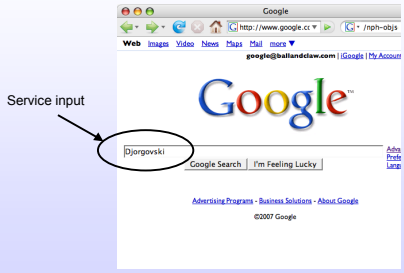
Service Architecture and XML

Roy Williams
California Institute of Technology

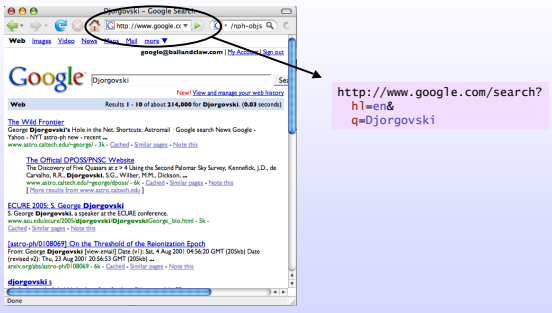
Agenda

- Services
  - Why and how
  - Mashup
  - Security and asynchrony
- XML
  - Why
  - VOTable
  - XSLT

Services drive the web



Query form



Making it happen

HTML form

```
<form action="/search">
  <input name=hl type=hidden value=en>
  <input name=q>
  <input type=submit value="Google Search">
</form>
```

Click submit
Request to server
Response comes back
Rendered by browser

Service Clients

Unix shell

```
curl -o out.html "http://www.google.com/q=Djorgovski"
```

Python calls Unix

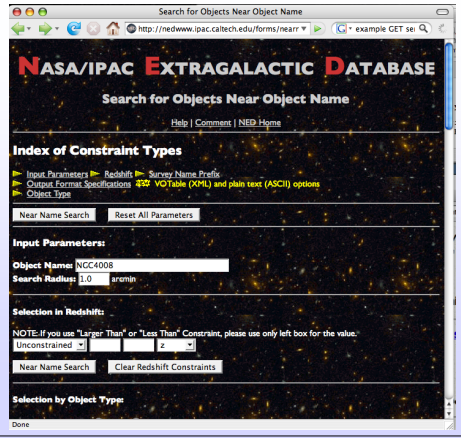
```
import os
searchstring = "Djorgovski"
url = "http://www.google.com/q=%s" % searchstring
cmd = "curl -o out.html %s" % url
os.system(cmd)
```

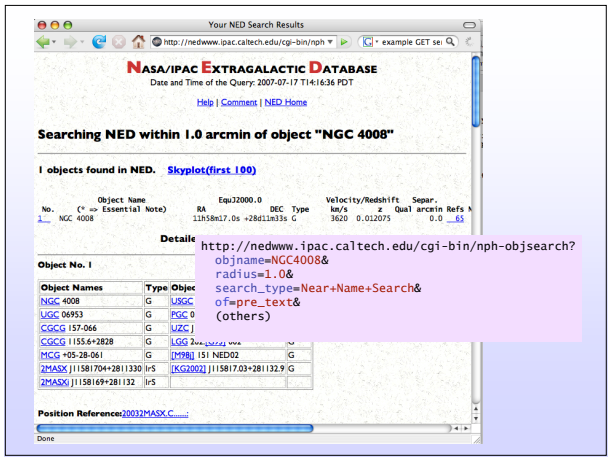
Python direct

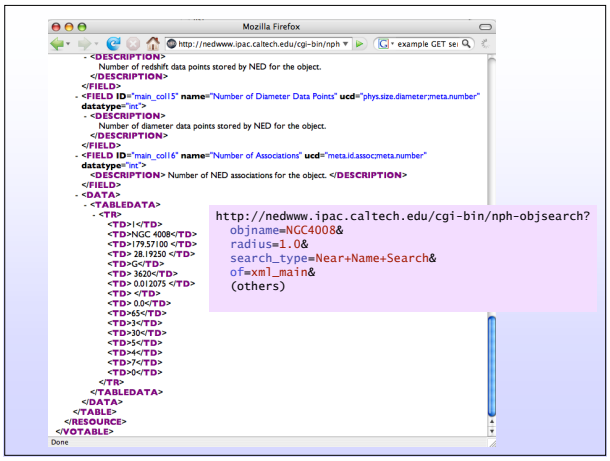
```
import urllib
try:
    stream = urllib.urlopen(url)
except IOError,e:
    print "Cannot open ",url
    print "Error is ", e
    continue
for line in readlines():
    print line
```

What is a web service?

