Building Mashups by Example

Rattapoom Tuchinda
Pedro Szekely
Craig A. Knoblock

Information Sciences Institute
University of Southern California
What’s a Mashup?

A website or application that combines content from more than one source into an integrated experience [wikipedia]

a) LA crime map  
b) zillow.com  
c) Ski bonk

Combined Data gives new insight / provides new services
Mashup Building Issues

Data Retrieval

Calibration
- source modeling
- cleaning

Integration

Display

Wrapper → Attribute → Clean → Combine → Customize Display

Wrapper → Attribute → Clean
Goal: Create Mashups without Programming
- Addresses syntax issues, but users still required to understand programming concepts

Existing Approaches

Yahoo’s Pipes

Widget Paradigm
- Widgets (i.e., 43 for Pipes, 300+ for MS) represents an operation on the data.
- Locating and learning to customize widget can be time consuming
- Most tools focus on particular issues and ignore others.

Can we come up with a framework that addresses all of the issues while still making the Mashup building process easy?
Our Integrated Approach

• Focus on data, not on the process
  – Users are already familiar with data.
  – Capture and model the Mashup building process from examples (PBD)

• Consolidate rather than Divide-And-Conquer
  – Solving one issue can help solve other issues.
  – Use one interaction platform -- a table

• Leverage existing database
  – Helps with source modeling, cleaning, and data integration.
Our system: Karma

**Embedded Browser**

**Table**

**Interaction Modes**
Our system: Karma

Embedded Browser

Table

Interaction Modes
Our system: Karma

Embedded Browser

Table

Interaction Modes
Extract

{Restaurant name, address, phone, Review}

Clean

{Restaurant name, address, phone, review, Date of Inspection, Score}

Map
Extract

{Restaurant name, address, phone, Review}

Clean

{Restaurant name, address, Date of Inspection, Score}

Clean

{Restaurant name, address, phone, review, Date of Inspection, Score}

Map
1. Japon Bistro
927 E Colorado Blvd., Pasadena, CA, 91106
Upscale yet affordable Japanese eatery offers the city’s largest sake selection.

2. Hokusai
8400 Wilshire Blvd., Beverly Hills, CA, 90211
Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.

3. Sushi Sasabune
12400 Wilshire Blvd. Ste 150, Los Angeles, CA, 90025
Sushi is the singular star at this Westside palace that bows only to the royalty of chef and fish.

4. Sushi Roku
8445 W 3rd St, Los Angeles, CA, 90048
High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

```
select one
<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Japon Bistro</td>
</tr>
<tr>
<td>Hokusai</td>
</tr>
</tbody>
</table>
```

Diagram:

```
<tboby>
  <tr>
    <td>
      1. Japon Bistro
      970 E Colorado St.
      Upscale yet affordable.
    </td>
    <td>
      2. Hokusai
      8400 Wilshire Blvd.
      Chic elegance.
    </td>
  </tr>
</tboby>
```
1. **Japon Bistro**
   970 E Colorado Blvd., Pasadena, CA, 91106
   Upscale yet affordable Japanese eatery offers the city’s largest sake selection.

2. **Hokusai**
   8400 Wilshire Blvd, Beverly Hills, CA, 90211
   Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.

3. **Sushi Sasabune**
   12400 Wilshire Blvd Ste 150, Los Angeles, CA, 90025
   Sushi is the singular star at this Zen Westside palace that bows only to the royalty of chef and fish.

4. **Sushi Roku**
   8445 W 3rd St, Los Angeles, CA, 90048
   High fashion, rock and roll and Hollywood buzz converge over innovative sushi.
Data Retrieval: Navigation

1. **Japon Bistro**
   927 E Colorado Blvd, Pasadena, CA, 91106
   Upscale yet affordable Japanese eatery offers the city’s largest sake selection.

2. **Hokusai**
   8400 Wilshire Blvd, Beverly Hills, CA, 90211
   Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.

3. **Sushi Sasabune**
   12400 Wilshire Blvd Ste 150, Los Angeles, CA, 90025
   Sushi is the singular star at this Zen Westside palace that bows only to the royalty of chef and fish.

4. **Sushi Roku**
   8445 W 3rd St, Los Angeles, CA, 90048
   High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

