Web Forms and SPARQL Queries

Bob DuCharme

Semantic Tech and Business Conference
Washington D.C.

November 30, 2011

TopQuadrant™

Introductions

- Presentation and all its URLs:
  http://snee.com/semtech/2011/dc/
- Me: SGML and XML at Moody's, LexisNexis, Innodata Isogen, TopQuadrant
- Weblog: http://www.snee.com/bobdc.blog
- Twitter: @bobdc

SPARQL

- **S**PARQL **P**rotocol **a**nd **R**DF **Q**uery **L**anguage

- “Trying to use the Semantic Web without SPARQL is like trying to use a relational database without SQL” – Tim Berners-Lee

- **SPARQL Endpoint**: a (web) service that accepts a SPARQL query and returns the result set

Learning SPARQL: The Book

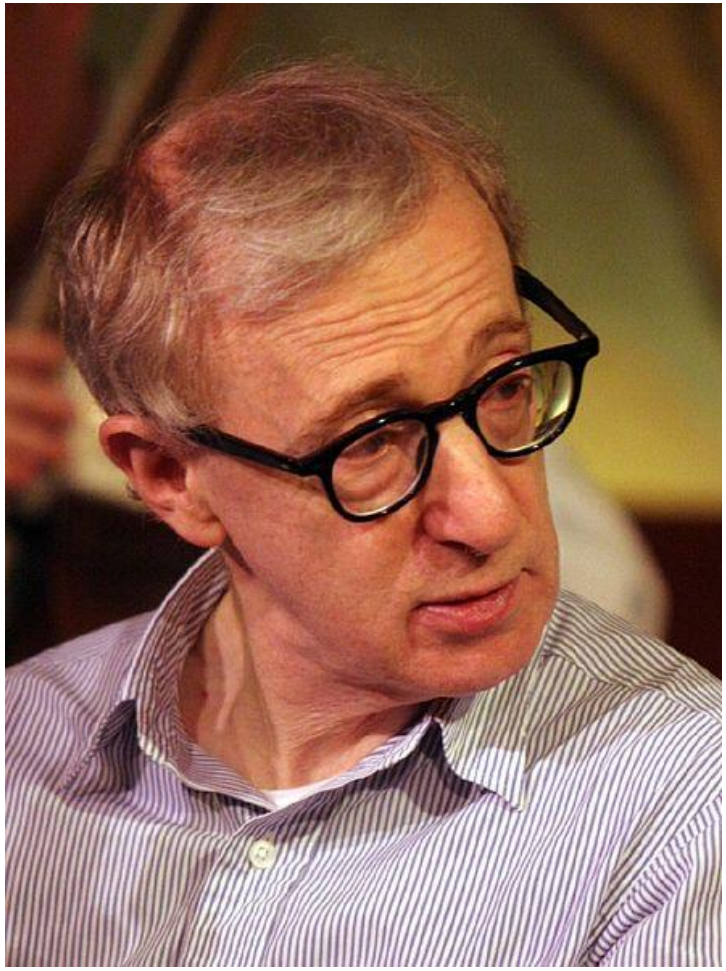
learningsparql.com

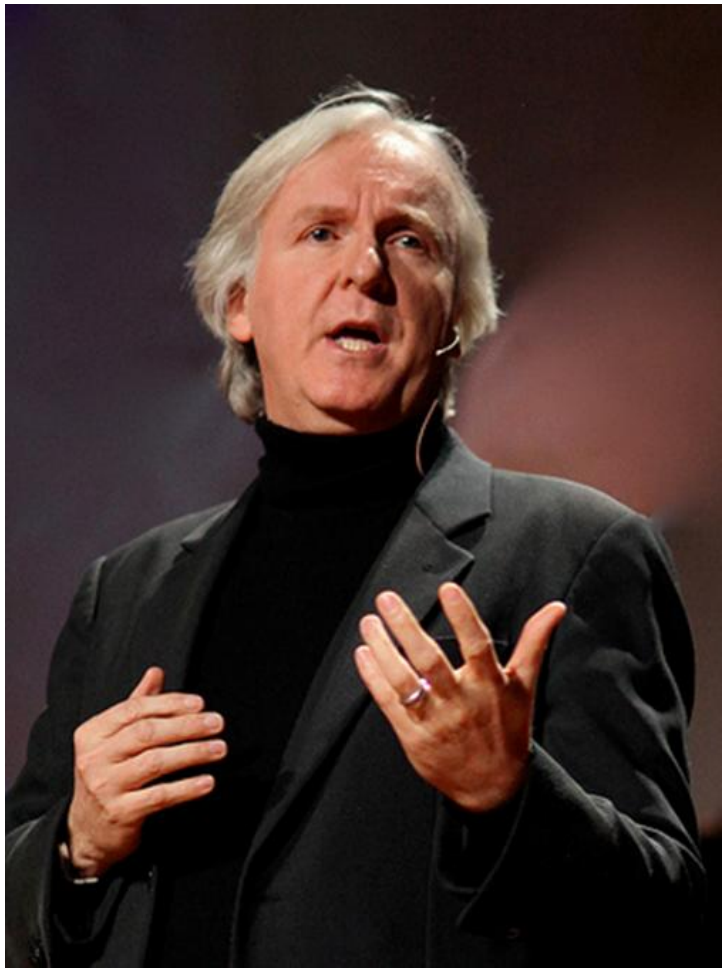
@learningsparql

Print bound copies, PDF, ePUB, Kindle, mobi…

TopQuadrant™

Actors directed by both?





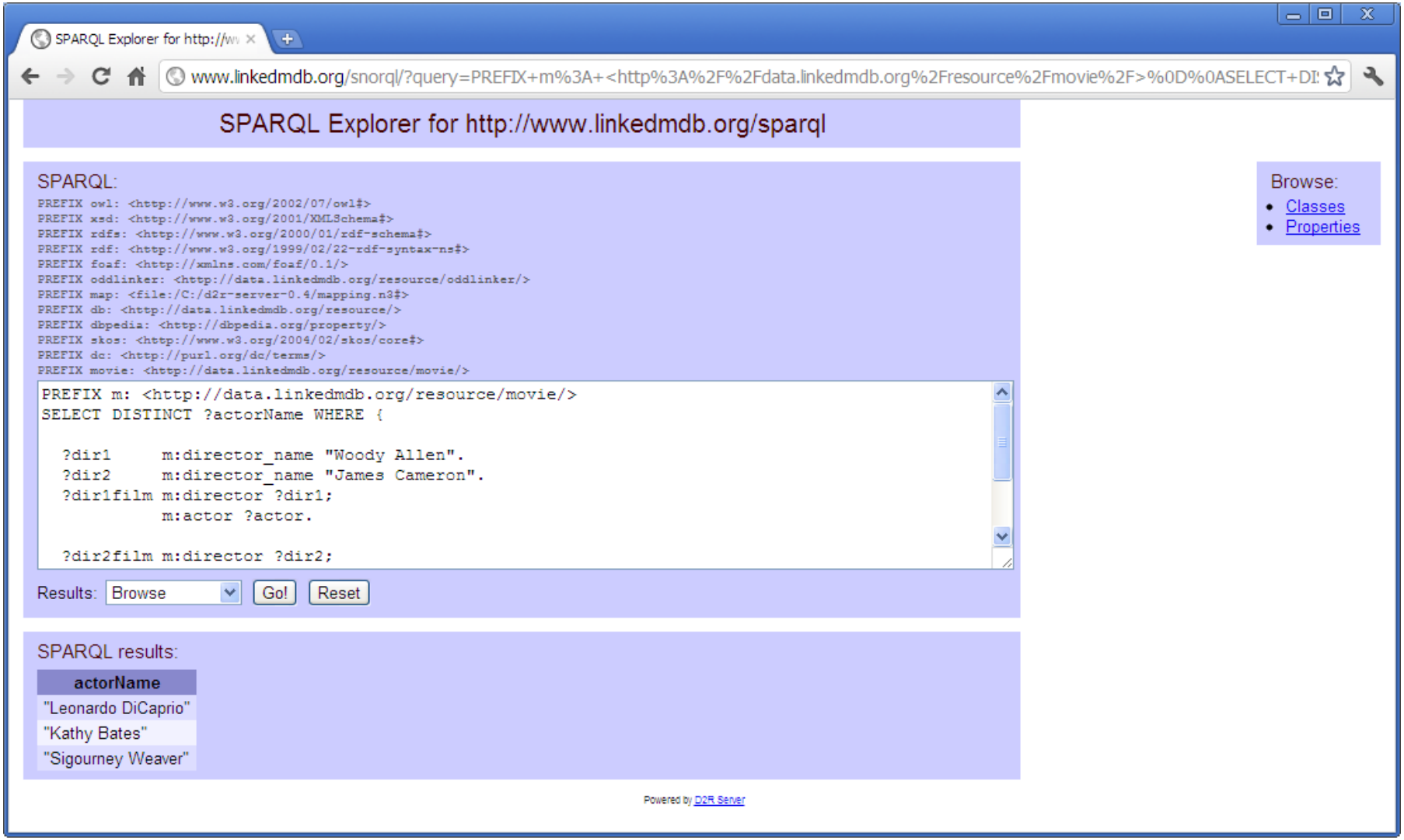
TopQuadrant™

The SPARQL Query


```
PREFIX m: <http://data.linkedmdb.org/resource/movie/>
SELECT DISTINCT ?actorName WHERE {

  ?dir1      m:director_name "Woody Allen".
  ?dir2      m:director_name "James Cameron".
  ?dir1film  m:director ?dir1;
             m:actor ?actor.

  ?dir2film  m:director ?dir2;
             m:actor ?actor.

  ?actor     m:actor_name ?actorName.
}
```