- "A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format." - W3C
  - ⇒ WSDL and SOAP conveyed using HTTP with an XML serialization + other Web-related standards
- Not a new idea:
  - RPC/RMI
  - CORBA
  - DCOM
  - XML-RPC







Messier Catalog with NED

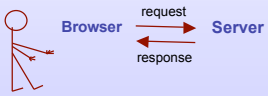
```
import urllib

urlbase = "http://nedwww.ipac.caltech.edu/cgi-bin/nph-objsearch?"
urlbase += "&radius=1.0&search_type=Near+Name+Search&of=xml_main"

for messierNumber in range(1,110):
    url = urlbase + "&objname=M%d" % messierNumber
    try:
        stream = urllib.urlopen(url)
    except IOError,e:
        print "Cannot open ",url

    outfile = open("messier%d.html" % messierNumber, "w")
    outfile.write(stream.read())
```

Messier Catalog with NED
(most popular galaxy)

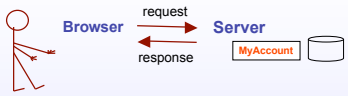
```
import urllib
import xml.dom.minidom

urlbase = "http://nedwww.ipac.caltech.edu/cgi-bin/nph-objsearch?"
urlbase += "&radius=1.0&search_type=Near+Name+Search&of=pre_text"

for messierNumber in range(1,110):
    url = urlbase + "&objname=M%d" % messierNumber
    try:
        stream = urllib.urlopen(url)
    except IOError,e:
        print "Cannot open ",url

    doc = xml.dom.minidom.parse(stream)
    for node in doc.getElementsByTagName("TR"):
        j = 0
        for node2 in node.getElementsByTagName("TD"):
        j += 1
        if j == 10:   # number of references is 10th column of table
            refs = ""
            for n in node2.childNodes:
                if n.nodeType == node.TEXT_NODE: refs += n.data
            print "%d references for Messier %d" % (refs, messierNumber)
```

Simple service

- Request
  - Keyword = value pairs
- Response
  - Document
    - Parsed by human
    - Parsed by machine

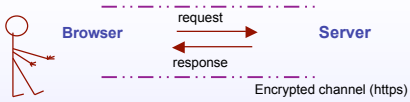


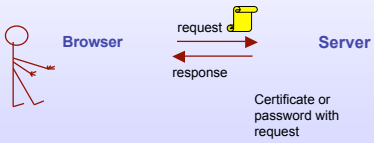
Services with State



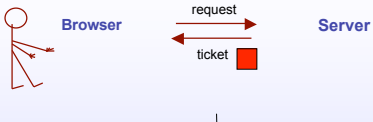
- Log in first, then
  - Shopping, banking, facebook, blog, etc
  - Science workbench

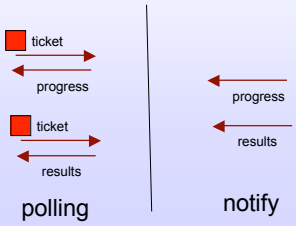
Secure Services





Asynchronous Services





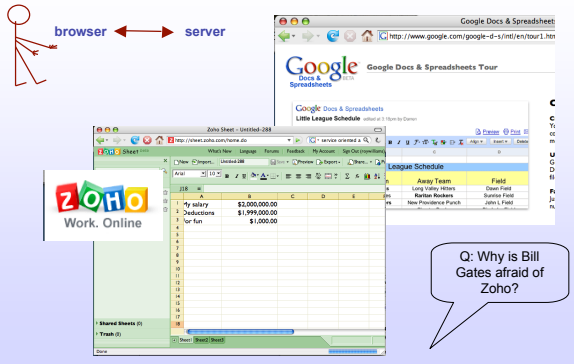
Service Oriented Architecture

- Service Encapsulation
  - Hiding details, coding, hide logic from the outside world
- Service autonomy
  - Services have control over the logic they encapsulate
- Loose coupling
  - Minimise dependencies, allow collaboration
- Service contract
  - Agreement on function, as defined collectively by one or more service description documents
- Service composability
  - Collections of services can be coordinated and assembled to form composite services
- Service discoverability
  - Services can be found and assessed via available discovery mechanisms

