A Framework for Dynamic Data Source Identification and Orchestration on the Web

Alexander Berezovskiy and Dr Leslie Carr

4th International Workshop on Web APIs and Service Mashups
European Conference on Web Services (ECOWS 2010)

1 December 2010, Ayia Napa, Cyprus
Problem

- Modern Web can be seen as a “universal data machine”
  - Multitude of web applications, services and data provides
  - Almost every user has vast amount of data stored online
- Data is often duplicated
  - Users need to register, provide their details, upload photos...
- Ideally, we would like a way to universally manipulate the data
  - APIs are different, functionality depends on data source
  - We need to know what data source to use
  - Make adjustments for every data source
Current solutions

- OpenSocial
  - Focused on social applications
  - Requires adjustments in target applications
- Integrate every possible resource
  - Very time consuming
  - No guarantee the user will be satisfied with the choice
  - User interaction required to choose the right source
Proposed solution

• Identify the most appropriate data resource
  • Task is given, nature of required data is known
  • Without any user intervention attempt to identify most suitable data resource to perform the given task

• Execute the task (process data request)
  • When data resource is identified, hand the request over to the resource and report on the results (or, return results)
  • Allow full CRUD (create, read, update, delete) operations on data
Proposed solution (cont'd)
Data resource identification

• One visit to a web page can tell a lot about the user
  • Country, language, browser, operating system, ...
  • We assume all these parameters affect the user preferences and their choice of applications

• Use two-dimensional model:
  • Information about the user
    – **Country, Language**, Age, Occupation, Marital Status, ...
  • Usage information
    – **Browser, Operating System**, Web and Local Applications, ...
  • Some information can be obtained from a single HTTP request with no user intervention required
Data resource identification (cont'd)

- User information is grouped in a single entity called Environment
- Data can be structured as a tree:

![Diagram showing data resource identification]

- Identification algorithm:
  Total score for data source $a$ is defined as:
  \[
  TSA(a) = TRS(a) + ERS(a) + TRAA(a)
  \]

\[
TRS(a) \quad - \text{Total Relevance Score for } a
\]

\[
ERS(a, u) \quad - \text{Environment Relevance Score for } a \text{ and user } u
\]

\[
TRAA(a) \quad - \text{Total Relevance Application-to-Application score for } a
\]
Data operations

• Each data resource serves data differently
• Data operations are performed by “bindings”
  • These are small chunks of code executed independently
  • We need at least one for each data resource
  • They can be written by anyone
Data operations (cont'd)

• Bindings return data in “raw” format
• The data can then be converted to almost any format
  • Currently available XML, JSON and Plain Text
  • Can be adjusted to serve RDF
• Example binding to return a user's name from Facebook:

```python
import urllib
import simplejson

interface = {
    'fields': {'username': {'required': 'yes', 'type': 'text'}},
    'formats': ['html', 'xml', 'txt']
}

def run_binding():
    url = 'http://graph.facebook.com/' + str(job.input_args['username'][0])
    response = urllib.urlopen(url)
    user = simplejson.loads(response.read())
    return user['name']
```
Demo

Applications

Use Service Map to navigate around things available at Nuvius, and narrow down the list of recommended applications.

Business
- Collaboration
- Finance
- Project management

Communication
- Authentication
- Blogging
- Messaging
- Social networking
  - Friend list
  - News
  - Profile

Information
- Encyclopedia
- News
- Article

Multimedia
- Audio
- Station
- Gaming
- Photo sharing
- Video
- Channel

Productivity
- Bookmarking
- Calendar
- Event
- Contact book
- Contact
- Data storage
  - Folder
- Notebook
  - Note
  - Office
  - Document
  - Presentation
  - Spreadsheet

Science & Education
- Distance Teaching
- Research
- Paper

About | Terms of Use
Demo

Applications

Recommended  Most Recent  Service Map

Use Service Map to navigate around things available at Nuvius, and narrow down the list of recommended applications.

Productivity

Calendar

Event

Attendee  Author  Date  Details  ID  Place  Title  Updated

Google Calendar

Google Calendar is a free time-management web application offered by Google. Its features include calendar sharing, device synchronization, SMS alarms and others. Users are required to have a Google Account in order to use the software.

Add to Library  Similar applications  Data for Event

Hardtree

Hardtree is an online, integrated business and IT management suite. Hardtree comes with a multitude of modules including Service Support, Project Tracking and Sales Management and other.

Add to Library  Similar applications  Data for Event

Tungle.me
Demo

Data for Event

- Using application: [Google Calendar]
- Binding: [GoogleCalendar by leboosh]
- Interface: [Calendar / Event: Generic Interface]
- Format: HTML

The result is:

1. attendee:
   - status:
   - picture:
   - name: leboosh@gmail.com

2. title: Mashups 10

3. author:
   - picture:
   - name: Alex Bezovskly
   - email: leboosh@gmail.com

4. updated: (2010, 11, 30, 11, 2, 56, 1, 334, 0)

5. place:
   - latitude:
   - name: None
   - longitude:

6. details: None

7. date:
   - start: (2010, 11, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
   - end: (2010, 11, 1, 17, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0)
Demo

Overview / Everpick

Twitter
Connect Twitter →

Facebook
Connect Facebook →

Applications:
Facebook
Twitter

Everpick is a giteso product
Demo

Time for a relaxing train journey back to Bath after a fun weekend at Southampton and Home.

Oliver Bills  30 November 2010, 11:22

Warming up for #online10 - "Aim for the Stars" no surprise its open data and the virtues of linking it #opendata #linkeddata #datagovuk

Nigel Shadbolt  30 November 2010, 11:12

"очень часто посещающие мысли :)

http://www.youtube.com/watch?v=bYw7PR86upk

Alina Klokova  30 November 2010, 10:32

Grr, reading documentation off a website which has just failed. Unhelpful!

Peter Brewer  30 November 2010, 10:17

-opendata Nigel Shadbolt calls for university computer science courses to prioritise open data and develop skills. http://livefook.org
Actor Leslie Nielsen Dies at 84

29 November 2010, 13:55

An anonymous reader writes "Famous actor Leslie Nielsen died of pneumonia at a Florida hospital Sunday evening. Leslie was renowned for his comedic roles in dozens of films and TV shows, such as The Naked Gun and Police Squad. His characteristic style and humor was always enjoyable, and he will be sorely missed." Of course you might also remember him from The Forbidden Planet, which is classic sci-fi by any measure.

Read more of this story at Slashdot.

Microsoft Patents Shape-Shifting Display

29 November 2010, 14:50

Scoobalou writes "In layman's terms, Microsoft's patent is for a special type of touch-screen display which includes a 'shape-memory' layer at its base. When activated by a special frequency of ultraviolet light, individual blocks — not coincidentally the same size as a pixel on the display port — can be raised or lowered, lending the displayed image physical texture."

Demo
Demo
Hardtree needs access to your profile

- Hardtree asks permission to write to your Contact book / Contact data
- Hardtree asks permission to read your Contact book / Contact data

With selected: [✓] Approve  [✗] Decline  [@] Remove
Discussion

• Dynamic discovery of data sources
  • Data mining can help us read data
  • How can we do full CRUD on the data?

• Universal data addressing
  • How can we universally address data based on its nature?

• Semantic Web application
  • Can we derive ontologies from the available data?
Thank you