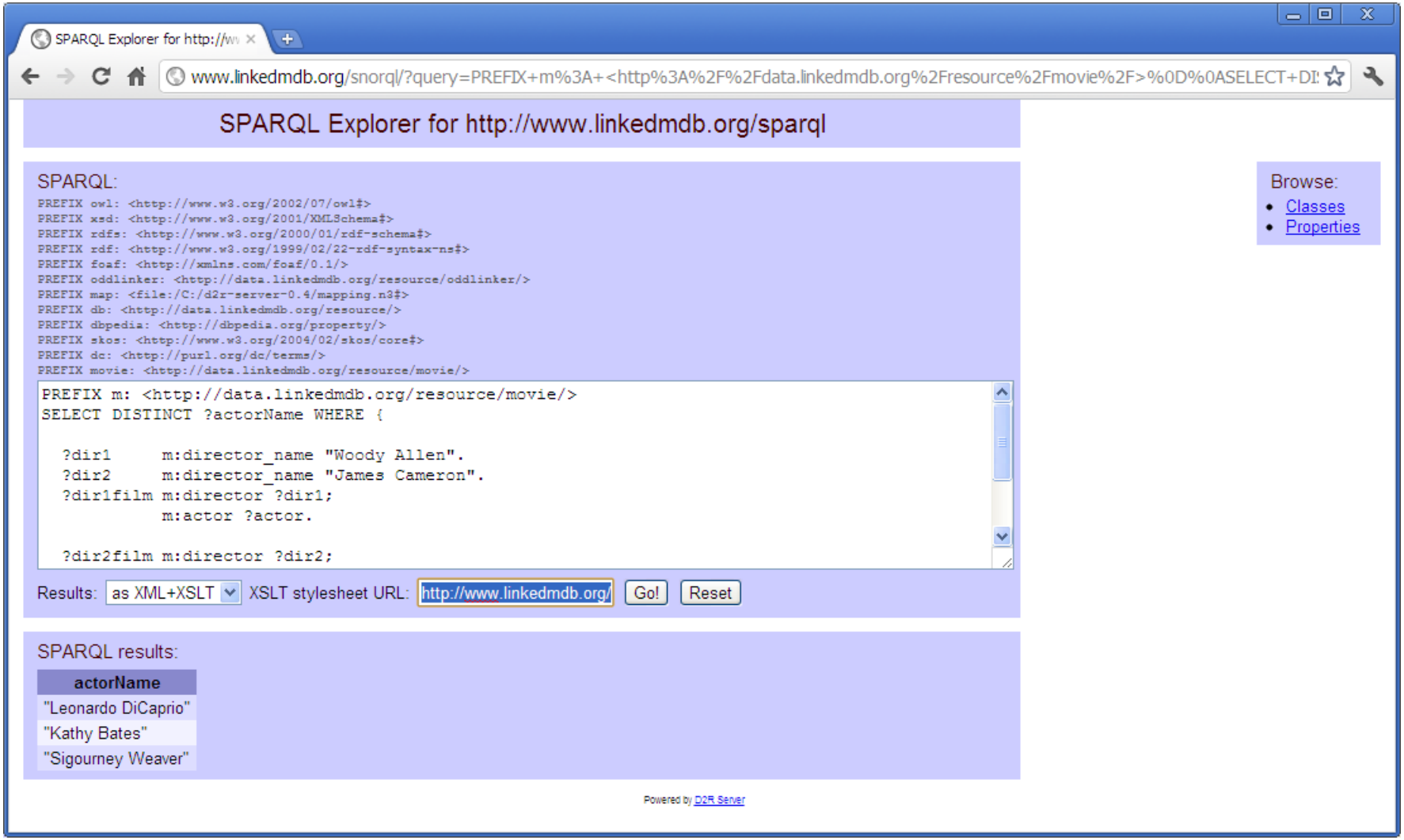
TopQuadrant™

Running query on SNORQL form



TopQuadrant™

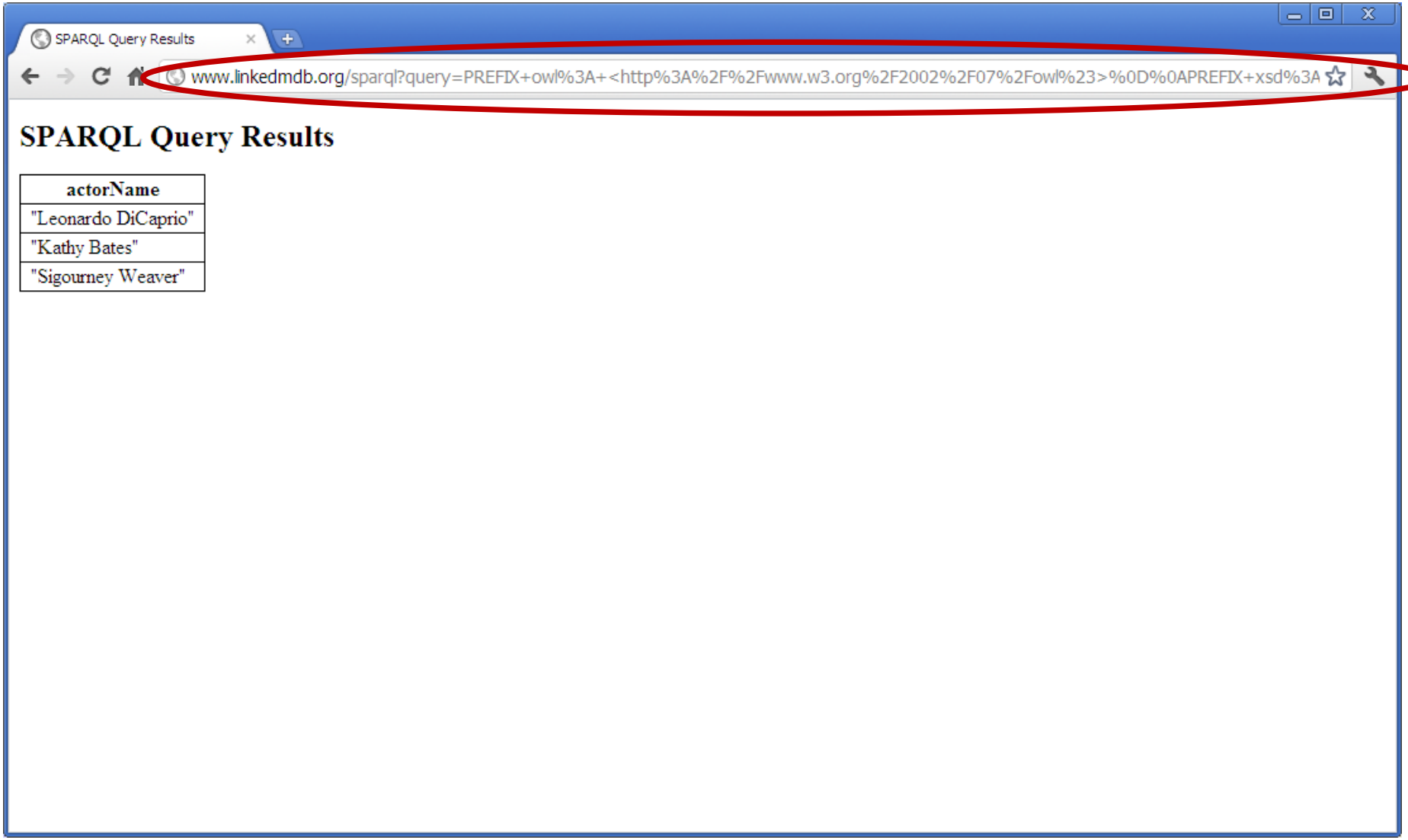
Running query on SNORQL form



URL: http://www.linkedmdb.org/snorql/xml-to-html.xsl

TopQuadrant™

XML+XSLT results



SPARQL Query Results

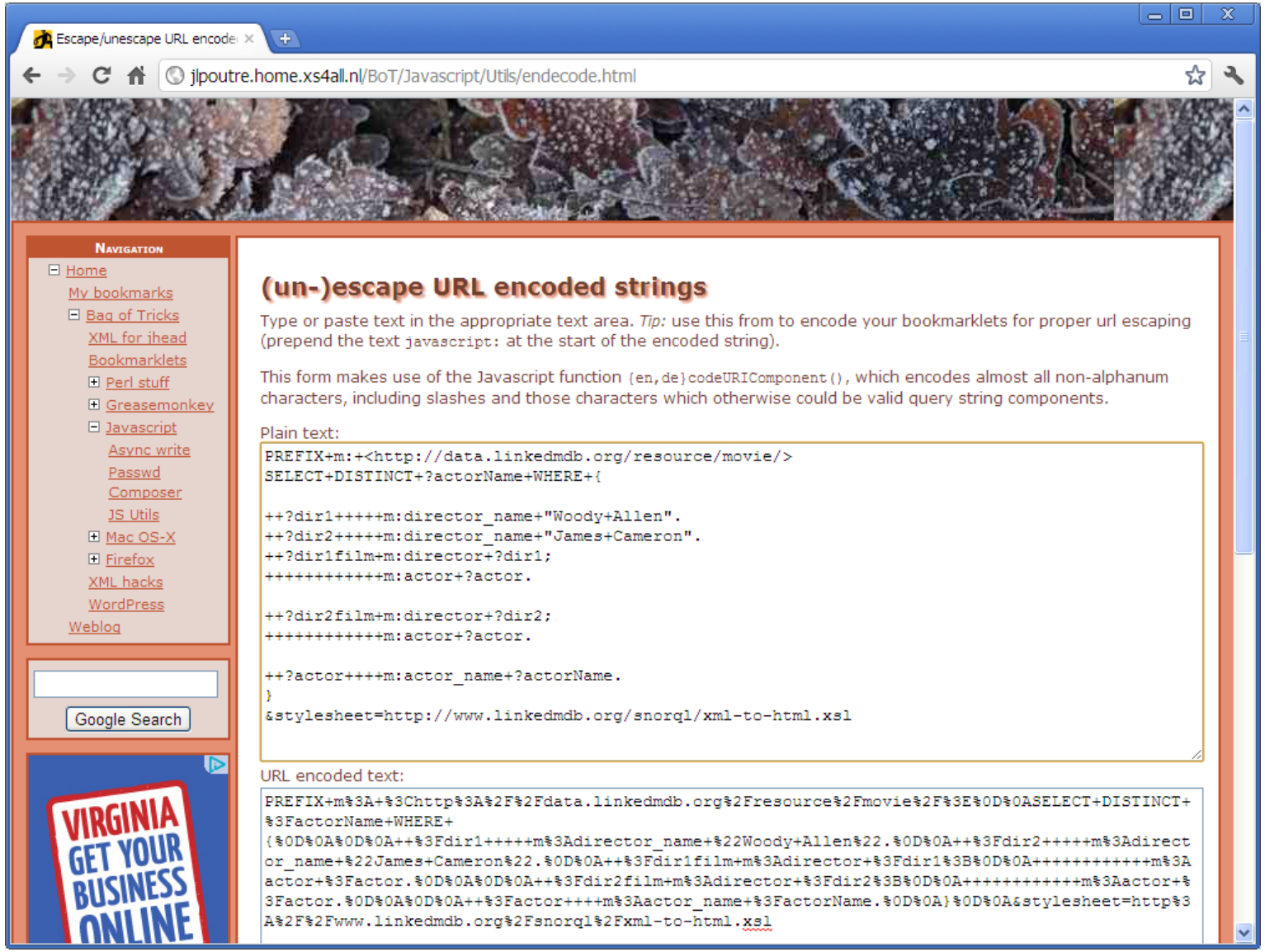
actorName
"Leonardo DiCaprio"
"Kathy Bates"
"Sigourney Weaver"

Full Linked MDB query URI

```
http://www.linkedmdb.org/sparql?query=PREFIX+m
%3A+%3Chttp%3A%2F%2Fdata.linkedmdb.org%2Fresou
rce%2Fmovie%2F%3E%0D%0ASELECT+DISTINCT+%3Facto
rName+WHERE+{%0D%0A%0D%0A++%3Fdir1+++++m%3Adir
ector_name+%22Woody+Allen%22.%0D%0A++%3Fdir2++
+++m%3Adirector_name+%22James+Cameron%22.%0D%0
A++%3Fdir1film+m%3Adirector+%3Fdir1%3B%0D%0A++
++++++++++m%3Aactor+%3Factor.%0D%0A%0D%0A++%3F
dir2film+m%3Adirector+%3Fdir2%3B%0D%0A+++++++
++++m%3Aactor+%3Factor.%0D%0A%0D%0A++%3Factor+
+++m%3Aactor_name+%3FactorName.%0D%0A}%0D%0A&s
tylesheet=http%3A%2F%2Fwww.linkedmdb.org%2Fsno
rql%2Fxml-to-html.xsl
```