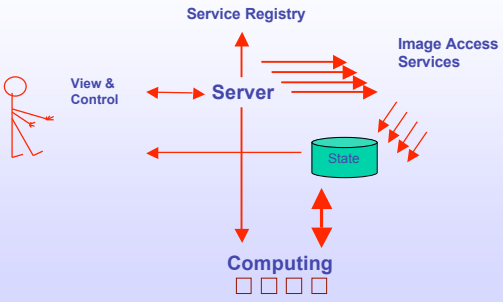
VO Services

- Cone Search
  - Input: position in the sky and radius
  - Output: Stars found in that region
- Image Access
  - Input: position in the sky and radius
  - Output: Images that cover that region
- … Spectra, registry services, compute services …

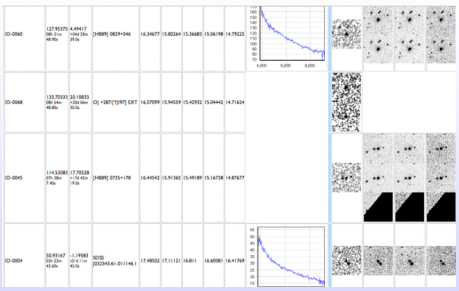
Service Based Tools



Services call Services (mashup)



VO Service mashup (VIM)



VO Data Services

- Cone Search
  - First standard NVO service:
    - radius+position ⇒ list of objects
    - encoded as VOTable
- Simple Image Access Protocol
  - "cone search for images"
  - images are referenced by URL
- Simple Spectrum Access Protocol
  - spectra have subtleties → protocol more complicated

VO Data Services

- Astronomical Data Query Language
  - For database queries
  - Core SQL functions plus astronomy-specific extensions
    - Sky region, Xmatch
- SkyNode
  - Exposes relational databases
  - Accepts ADQL query
  - "Full" SkyNodes support positional cross-match function
  - OpenSkyQuery portal
    - show database structure
    - query tools

XML

- Question:
  - How can a computer use a service?
- Answer:
  - Because the response is XML

I ♥ XML

In HTML :
For example

```
<b>M31<b>
<i>2900<i>
<i>3.4<i>
```

could mean anything

In XML we represent it as.

```
<Source>
        <Galaxy>
            <Name>M31</Name>
            <Distance> 2900 </ Distance >
            <Brightness>3.4</ Brightness >
        </ Galaxy >
    </ Source >
```

Advantages of XML

- Readability
  - XML document is plain text and human readable, To edit/view XML documents any simple text editor will suffice .
- Hierarchical
  - XML document has a tree structure which is powerful enough to express complex data and simple enough to understand
- Language Independent
  - A Java program can generate a XML which can be parsed by a program written in C++ or Perl.
- Structured
  - Schema allows independent check of XML documents.
- OS Independent
  - XML files are Operating System independent.

Uses of XML

- Messaging
  - Where applications or organizations exchanges data between them
- Database
  - The data extracted from the database can be preserved with original information and can be used for more than one application in different ways. One application might just display the data and the other application might perform some complex calculation on this data
- Service Oriented Architecture (SOA)
  - The neutral and generalized format are ideal for data exchange because to simplify reuse of program components the individual services need to send and receive data in general formats.

Separation of structure from presentation

```
<From>Antonio Stadivarius</From>
<To>Domenico Scarlatti</To>
<Date>
        <Day>13</Day>
        <Month>4</Month>
        <Year>1723</Year>
</Date>
<Body>
Io bisogno una appartamento
acoglienti a Cremona …
</Body>
```

4/13/23

April 13, 1723

17. iv.1723

The computer can read the document and answer queries like this:
"Find all memos from April 1723"

XML

- Documents and data
- Human readable, editable, mailable
- Schema constrains structure
  - -- can encode data models
- Can be transformed (XSLT)
  - -- other xml
  - -- html/pdf/excel etc
- Tools
  - Parsers in Java, C, C++, Perl, Python, ...
  - Browsers and editors
  - XML databases
  - Binding to make API
- For serialization, mediation, brokers

XML for science

XML is a comfortable vehicle for our metadata and data models

But the real challenge is:

To define NVO-specific data objects
And how they are used

We need consensus
more than either software or hardware

VOTable
VOResource
services -- WSDL

XML example (no schema)

```
<?xml version="1.0"?>
<BookCatalogue>
  <Book>
    <Title>The Cambridge Star Atlas</Title>
    <Author>Wil Tirion</Author>
    <Publisher>Cambridge UP</Publisher>
  </Book>
  <Book>
    <Title> Parallel Computing Works!</Title>
    <Author>Geoffrey C. Fox</Author>
    <Author>Roy D. Williams</Author>
    <Author>Paul C. Messina</Author>
    <ISBN>1-55860-253-4</ISBN>
    <Publisher>Morgan Kaufmann</Publisher>
  </Book>
</BookCatalogue>
```

