Taverna Xpath service
Getting Data out of XML

Stian Soiland-Reyes and Christian Brenninkmeijer
University of Manchester
material by Mr Alan Williams

http://orcid.org/0000-0001-9842-9718
http://orcid.org/0000-0002-2937-7819
http://orcid.org/0000-0003-3156-2105

Bonn University, 2014-09-01 / 2014-09-03
http://www.taverna.org.uk/

This work is licensed under a
Creative Commons Attribution 3.0 Unported License
XML – eXtensible Markup Language

- Designed for the storage and transport of data
  - This includes passing data between services or retrieving data from a Web page

- Provides a machine readable dataset

- Many service providers export data in XML
<?xml version="1.0"?>
<note>
  <to>Katy</to>
  <from>Helen</from>
  <heading>Reminder</heading>
  <body>Don't forget about Bonn Trip!</body>
</note>

- The following website has lost of information about XML and tutorials:  http://www.w3schools.com/xml
Extracting XML Elements

- Often, we need to extract particular elements of the XML to feed into the next services in the workflow.
- Taverna has a dedicated service to help you do this.
- Go to the service panel and find “Xpath Service” in the Service Templates section.
- We will use this service to explore the contents of an SBML file from the BioModels database, but first we need to find the BioModels services.
SBML is the Systems Biology Mark-up Language. It is the community standard for describing systems biology models.

SBML models contain information about the reactions between enzymes, genes, and metabolites, and their interactions in living systems.

Well-annotated SBML models will contain descriptions of model entities and official IDs to link those entities back to their sources (e.g. Uniprot IDs for proteins/enzymes, KEGG IDs or Reactome IDs for reactions etc.)
We will find all the BioModels constructed in yeast (Saccharomyces cerevisiae), and find the publications about them.

In Biocatalogue, search for the ‘BioModels’:
- `getModelsIDByTaxonomy`
- `getModelSBMLById`

And add these services to the Service Panel and your workflow.

Hint: As you did in the REST exercise.
Extracting XML Elements from SBML

- Add a input port (rename the service to TaxonomySearch) and connect it with getModelsIdByTaxonomy
- Then connect the output (Return) from getModelsIdByTaxonomy with the input for getModelSBMLById
- Create an output port
The workflow should look like this
Run the workflow using the value *Saccharomyces cerevisiae*

*Hint: As you only need 1 Value you can “Cancel” the workflow as soon as a few Values appear*

Save one of the SBML models as an XML file

*Hint: In the results view select one of the Values from the List for example “Value 1”

*Click on “Save Value”

■ Using a .xml file extension will make it easier to find
Now import the Xpath Service from Taverna

In the pop-up window, enter the path to the saved SBML file and click the green arrow to load it into the service

By expanding the + icons, you can explore the file

Expand the XML under ‘annotation’, until you reach RDF:resource
Extracting XML Elements from SBML

- Click on ‘Generate Xpath From Expression’
- Then Click on ‘Run Xpath’ and click ‘Apply’
- This service will now pull out all the reactions from each model
Extracting XML Elements from SBML

- Connect it to the output of ‘getModelSBMLByld’
- Add an output port to the Xpath service “nodeListAsXML”
- Save and run the workflow again
We can change the output of the Xpath service into a flattened list.

(In Design View) Go to the Services Panel and add a Flatten List (from “Available Services”/Local Services”/“list”).
Connect the output of the Xpath service to the “Flatten List”

Connect the output of the “Flatten List” with the output port
Run the workflow again

- Note: Flatten_List will not show any values until the workflow is finished as its output is a single list

Optional: Try the service with a different Xpath Query

- In Design View right click the Xpath Service
- Select Configure Xpath Service
- This brings back up the configuration Screen you saw earlier.