TopQuadrant™

Unescaped version of blue text



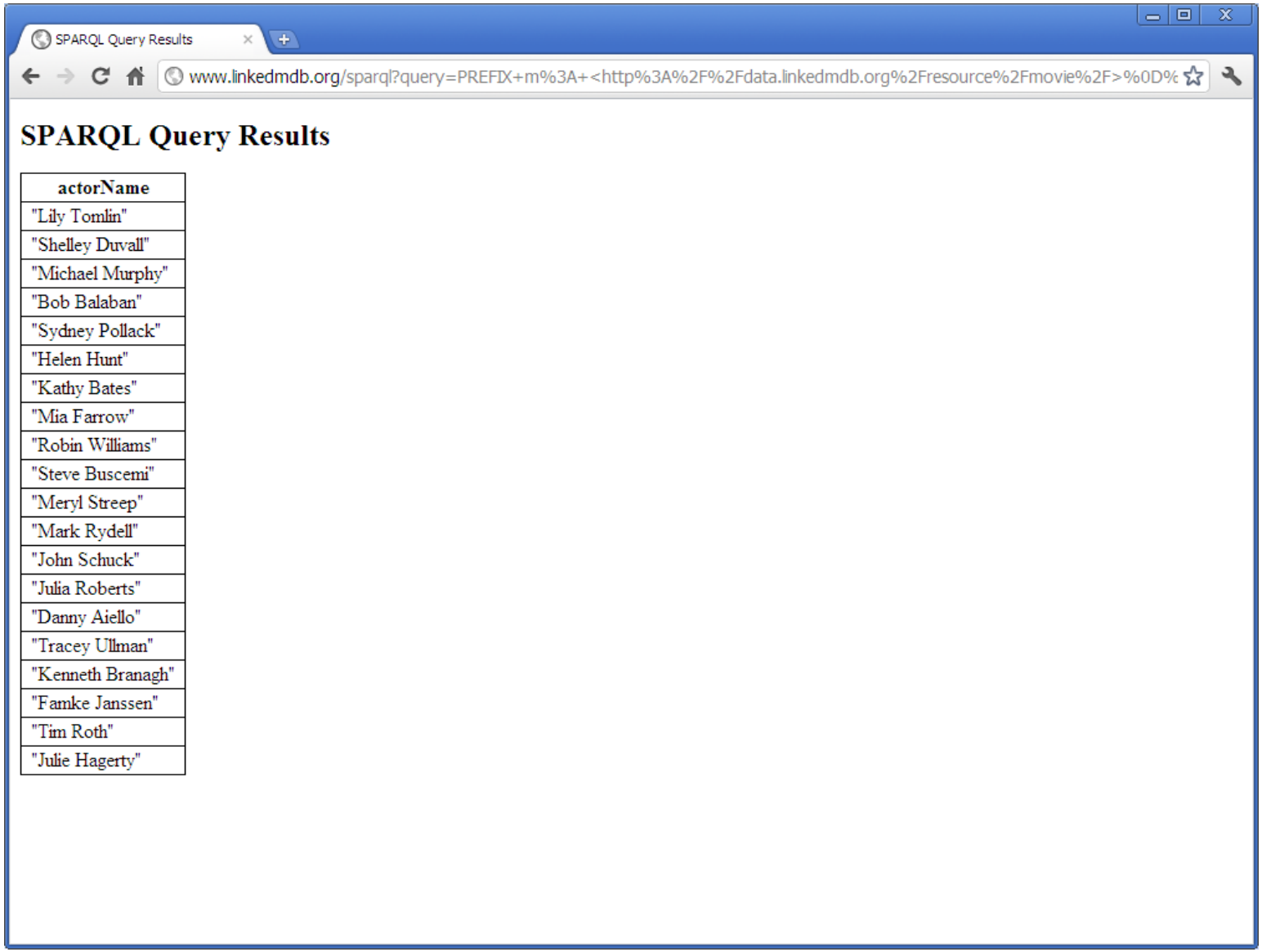
TopQuadrant™

Revised Linked MDB query URI

http://www.linkedmdb.org/sparql?query=PREFIX+m%3A+%3Chttp%3A%2F%2Fdata.linkedmdb.org%2Fresource%2Fmovie%2F%3E%0D%0ASELECT+DISTINCT+%3FactorName+WHERE+{%0D%0A%0D%0A++%3Fdir1+++++m%3Adirector_name+%22Woody+Allen%22.%0D%0A++%3Fdir2+++++m%3Adirector_name+%22Robert+Altman%22.%0D%0A++%3Fdir1film+m%3Adirector+%3Fdir1%3B%0D%0A++++++++++m%3Aactor+%3Factor.%0D%0A%0D%0A++%3Fdir2film+m%3Adirector+%3Fdir2%3B%0D%0A+++++++++++m%3Aactor+%3Factor.%0D%0A%0D%0A++%3Factor+++m%3Aactor_name+%3FactorName.%0D%0A}%0D%0A&stylesheet=http%3A%2F%2Fwww.linkedmdb.org%2Fsnorql%2Fxml-to-html.xsl

TopQuadrant™

Woody Allen/Robert Altman actors



SPARQL Query Results

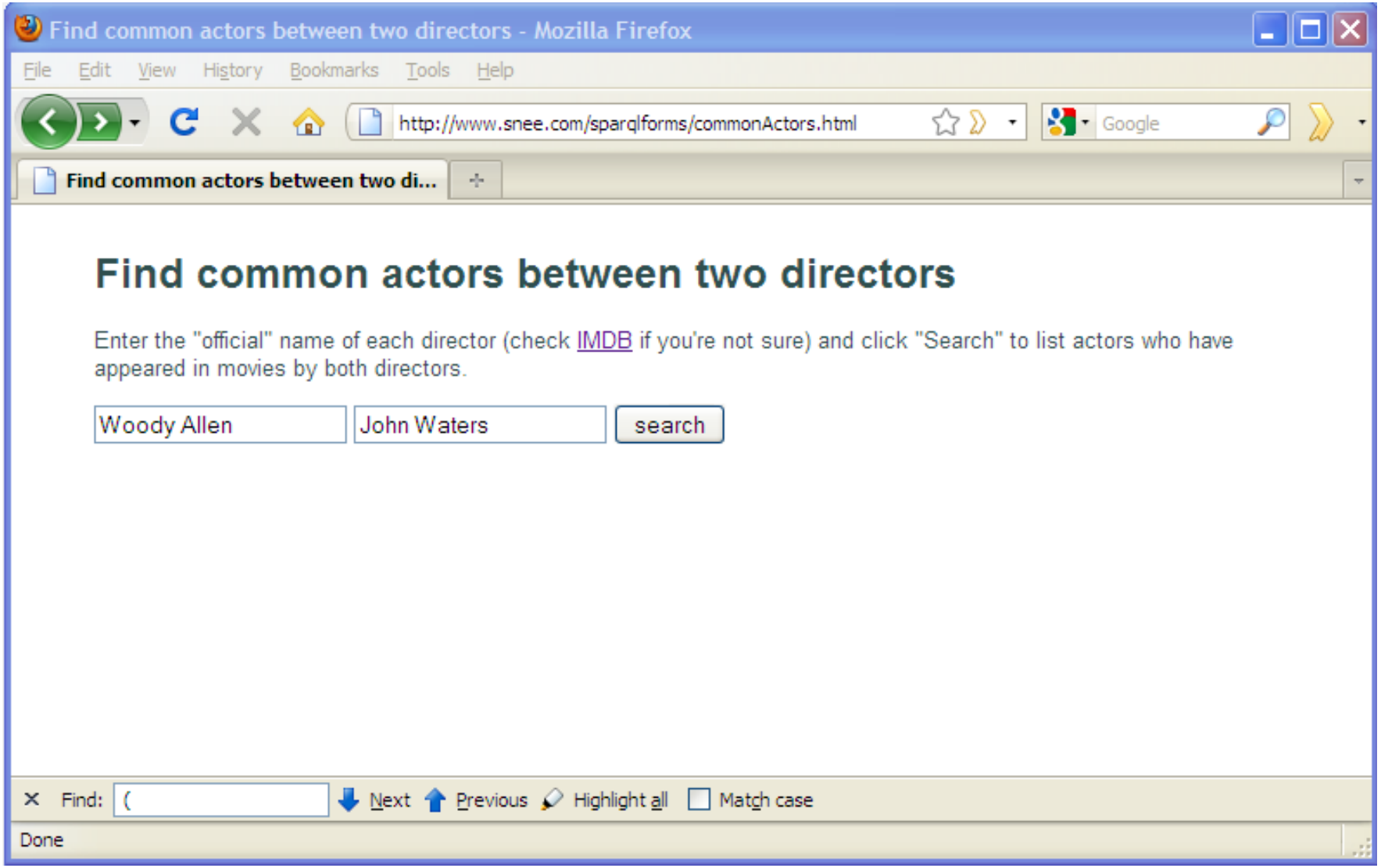
actorName
"Lily Tomlin"
"Shelley Duvall"
"Michael Murphy"
"Bob Balaban"
"Sydney Pollack"
"Helen Hunt"
"Kathy Bates"
"Mia Farrow"
"Robin Williams"
"Steve Buscemi"
"Meryl Streep"
"Mark Rydell"
"John Schuck"
"Julia Roberts"
"Danny Aiello"
"Tracey Ullman"
"Kenneth Branagh"
"Famke Janssen"
"Tim Roth"
"Julie Hagerty"

Using cURL or wget to retrieve results

```
wget -O actors.xml "big url from earlier slide"

curl -o actors.xml "url from earlier slide"
```

TopQuadrant™

Form-based SPARQL app



HTML source of query form

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Find common actors between two directors</title>
    <link href="simple.css" type="text/css" rel="stylesheet" />
  </head>
  <body>
    <h1>Find common actors between two directors</h1>
    <form action="twodirectors" method="get">

      <p>Enter the "official" name of each director (check <a
href="http://www.imdb.com">IMDB</a> if you're not sure) and click
"Search" to list actors who have appeared in movies by both
directors.</p>
      <p>
        <input type="text" name="dir1"/>
        <input type="text" name="dir2"/>
        <input type="submit" value="search"/>
      </p>