XML Parsing

SAX: Event-Based

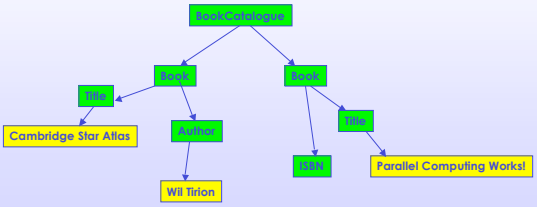
Handlers functions for StartElement, Text, EndElement, etc.

Found element BookCatalogue
Found element Book
Found Element Title
Found Text The Cambridge Star Atlas
Found End Element Title
....

Parsing

DOM: Document Object Model

Returns a tree-like Document object with data attached



XML Schema

```
<?xml version="1.0"?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema"
        xmlns:cat="uri://BookCatalogue">

  <element name="BookCatalogue">
    <complexType>
      <sequence>
          <element ref="cat:Book" minOccurs="0" maxOccurs="unbounded"/>
      </sequence>
    </complexType>
  </element>
  <element name="Book">
    <complexType>
      <sequence>
        <element ref="cat:Title"     minOccurs="1" maxOccurs="1"/>
        <element ref="cat:Author"    minOccurs="1"/>
        <element ref="cat:Date"      minOccurs="0" maxOccurs="1"/>
        <element ref="cat:ISBN"      minOccurs="1" maxOccurs="1"/>
        <element ref="cat:Publisher" minOccurs="1" maxOccurs="1"/>
      </sequence>
    </complexType>
  </element>
  <element name="Title" type="string"/>
  <element name="Author" type="string"/>
  <element name="Date" type="string"/>
  <element name="ISBN" type="string"/>
  <element name="Publisher" type="string"/>
</schema>
```

Book.xsd = Xml-Schema Definition

VOTable

- Full metadata representation
- Hierarchy of RESOURCEs
- containing PARAMs and TABLEs
- UCD (unified content descriptor)
  - a has **unit** meter
  - a has **UCD** *ORBIT_ SIZE_ SMAJ* (Semi-major axis of the orbit)
- Can reference remote and/or binary streams
  - Table can be
    - Pure XML
    - "Simple Binary"
    - FITS Binary Table

Sample VOTable

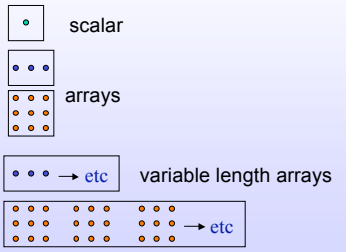
```
<?xml version="1.0"?>
<!DOCTYPE VOTABLE SYSTEM "http://us-vo.org/xml/VOTable.dtd">
<VOTABLE version="1.0">
 <DEFINITIONS>
  <COOSYS ID="myJ2000" equinox="2000." epoch="2000." system="eq_FK5"/>
 </DEFINITIONS>
  <RESOURCE>
  <PARAM name="Observer" datatype="char" arraysize="*" value="William Herschel">
   <DESCRIPTION>This parameter is designed to store the observer's name
   </DESCRIPTION>
  </PARAM>
  <TABLE name="Stars">
   <DESCRIPTION>Some bright stars</DESCRIPTION>
   <FIELD name="Star-Name" ucd="ID_MAIN" datatype="char" arraysize="10"/>
   <FIELD name="RA" ucd="POS_EQ_RA" ref="myJ2000" unit="deg"
          datatype="float" precision="F3" width="7"/>
   <FIELD name="Dec" ucd="POS_EQ_DEC" ref="myJ2000" unit="deg"
          datatype="float" precision="F3" width="7"/>
   <FIELD name="Counts" ucd="NUMBER" datatype="int" arraysize="2x3x*"/>
   <DATA>
    <TABLEDATA>
    <TR>
     <TD>Procyon</TD><TD>114.827</TD><TD>5.227</TD>
     <TD>4 5 3 4 3 2 1 2 3 3 5 6</TD>
    </TR>
    <TR>
     <TD>Vega</TD><TD>279.234</TD>
     <TD>38.782</TD><TD>8 7 8 6 8 6</TD>
    </TR>
    </TABLEDATA>
   </DATA>
  </TABLE>
 </RESOURCE>
</VOTABLE>
```

