WEB API: WHY THEY MATTER

ECOL 453/553 2012
Nirav Merchant
nirav@email.arizona.edu
WHAT IS A WEB API

- API: Application Programming Interface
- Calling API over the web (http)
- What is this http thing you keep talking about?
- So why do you want to call API over the web?
  - Data size
  - Data Complexity
  - Linked data
  - Specialized databases and data structure
- It allows for automation (mashup, workflows)
- How is this different then a perl module?
WHERE HAVE YOU SEEN WEB API

- Internet apps
- Science apps
- [http://www.programmableweb.com/apis/directory/1?apicat=Medical&protocol=REST](http://www.programmableweb.com/apis/directory/1?apicat=Medical&protocol=REST)
- We will use EBI and TogoWS for our class today

- Getting started …
TogoWS: integrated SOAP and REST APIs for interoperable bioinformatics Web services

Toshiaki Katayama, 1,* Mitsutenu Nakao, 2,3,* and Toshihisa Takagi 2,4

ABSTRACT

Web services have become widely used in bioinformatics analysis, but there exist incompatibilities in interfaces and data types, which prevent users from making full use of a combination of these services. Therefore, we have developed the TogoWS service to provide an integrated interface with advanced features. In the TogoWS REST (REpresentative State Transfer) API (application programming interface), we introduce a unified access method for major database resources through intuitive URLs that can be used to search, retrieve, parse and convert the database entries. The TogoWS SOAP API resolves compatibility issues found on the server and client-side SOAP implementations. The TogoWS service is freely available at:

http://togows.dbcls.jp/

INTRODUCTION

In recent years, major bioinformatics centers have begun providing SOAP-based Web services that enable users to use these database resources with client programs in an automated manner. These include the E-Utilities service (1) provided by the National Center for Biotechnology Information (NCBI), Web services provided by the European Bioinformatics Institute (EBI) (2,3), the Web API for Bioinformatics (WABI) from the DNA Data Bank of Japan (DDBJ) (4–7), the Protein Data Bank Japan’s (PDBj) Web services (8) and the KEGG API service from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (9). Thanks to these services, users can easily perform various bioinformatics tasks through their choice of client software and can reproduce each procedure as a workflow.
SOME TERMS

- RESTful (Representational State Transfer)
- Architectural Styles and the Design of Network-based Software Architectures
  Fielding, Roy Thomas (2000)
  Doctoral dissertation, University of California, Irvine

Some concepts

- GET
- POST
- PUT
- DELETE
JSON (AND SOME XML)

- JSON (JavaScript Object Notation)
  lightweight data-interchange format
- It is easy for humans to read and write
- It is easy for machines to parse and generate
- It is based on a subset of the JavaScript Programming Language

Built on two structures:
- A collection of name/value pairs.
  (object, record, struct, dictionary, hash table, keyed list, or associative array)
- An ordered list of values
  (array, vector, list, or sequence)
```json
{
    "firstName": "John",
    "lastName": "Smith",
    "age": 25,
    "address": {
        "streetAddress": "21 2nd Street",
        "city": "New York",
        "state": "NY",
        "postalCode": "10021"
    },
    "phoneNumber": [
        {
            "type": "home",
            "number": "212 555-1234"
        },
        {
            "type": "fax",
            "number": "646 555-4567"
        }
    ]
}
```
EXERCISE (MORE TERMS)

- Go to http://togows.dbcls.jp/site/en/rest.html
- Let’s follow some URL
- Now can you do this for 100 id’s?
- Log on to your HPC account
- Utilities you need
  - wget
  - curl
  - JSON pretty printing: python –mjson.tool
  - Unix pipe | to connect things together
## Ensembl REST API Endpoints

### Comparative Genomics

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET genetree/id:id</td>
<td>Retrieves Gene Tree dumps for a given Gene Tree stable identifier</td>
</tr>
<tr>
<td>GET homology/id:id</td>
<td>Retrieves homology information by ensembl gene id</td>
</tr>
<tr>
<td>GET homology/symbol/:species/:symbol</td>
<td>Retrieves homology information by symbol</td>
</tr>
</tbody>
</table>

### Cross References

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET xrefs/id:id</td>
<td>Perform lookups of Ensembl Identifiers and retrieve their external cross reference</td>
</tr>
<tr>
<td>GET xrefs/name/:species/:name</td>
<td>Performs a lookup based upon the primary accession or display label of an external entry</td>
</tr>
<tr>
<td>GET xrefs/symbol/:species/:symbol</td>
<td>Looks up an external symbol and returns all Ensembl objects linked to it. This can be a synonym or an externally linked reference. If a Gene’s transcript is linked to the Transcript (it supports transient links).</td>
</tr>
</tbody>
</table>

### Features
## Example Requests

### /feature/region/human/7:140424943-140624564?
```plaintext
feature=gene;feature=transcript;feature=cds;feature=exon;content-type=application/json
```

### /feature/region/human/7:140424943-140624564?
```plaintext
feature=gene;feature=transcript;feature=cds;feature=exon;content-type=text/x-gff3
```

1. `wget -q --header='Content-type:application/json' 'http://beta.rest.ensembl.org/feature/region/human/7:140424943-140624564?feature=gene;feature=transcript;feature=cds;feature=exon' -O -`

### Example output

```json
{
  "ID": "ENSG00000207040",
  "logic_name": "ncrna"
}
```
What modules do we need
use HTTP::Tiny;
use JSON;
use strict;
use warnings;
use Data::Dumper;
LETS GET THE ONES WE ARE MISSING

► http://search.cpan.org
HOW DO WE GET THE MODULE

- wget to rescue
- Copy and paste url!
- tar –xvzf <downloaded file>
- Make a directory in your home perl-module
- Go to downloaded dir (HTTP-Tiny-0.024)
- perl Makefile.PL PREFIX=~/perl-module
- make
- make test
- make install
- check perl-module for installed files!
HOW DO YOU USE IT

- setenv PERL5LIB ~/perl-module/share/perl5/
- Now you can run the program
- To make your life easy I have already built the modules that you can use
- Its in /genome/perl-lib/share/perl5/

SO:

- set PERL5LIB /genome/perl-lib/share/perl5/

Now you can run the sample perl scripts from ensembl rest services!