    </form>
  </body>
</html>
```

URL this will try to retrieve

```
http://www.snee.com/sparqlforms/twodirectors?dir1=Woody+Allen&dir2=John+Waters
```

Can we somehow get this instead?

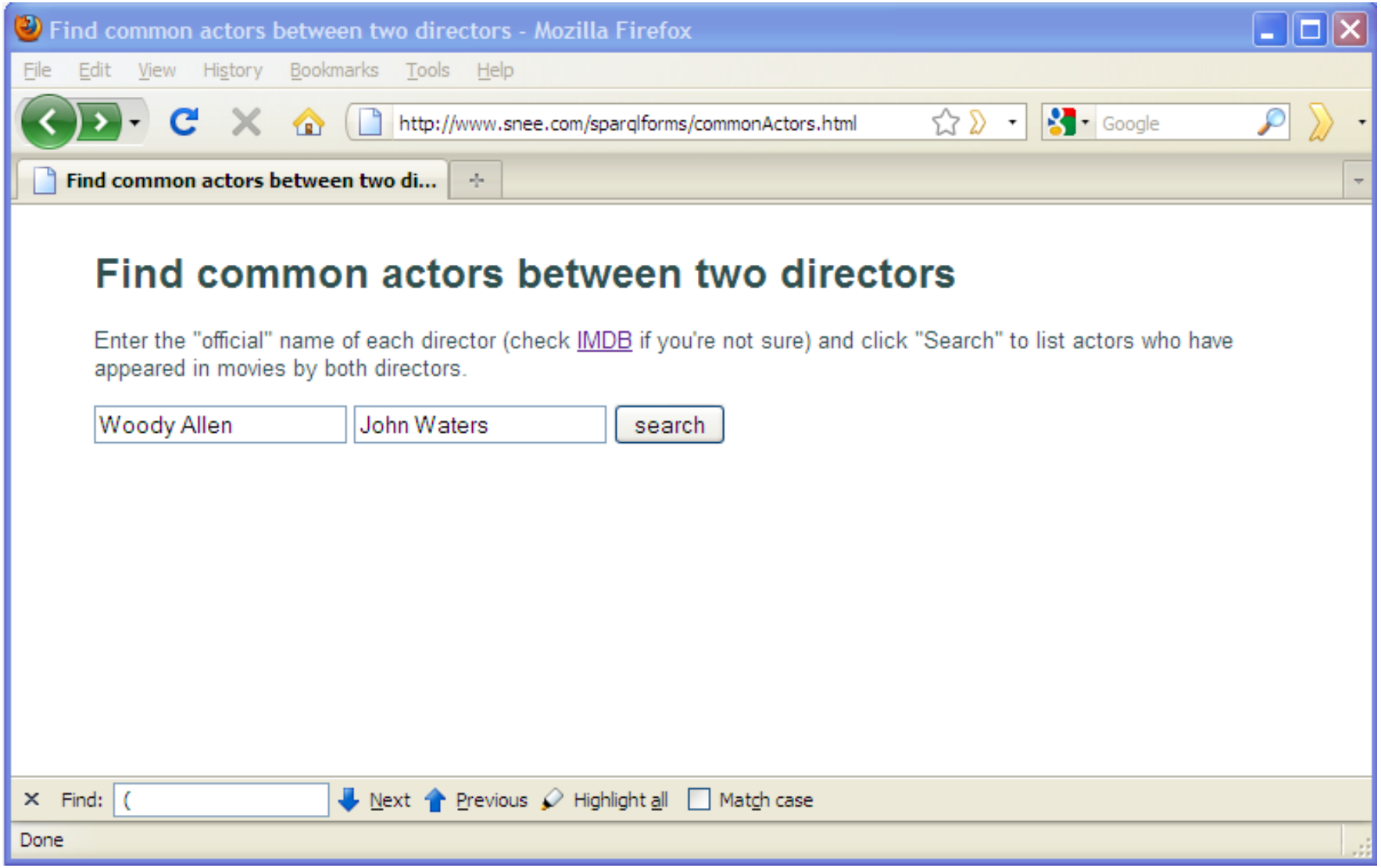
```
http://www.linkedmdb.org/sparql?query=PREFIX+m%3A+%3Chttp%3A%2F%2Fdata.linkedmdb.org%2Fresource%2Fmovie%2F%3E%0D%0ASELECT+DISTINCT+%3FactorName+WHERE+{%0D%0A%0D%0A++%3Fdir1+++++m%3Adirector_name+%22Woody+Allen%22.%0D%0A++%3Fdir2+++++m%3Adirector_name+%22John+Waters%22.%0D%0A++%3Fdir1film+m%3Adirector+%3Fdir1%3B%0D%0A++++++++++++m%3Aactor+%3Factor.%0D%0A%0D%0A++%3Fdir2film+m%3Adirector+%3Fdir2%3B%0D%0A+++++++++++m%3Aactor+%3Factor.%0D%0A%0D%0A++%3Factor++++m%3Aactor_name+%3FactorName.%0D%0A}%0D%0A&stylesheet=http%3A%2F%2Fwww.linkedmdb.org%2Fsnorql%2Fxml-to-html.xsl
```

With an .htaccess redirect

```
RewriteEngine On

ErrorDocument 404 /404page.html

redirect 301 old_filename.html new_filename.html

RewriteCond %{QUERY_STRING} dir1=(.*)&dir2=(.*)

# The rest is actually one line

RewriteRule ^.*twodirectors.*$
http://www.linkedmdb.org/sparql?query=PREFIX+m:+<http:/
/data.linkedmdb.org/resource/movie/>+SELECT+DISTINCT+?a
ctorName+WHERE+{+?dir1+m:director_name+"%1".+?dir2+m:di
rector_name+"%2".+?dir1film+m:director+?dir1;+m:actor+?
actor.+?dir2film+m:director+?dir2;+m:actor+?actor.+?act
or+m:actor_name+?actorName.+}+&stylesheet=http://www.li
nkedmdb.org/snorql/xml-to-html.xsl
```

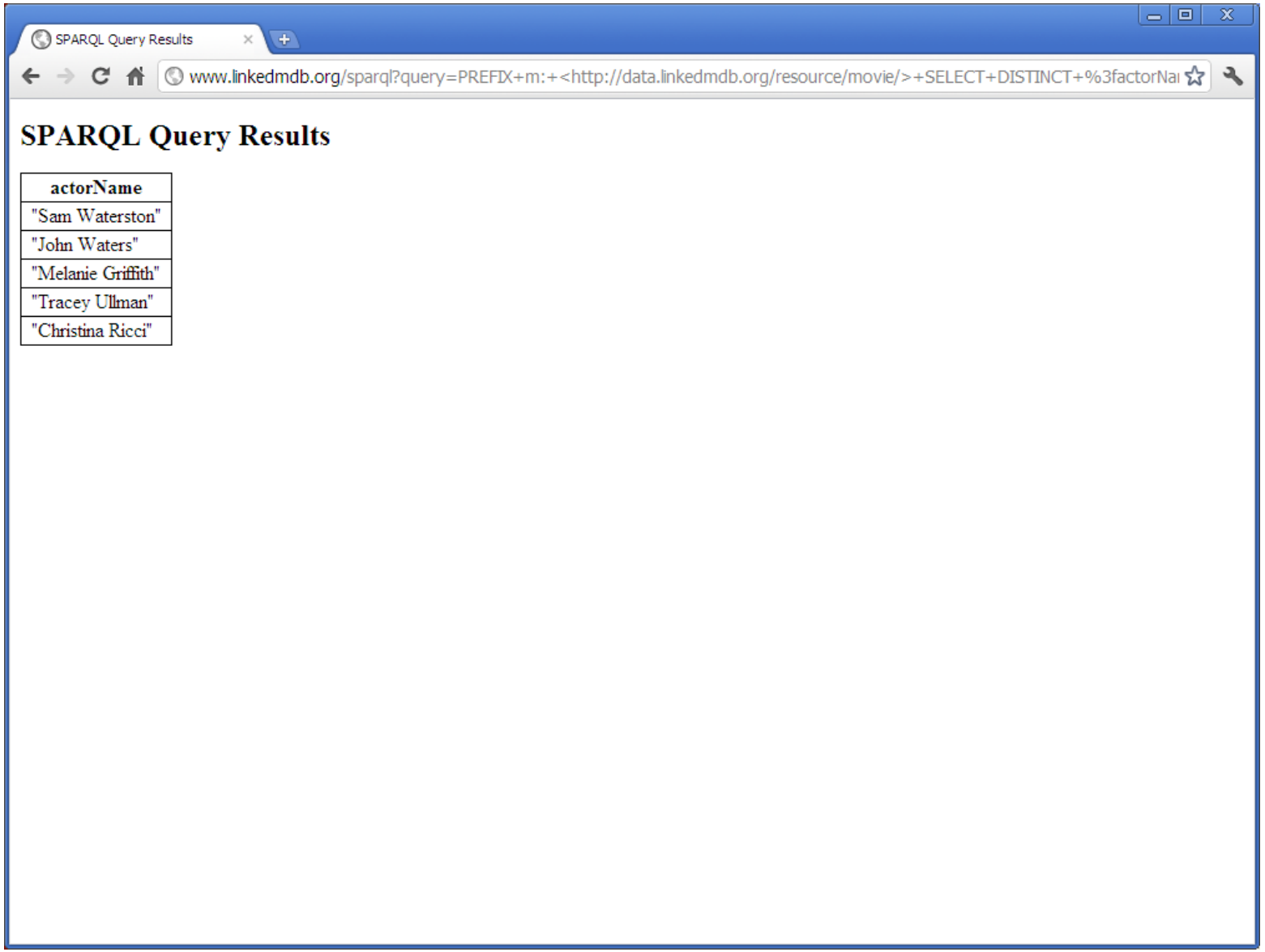
TopQuadrant™

Entering this query…



TopQuadrant™

Gets us this result.



HTML source of query form

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Find common actors between two directors</title>
    <link href="simple.css" type="text/css" rel="stylesheet" />
  </head>
  <body>
    <h1>Find common actors between two directors</h1>
    <form action="commonActors.cgi" method="get">

      <p>Enter the "official" name of each director (check <a
href="http://www.imdb.com">IMDB</a> if you're not sure) and click
"Search" to list actors who have appeared in movies by both
directors.</p>
      <p>
        <input type="text" name="dir1"/>
        <input type="text" name="dir2"/>
        <input type="submit" value="search"/>
      </p>

    </form>
  </body>
</html>
```

commonactors.cgi main() part 1

```
#!/usr/local/bin/python

import sys
# following needed for hosted version
sys.path.append('/usr/home/bobd/lib/python/')
from SPARQLWrapper import SPARQLWrapper, JSON
import string
import urllib
import cgi

def main():
  form = cgi.FieldStorage()
  dir1name = form.getvalue('dir1')
  dir2name = form.getvalue('dir2')

  sparql = SPARQLWrapper("http://data.linkedmdb.org/sparql")
```

