

Building Mashups by Example

Rattapoom Tuchinda

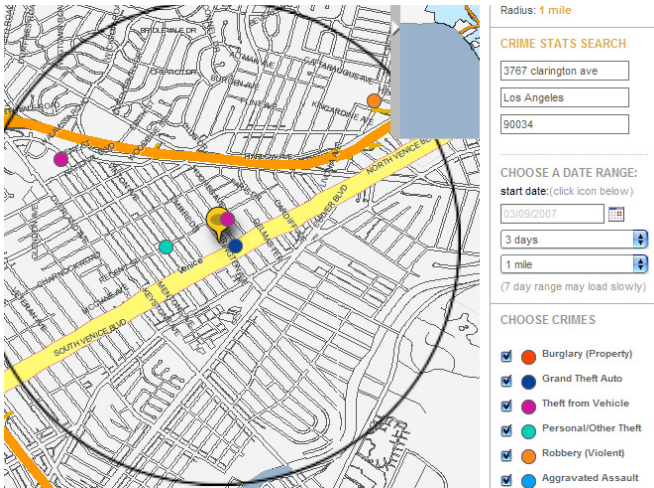
Pedro Szekely

Craig A. Knoblock

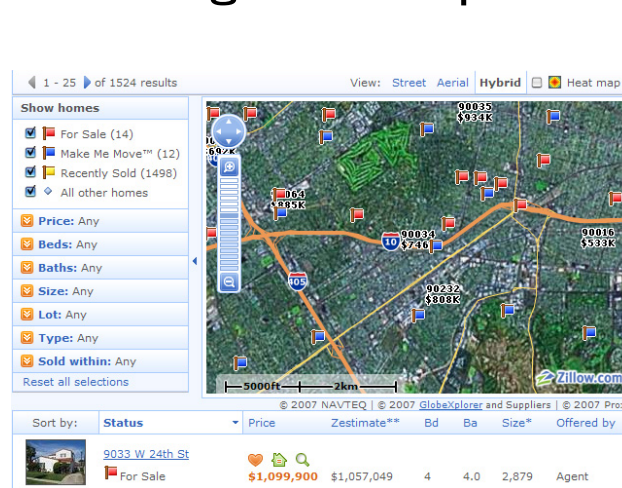
Information Sciences Institute
University of Southern California

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



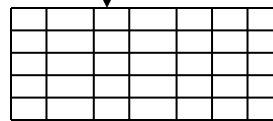
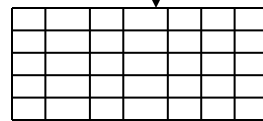
Mashup Building Issues

Data Retrieval



Wrapper

Wrapper



Calibration
-source modeling
-cleaning

Attribute

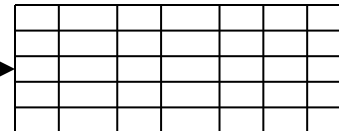
Attribute

Clean

Clean

Integration

Combine



Display

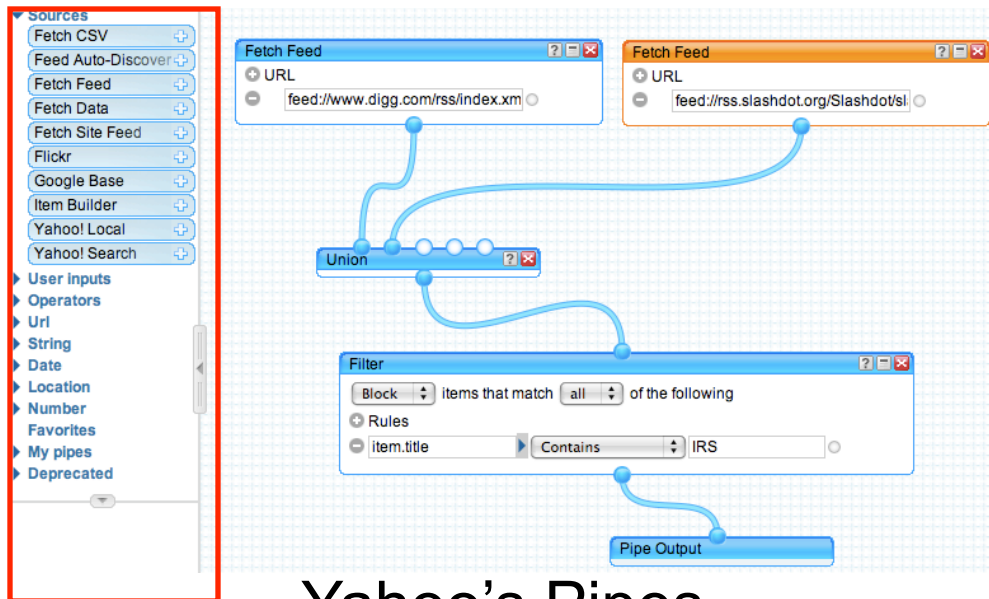
Customize Display



Existing Approaches

Goal: Create Mashups without Programming

- Addresses syntax issues, but users still required to understand programming concepts



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

The screenshot displays the Karma system interface, which is divided into two main sections: an embedded browser on the left and a table/data management interface on the right.