Observer = Herschel

	RA	Dec	

Table Cell

follows FITS binary table
does NOT follow XML schema



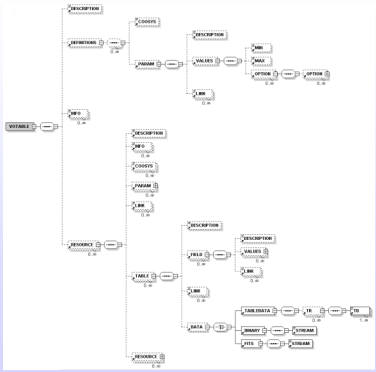
Primitives

- boolean
- bit
- unsignedByte
- short
- int
- long
- char
- unicodeChar
- float
- double
- floatComplex
- doubleComplex

VOTable is Flexible

- eg Table of images
  - UCD="meta.code.mime; image.jpeg" datatype="unsignedByte" arraysize="*"
- eg Table of URL links
  - UCD="meta.ref.url" datatype="char" arraysize="*"

VOTable Schema (xsd)



XSLT

- A language to filter XML documents
  - Eg XML → HTML
- Declarative, not procedural
- Written in XML

XSLT Example

```
<VOTABLE version="1.0">
 <DESCRIPTION>Output from the messier catalog at VirtualSky.org</DESCRIPTION>
 <RESOURCE type="results">
  <PARAM ID="RA" datatype="E" value="200.0" />
  <PARAM ID="DE" datatype="E" value="40.0" />
  <PARAM ID="SR" datatype="E" value="30.0" />
  <PARAM ID="PositionalError" datatype="E" value="0.1" />
  <PARAM ID="Credit" datatype="A" arraysize="*" value="Charles Messier, Richard Gelderman" />
  <TABLE>
   <DESCRIPTION>Output from messier Catalog Server</DESCRIPTION>
   <FIELD ID="I" name="Messier Number" datatype="char" arraysize="*" ucd="ID_MAIN">
    <DESCRIPTION>Messier Number</DESCRIPTION>
   </FIELD>
   <FIELD ID="RA" name="Right Ascension" datatype="float" unit="degrees" ucd="POS_EQ_RA_MAIN">
    <DESCRIPTION>Right Ascension J2000</DESCRIPTION>
   </FIELD>
....
   <DATA>
    <TABLEDATA>
     <TR>
      <TD>3</TD>  <TD>205.5</TD>  <TD>28.402</TD>  <TD />
      <TD>16.2</TD>  <TD>6.4004</TD>  <TD>Globular Cluster</TD>
      <TD>Canes Venatici</TD>  <TD>M3 is one of more heavily studied globular clusters due to its position in the galaxy, putting it far from interstellar absorbtion. More than 200 variable stars have been observed out of a total of near 50,000. Being one of the brightest clusters, M3 is</TD>
     </TR>
```

XSLT Result

this table is the result of a conesearch

Parameters

Name	Value
RA	200.0
DE	40.0
SR	30.0
PositionalError	0.1
Credit	Charles Messier, Richard Gelderman

Data

Messier Number	Right Ascension	Declination	Common Name	Object Size	Magnitude	Object classification	Constellation	Description
3	205.5	28.402		16.2'	6.4004	Globular Cluster	Canes Venatici	M3 is one of more heavily studied globular clusters due to its position in the galaxy, putting it far from interstellar absorbtion. More than 200 variable stars have been observed out of a total of near 50,000. Being one of the brightest clusters, M3 is
40	185.5999	58.07998	Double Star WHC4		9.599996	Double Star	Ursa Major	
85	186.31	18.19999		7.4'x5.7'	9.2011	Elliptical Galaxy	Coma Berenices	M85 is a fairly bright irregular galaxy of the Virgo cluster. It is believed to be about 40,000 light years across, with a mass of 100 billion suns.
								M51 is better known as the "Whalpool Galaxy" due to its pronounced spiral form. This was

XSLT Program

```
<h2>Data</h2>
        <table border="1">
        <xsl:for-each select="FIELD">
                <td><b><xsl:value-of select="@name" /> </b></td>
        </xsl:for-each>
        <xsl:for-each select="DATA">
        <xsl:for-each select="TABLEDATA">
                <xsl:for-each select="TR">
                        <tr>
                        <xsl:for-each select="TD">
                                <td width="100"><xsl:value-of select="." /></td>
                        </xsl:for-each>
                        </tr>
                </xsl:for-each>
        </xsl:for-each>
        </xsl:for-each>
        </table>
```

Questions?