commonactors.cgi main() part 2

```
queryString = """

PREFIX m: <http://data.linkedmdb.org/resource/movie/>
SELECT DISTINCT ?actorName WHERE {

  ?dir1     m:director_name "DIR1-NAME".
  ?dir2     m:director_name "DIR2-NAME".
  ?dir1film m:director ?dir1;
            m:actor ?actor.

  ?dir2film m:director ?dir2;
            m:actor ?actor.

  ?actor    m:actor_name ?actorName.
}
  """
```

commonactors.cgi main() part 3

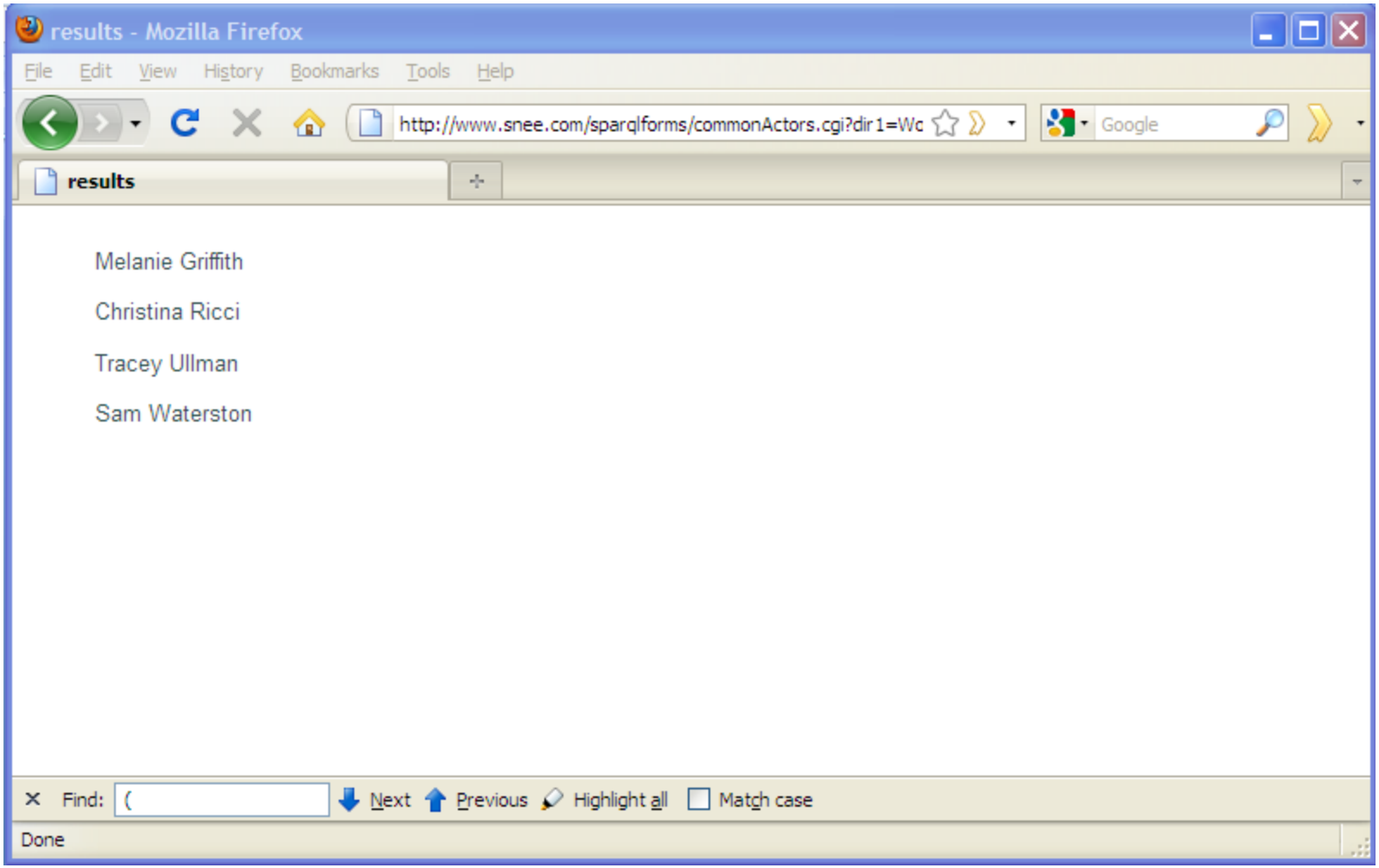
```
queryString = queryString.replace("DIR1-NAME",dir1name)
queryString = queryString.replace("DIR2-NAME",dir2name)

sparql.setQuery(queryString)
sparql.setReturnFormat(JSON)

try:
  ret = sparql.query()
  results = ret.convert()
  requestGood = True
except Exception, e:
  results = str(e)
  requestGood = False
```

commonactors.cgi main() part 4

```
  print """Content-type: text/html

    <html>
      <head>
        <title>results</title>
          <link href="simple.css" type="text/css" rel="stylesheet" />
      </head>
      <body>
"""

  if requestGood == False:
    print "<h1>Problem communicating with the server</h1>"
    print "<p>" + results + "</p>"
  elif (len(results["results"]["bindings"]) == 0):
      print "<p>No results found.</p>"

  else:
    for result in results["results"]["bindings"]:
      print "<p>" + result["actorName"]["value"] + "</p>"
  print "</body></html>"
```