Embedded Browser: The browser window shows a search result for "Best Sushi 2007" in Los Angeles. The URL is "losangeles.citysearch.com/bestof/winners/sushi?init_search=1". The page content includes a "Best Sushi 2007" header, a "Send to a friend" link, and a list of seven restaurants. The first restaurant, "Japon Bistro", is highlighted with a pink box. A pink arrow points from this box to the table on the right.

Table: The table on the right contains a list of restaurant names in the first column and several empty columns. The names listed are: Japon Bistro, Sushi Dokor..., Hokusai, Sushi Sasab..., Sushi Roku, Hide Sushi, Fat Fish, Sushi Katsu-ya, Gindi Thai / ..., Katana, and Echigo.

Data Management Interface: Below the table, there are tabs for "Table", "Attributes", "Data cleaning", "Integration", "Save", and "Load". The "Data cleaning" tab is active, showing a "Clean column" dropdown set to "Column 0", and "Start cleaning" and "Finish" buttons. Below this, there are checkboxes for "Final result is:" with options: "Use extracted values" (checked), "Use suggested values (override extracted values)", and "Use user defined values (override the first two options)". An "Update" button is also present.

Interaction Modes



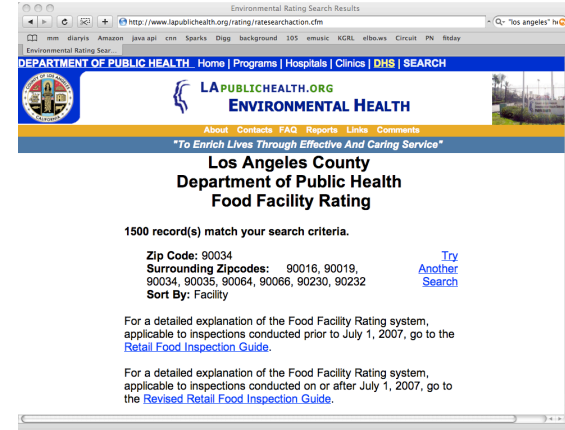
Extract

{Restaurant name, address, phone, Review}

Clean

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

Map