---

**Japon Bistro**
927 E Colorado Blvd
Pasadena, CA 91106
Phone: (626) 744-1751

---

<table>
<thead>
<tr>
<th>select one</th>
<th>address</th>
<th>select one</th>
<th>Select one</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japon Bistro</td>
<td>927 E Coloro..</td>
<td>Upscale yet affordab</td>
<td>31 Reviews</td>
</tr>
<tr>
<td>Sushi Dokoro Ki Ra La</td>
<td>9777 S Santa Monica Blvd, Beverly Hills, CA 90211</td>
<td>Intimate and charming Japanese restaurant offers wide range of hand-selected sushi and sashimi.</td>
<td></td>
</tr>
<tr>
<td>Hokusai</td>
<td>8400 Wilshire...</td>
<td>Chic elegance...</td>
<td></td>
</tr>
<tr>
<td>Sushi Sasabune</td>
<td>12400 Wilshire...</td>
<td>Authentic Ja...</td>
<td></td>
</tr>
<tr>
<td>Sushi Roku</td>
<td>8445 W 3rd...</td>
<td>High fashion...</td>
<td></td>
</tr>
<tr>
<td>Hide Sushi</td>
<td>2040 Sawtel...</td>
<td>No fuss, jus...</td>
<td></td>
</tr>
<tr>
<td>Fat Fish</td>
<td>616 N Rober...</td>
<td>Inventive ro...</td>
<td></td>
</tr>
<tr>
<td>Sushi Katsu-ya</td>
<td>11680 Vent...</td>
<td>The MOCA o...</td>
<td></td>
</tr>
<tr>
<td>Gindi Thai</td>
<td>4017 W Riv...</td>
<td>Burbank res...</td>
<td></td>
</tr>
<tr>
<td>Katana</td>
<td>8439 W Sun...</td>
<td>Rustic Japa...</td>
<td></td>
</tr>
<tr>
<td>Echigo</td>
<td>11217 Sant...</td>
<td>Stellar sushi...</td>
<td></td>
</tr>
</tbody>
</table>
Data Retrieval: Navigation

1. **Japon Bistro**
   927 E Colorado Blvd, Pasadena, CA, 91106
   Upscale yet affordable Japanese eatery offers the city's largest sake selection.

2. **Hokusai**
   8400 Wilsshire Blvd, Beverly Hills, CA, 90211
   Chic elegance and modern Zen style surround Japanese French this paeon to haute cuisine and stylized sushi.

3. **Sushi Sasabune**
   12400 Wilshire Blvd Ste 150, Los Angeles, CA, 90025
   Sushi is the singular star at this Zen Westside palace that bows only to the royalty of chef and fish.

4. **Sushi Roku**
   8445 W 3rd St, Los Angeles, CA, 90048
   High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

---

### Japon Bistro

**970 E Colorado..**
Upscale yet affordable

---

### Japon Bistro

**927 E Colora..**
Upscale yet affordable

---

### Hokusai

**8400 Wilshir..**
Chic elegance...

---

<table>
<thead>
<tr>
<th>select one</th>
<th>address</th>
<th>select one</th>
<th>Select one</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japon Bistro</td>
<td>927 E Colora..</td>
<td>Upscale yet</td>
<td>31 Reviews</td>
</tr>
<tr>
<td>Sushi Dokor..</td>
<td>9777 S Sant..</td>
<td>Intimate an..</td>
<td></td>
</tr>
<tr>
<td>Hokusai</td>
<td>8400 Wilshir..</td>
<td>Chic eleganc..</td>
<td></td>
</tr>
<tr>
<td>Sushi Sasabune</td>
<td>12400 Wilshi..</td>
<td>Authentic Ja..</td>
<td></td>
</tr>
<tr>
<td>Sushi Roku</td>
<td>8445 W 3rd..</td>
<td>High fashion..</td>
<td></td>
</tr>
<tr>
<td>Hide Sushi</td>
<td>2040 Sawtel..</td>
<td>No fuss, jus..</td>
<td></td>
</tr>
<tr>
<td>Fat Fish</td>
<td>616 N Rober..</td>
<td>Inventive ro..</td>
<td></td>
</tr>
<tr>
<td>Sushi Katsu-ya</td>
<td>11680 Vent..</td>
<td>The MOCA o..</td>
<td></td>
</tr>
<tr>
<td>Gindi Thai /..</td>
<td>4017 W Riv..</td>
<td>Burbank res..</td>
<td></td>
</tr>
<tr>
<td>Katana</td>
<td>8439 W Sun..</td>
<td>Rustic Japa..</td>
<td></td>
</tr>
<tr>
<td>Echigo</td>
<td>11217 Sant..</td>
<td>Stellar sushi..</td>
<td></td>
</tr>
</tbody>
</table>
Data Retrieval: Navigation

1. **Japon Bistro**
   927 E Colorado Blvd., Pasadena, CA, 91106
   Upscale yet affordable Japanese eatery offers the city's largest sake selection.

2. **Hokusai**
   8400 W crisps Blvd., Beverly Hills, CA, 90211
   Chic elegance and modern Zen style surround Japanese French this paeo to haute cuisine and stylized sushi.

3. **Sushi Sasabune**
   12400 W 3rd St., Los Angeles, CA, 90025
   Sushi is the singular star at this Zen Westside palace that bows only to the royalty of chef and fish.