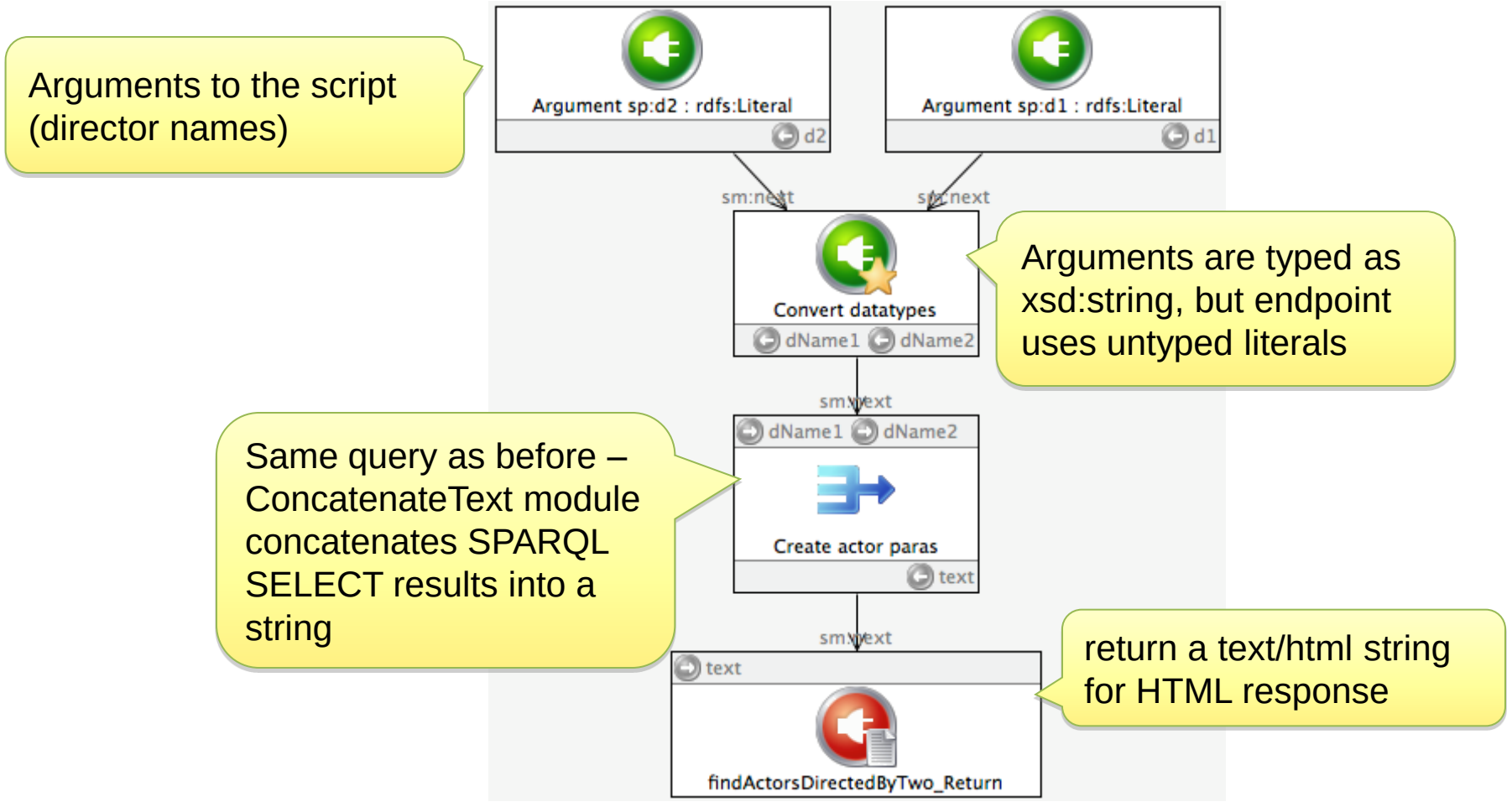
TopQuadrant™

Form-based SPARQL app: results



TopBraid SPARQLMotion service

- SPARQLMotion Script: get arguments, process query, return text/html



Calling SPARQLMotion Web Service

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
    <title>Find common actors between two directors</title>
    <link href="simple.css" type="text/css" rel="stylesheet" /> </head>
<body>
    <h1>Find common actors between two directors</h1>
    <form action="http://localhost:8083/tbl/sparqlmotion">
        <p>Enter the "official" name of each director (check
        <a href="http://www.imdb.com">IMDB</a> if you're not sure)
            and click "Search" to list actors
        who have appeared in movies by both directors.</p>
        <p> <input type="text" name="d1"/>
                <input type="text" name="d2"/>
                <input type="submit" value="search"/></p>
        <input type="hidden" name="id" value="FindActorsDirectedByTwo"/>
    </form>
</body></html>
```

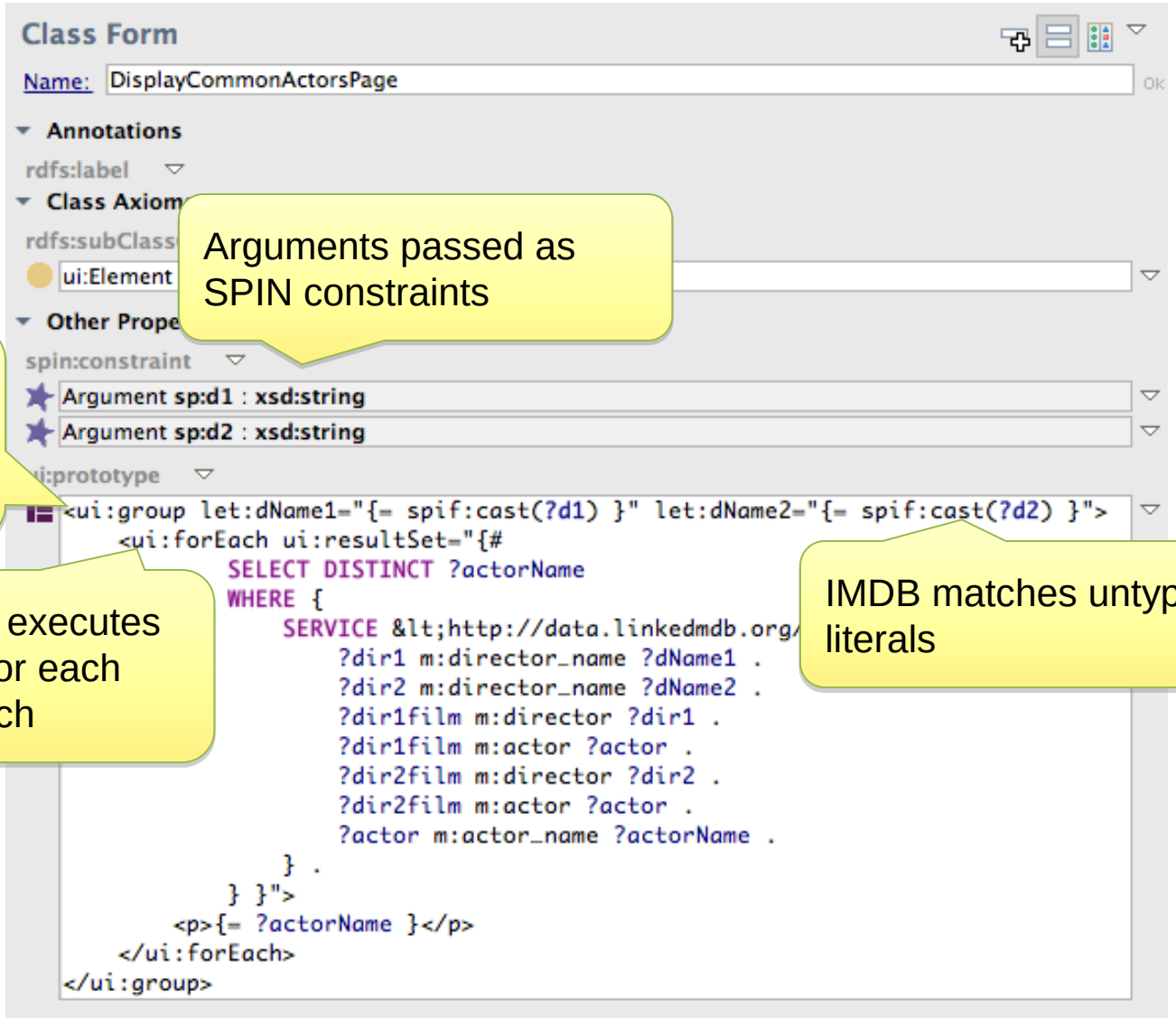
Example:

http://localhost:8083/tbl/sparqlmotion?id=FindActorsDirectedByTwo&d1=John Waters&d2=Woody Allen

Live server URL — **Function name** — **parameters**

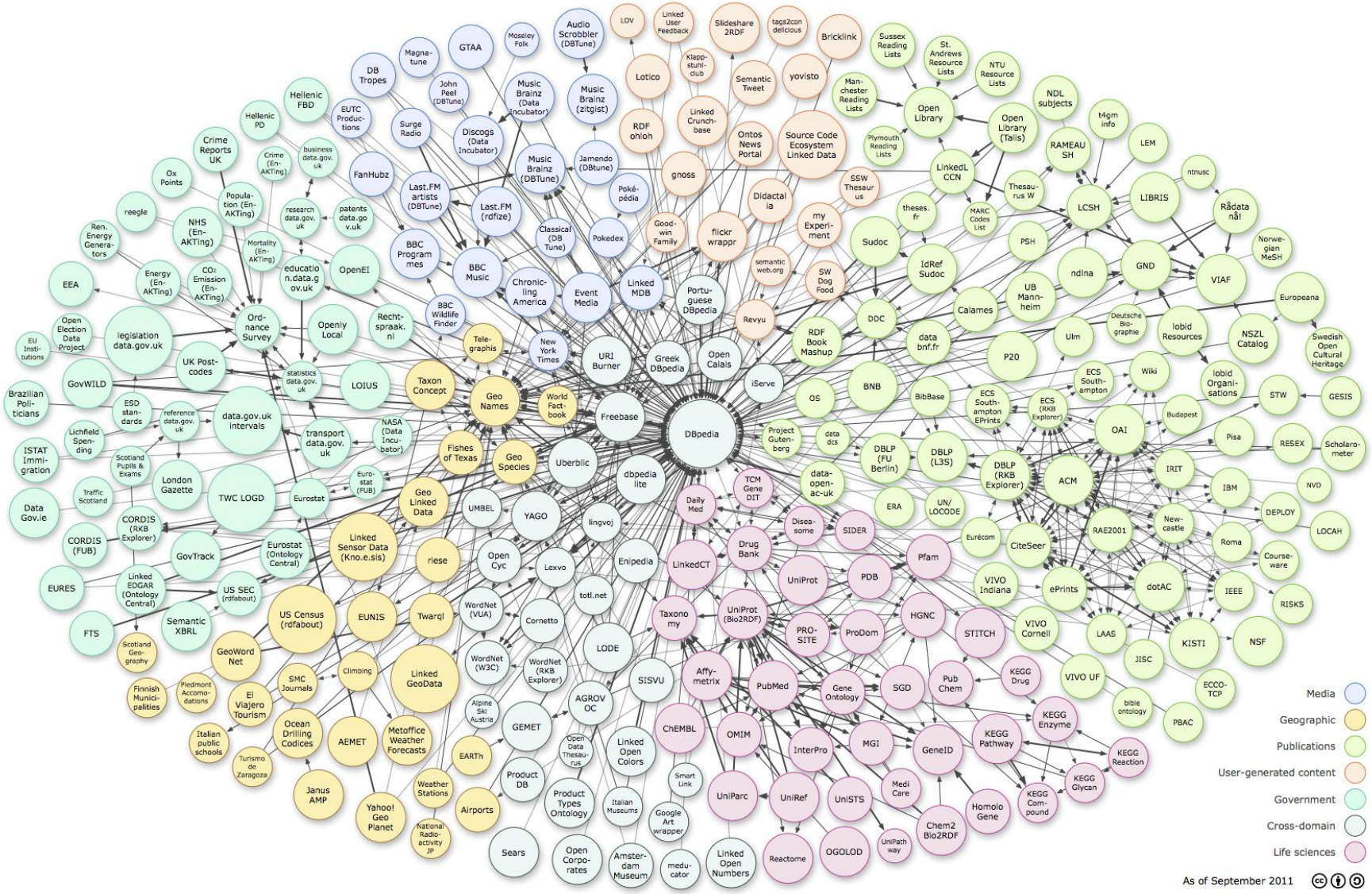
Using SPARQL Web Pages

- SWP template



TopQuadrant™

Obligatory Linked Data Cloud slide



Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences

As of September 2011

TopQuadrant™

Is SPARQL Difficult?

File Edit View History Bookmarks Tools Help

http://www.cnn.com/2008/TECH/12/17/db.semanticweb/

Google

Making sense of the 'semantic Web' ...

digitalbiz

IN ASSOCIATION WITH KONICA MINOLTA

updated 10:01 a.m. EST, Thu December 18, 2008

Making sense of the 'semantic Web'

STORY HIGHLIGHTS

- The "semantic web" could make the Internet more interactive and rewarding
- Things on the Web will be described in languages that computers can understand
