の型をなんとかしたい）
 - ありがちなツールへの対応
 - BLAST などの実行と結果の整理
- BioRuby でいいのかも？
- メソッド名に統一的なルール
 - Rails 風の規約 - CoC
 - 覚えることは少なく、コード量も少なく

BioRubyの進化

- 第0世代：GenBank パーザ (2000)
 - ↓
- 第1世代：各種 DB とアプリを扱うライブラリ集 (2001-2004)
 - ↓
- 第2世代：BioRuby シェル - CUI (2005 : IPA[®] 未踏ソフト)
 - ↓
- 第3世代：Ruby on Rails を無理やり内蔵 - GUI (2006)
 - ↓
- 第4世代：シェル機能を Ruby on Rails のプラグイン化 (2007)
 - ↓
- 第5世代：ワークフローのできるウェブサービス環境？

相互運用性はバイオインフォだけの問題か

- サーバ資源の共有
 - 日々生み出されるゲノム情報に匹敵する巨大なデータ
 - 世界中に分散したデータベースや様々なアプリケーション
- Phyloinformatics
 - 大量の tree の計算
 - CIPRES - CORBA ベース
- 気象、宇宙、物理などでも同様の状況はありそう
 - 電脳Ruby - Gfdnavi (Rails で扱う地球流体DB, 解析, 可視化)

今後の予定

- 賛同者がいればテンプレート構築開始
(とりあえず「KEGG on Rails」や「BioRuby シェル on Rails」のプロモーション)
- 国際交流と標準化の必要性
 - BioMOBY
 - セマンティックウェブとか?

The screenshot shows a web browser window with the title "NETTAB 2007 Workshop". The address bar displays the URL <http://www.nettab.org/2007/>. The page header features the NETTAB logo, which includes a stylized orange and blue ribbon-like graphic and the text "NETTAB". Below the logo, the tagline "Network Tools and Applications in Biology" is visible. The top navigation bar contains several logos of partner institutions: Università di Camerino (UNICAM), Istituto Superiore di Sanita (ISS), University of Pisa (UIP), and W3C Italia. To the right of the logos, the text "NETTAB 2007 focused on:" is followed by the main theme: "**A Semantic Web for Bioinformatics: Goals, Tools, Systems, Applications**". Below this, the specific dates and location are mentioned: "June 12-15, 2007, University of Pisa, Italy". At the bottom of the page, there is a welcome message: "Welcome to the web site of the NETTAB 2007 workshop!", a "Home Page" link, and two red links for the "Scientific Programme" (with sub-links "Overview - Invited Speakers" and "Programme - Abstracts").