4. **Sushi Roku**
   8445 W 3rd St., Los Angeles, CA, 90048
   High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

---

**Table**: Menu Items

<table>
<thead>
<tr>
<th>Select One</th>
<th>Address</th>
<th>Select One</th>
<th>Select One</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japon Bistro</td>
<td>927 E Colora..</td>
<td>Upscale yet affordab</td>
<td>31 Reviews</td>
</tr>
<tr>
<td>Sushi Dokoro</td>
<td>9777 S Santa Mon..</td>
<td>Intimate an..</td>
<td>3 Reviews</td>
</tr>
<tr>
<td>Hokusai</td>
<td>8400 Wilshir..</td>
<td>Chic eleganc..</td>
<td>60 Reviews</td>
</tr>
<tr>
<td>Sushi Sasab..</td>
<td>12400 Wilsh..</td>
<td>Authentic Ja..</td>
<td>66 Reviews</td>
</tr>
<tr>
<td>Sushi Roku</td>
<td>8445 W 3rd..</td>
<td>High fashion..</td>
<td>52 Reviews</td>
</tr>
<tr>
<td>Hide Sushi</td>
<td>2040 Sawtel..</td>
<td>No fuss, jus..</td>
<td>25 Reviews</td>
</tr>
<tr>
<td>Fat Fish</td>
<td>616 N Rober..</td>
<td>Inventive ro..</td>
<td>38 Reviews</td>
</tr>
<tr>
<td>Sushi Katsu-ya</td>
<td>11680 Vent..</td>
<td>The MOCA o..</td>
<td>49 Reviews</td>
</tr>
<tr>
<td>Gindi Thai</td>
<td>4017 W Riv..</td>
<td>Burbank res..</td>
<td>29 Reviews</td>
</tr>
<tr>
<td>Katana</td>
<td>8439 W Sun..</td>
<td>Rustic Japa..</td>
<td>96 Reviews</td>
</tr>
<tr>
<td>Echigo</td>
<td>11217 Sant..</td>
<td>Stellar sushi..</td>
<td>49 Reviews</td>
</tr>
</tbody>
</table>
Source Modeling (Attribute selection)

Possible Attribute
\{a \mid a, s: a \in att(s) \land (val(a, s) \subseteq V)\}

restaurant name (3)
artist name (1)
Data Cleaning: using existing values

Newly extracted data

- Japon Bistro
- Hokusai
- Sushi Sasabune
- Sushi Roka

Data repository

<table>
<thead>
<tr>
<th>restaurant name</th>
<th>Address</th>
<th>...</th>
<th>Health Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hokusai</td>
<td>8400..</td>
<td>...</td>
<td>90</td>
</tr>
<tr>
<td>Katana</td>
<td>8439..</td>
<td>...</td>
<td>99</td>
</tr>
<tr>
<td>Japon Bistro</td>
<td>927 E..</td>
<td>...</td>
<td>95</td>
</tr>
</tbody>
</table>

Zagat

<table>
<thead>
<tr>
<th>restaurant name</th>
<th>zagat Rating</th>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sushi Sasabune</td>
<td>27</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Sushi Roku</td>
<td>25</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Katana</td>
<td>23</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Data Cleaning: using predefined rules

28 Reviews → 28

Subset Rule:
\((s_1 s_2 \ldots s_k) \rightarrow (d_1 d_2 \ldots d_t) \land (k \leq t) \land s_i \in \{d_1, d_2, \ldots, d_t\} \land d_i \neq d_j\)
Data Integration
Based on [tuchinda 2007]
\( \{a\}_R \) = possible new attribute selection for row \( i \).
\( \{x\} \) = Set intersection(\( \{a\} \)) over all the value rows.
\( \{v\} = \text{val}(a, s) \text{ where } a \{x\} \)
\( s \) is any source where \( att(s) \{x\} \neq \{\} \)
Map Generation

restauran_name: Hokusai
address: 8400 Wilshire Blvd, Beverly Hills, CA, 90211
description: Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.
number_of_reviews: 30
health_rating: 90
Related Work

• Building Blocks
  – Programming by Demonstration [Cypher 1993, Lau 2001]
  – Simile [Huynh 2005]
  – Potter’s Wheel [Raman 2001]
  – Building Queries by Demonstration [Tuchinda 2007]

• Data Extraction
  – Simile, Dapper, D.Mix [Hartman 2007], OpenKapow

• Widget Approach
  – Yahoo’s Pipes, Microsoft’s Popfly, IBM’s QED Wiki, Bungee Labs, Proto Software, Marmite [Wong 2007]

• Misc.
  – Intel’s Mashmaker [Ennals 2007]
  – Google MyMap
3 Mashup building tasks
Compare Karma (K) and Dapper + Yahoo (DP).
Record the number of steps:
- typing values in textbox, clicking a button, select an option from a list, dragging and dropping, and connecting one widget to one another
Conclusion

• Contribution: An approach to build Mashups by combining four common information integration techniques into a unified framework.
  – Data extraction
  – Source modeling
  – Data Cleaning
  – Data Integration

• Future work:
  – Data display
  – In depth evaluation
  – Dynamic web data
  – PBD issues