- An experimental kiosk lets users see the potential of this Web 3.0 application
- Widespread adoption of the semantic Web is still a long way off

Next Article in Technology »

By Steve Mollman
For CNN

TEXT SIZE

(CNN) -- The "semantic Web" does not sound like it's fun and easy to use, but it could make surfing Web 3.0 a more rewarding and interactive experience. Some believe it could even lead to a new form of artificial intelligence.

Spotlets pop up on a 'semantic Web' kiosk run by a pair of German researchers.

The idea behind the semantic Web, very broadly, is that things on the Internet will be described with descriptor languages so that computers can "understand" what they are.

An object might be a marked as a car part or a person, for instance. If objects were thus identified, an enormous network of linked data would emerge and machines, with their vast processing speeds, could suggest surprising and useful links that the human mind could never come up with, posing the possibility of a new sort of artificial intelligence.

The semantic Web is considered a key part of the upcoming "Web 3.0." It's starting to occur here and there, but widespread adoption is still a long way off.

A pair of German researchers have created an experimental kiosk that lets you easily use semantic Web capabilities -- even if you have no idea what they are. All that is needed is an iPhone and a finger with which to drag icons around on the kiosk's touch screen.

Find: Next Previous Highlight all Match case

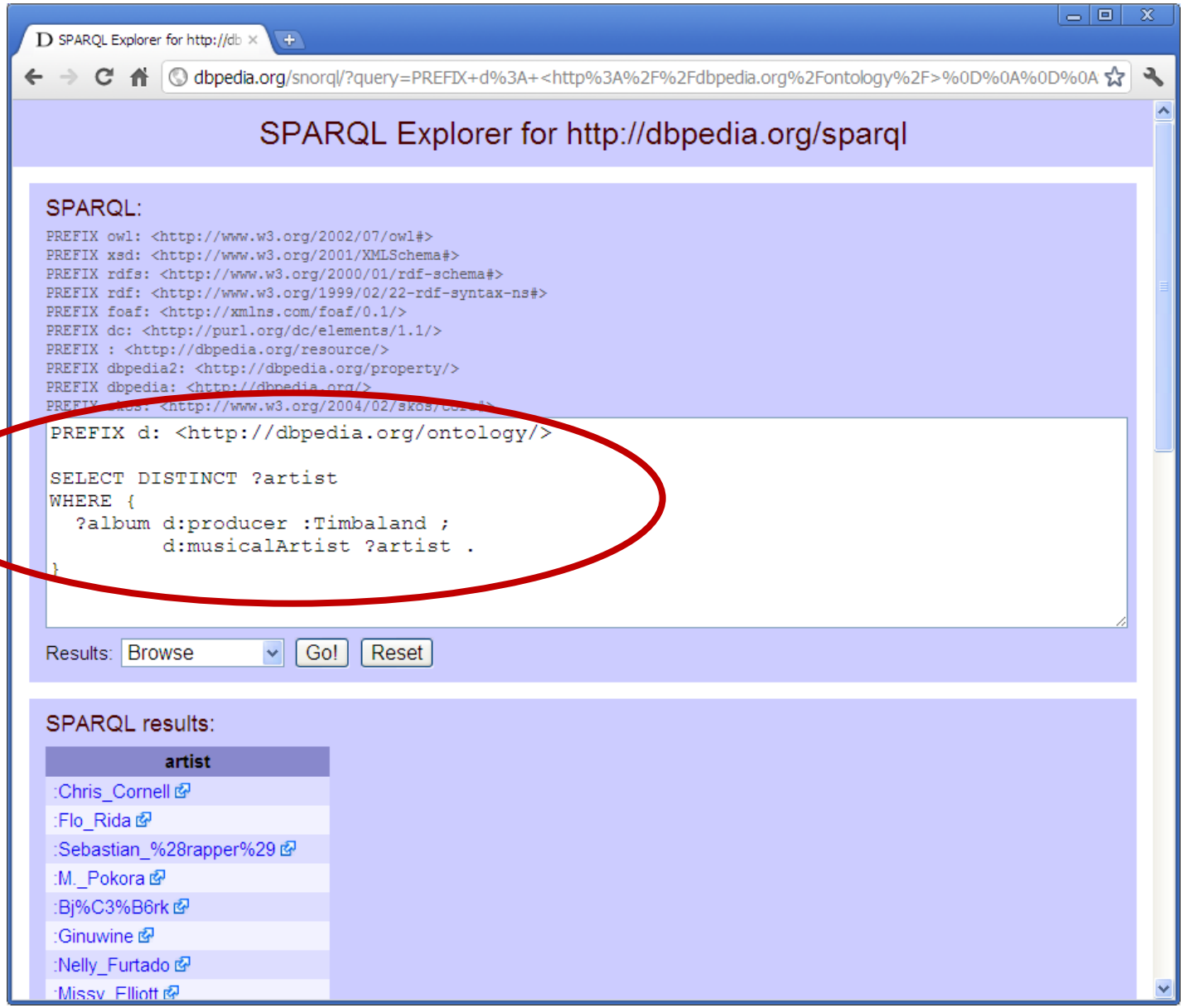
Done

Is SPARQL Difficult?

"Consider, for instance, SPARQL, a query language. To find, say, music artists associated with the producer Timbaland, you'd have to type a long piece of convoluted code that most of us wouldn't bother to do."

TopQuadrant™

Artists Timbaland produced



On the other hand…

Some JavaScript from a View Source of that same CNN page:

```
if(cnnWinExtraRegExp.test(cnnWinExtra)){var cnnOmniExtra =
cnnWinExtraRegExp.split(cnnWinExtra);cnnWinLoc = cnnWinLoc +
cnnOmniExtra[0];} else {cnnWinLoc = cnnWinLoc +
cnnWinExtra;}} if (typeof(cnnPageName) != "undefined")
{s.pageName = cnnPageName;s.eVar1 = cnnPageName;} else
{s.pageName = cnnWinLoc;s.eVar1 = cnnWinLoc;} if
(typeof(cnnSectionName) != "undefined")
{s.channel=cnnSectionName;s.eVar2=cnnSectionName;} else
{s.channel="Nonlabeled";s.eVar2="Nonlabeled";} if
(typeof(cnnSubSectionName) != "undefined")
{s.server=cnnSubSectionName;s.eVar3=cnnSubSectionName;} else
{s.server="";s.eVar3="";} if (typeof(cnnSectionFront) !=
"undefined") {s.prop1=cnnSectionFront;} if
(typeof(cnnContentType) != "undefined")
{s.prop4=cnnContentType;s.prop6=s.pageName;}
```

Summary

- Simple web forms can send parameterized SPARQL queries to SPARQL endpoints:
  - Using .htaccess files
  - With CGI scripts
  - Using REST web services as shown by TopBraid SPARQLMotion script
  - Embedding SPARQL queries into HTML markup as shown by SPARQL Web Pages with TopBraid

Thank you!

TopBraid Suite™

TopBraid
Enterprise Vocabulary Net™

"A wonderful harmony is created when we join together the seemingly unconnected."

- Heraclitus

Woody Allen picture: http://www.flickr.com/photos/thomasthomas/504401865/ by Thomas Thomas some rights reserved; James Cameron picture http://www.flickr.com/photos/jurvetson/4357736383 by Steve Jurvetson some rights reserved

bducharme@topquadrant.com
info@topquadrant.com