Extract

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

Clean

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

Map



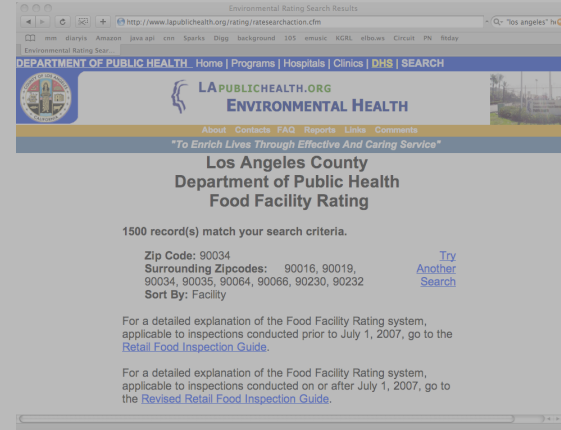
Extract

{Restaurant name, address, phone, Review}

Clean

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

Map



Extract

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

Clean

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

Map

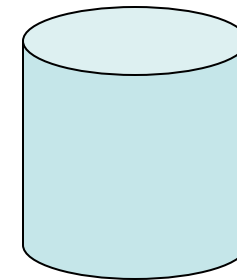


Extract

{Restaurant name, address, phone, Review}

Clean

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



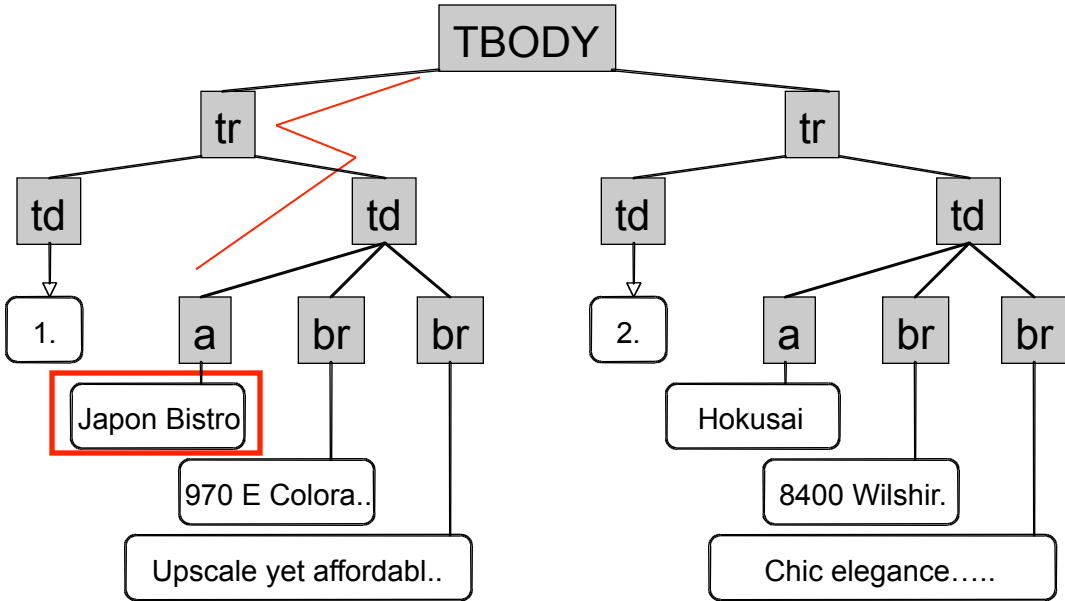
Database

Map

Data Retrieval: Extraction

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 paeen 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.

select one			
Japon Bistro			



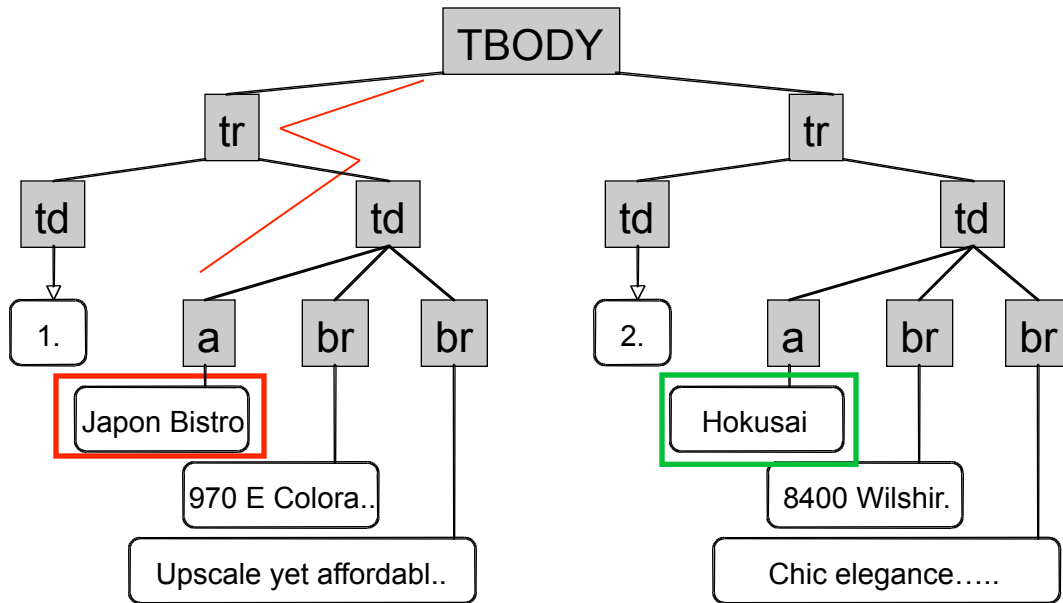
Tbody/tr[1]/td[2]/a

Data Retrieval: Extraction

- Japon Bistro**
 927 E Colorado Blvd , Pasadena , CA , 91106
 Upscale yet affordable Japanese eatery offers the city's largest sake selection.
- Hokusai**
 8400 Wilshire Blvd , Beverly Hills , CA , 90211
 Chic elegance and modern Zen style surround Japanese French this paeen to haute cuisine and stylized sushi.
- 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.
- Sushi Roku**
 8445 W 3rd St , Los Angeles , CA , 90048
 High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

Sushi Dokoro Ki Ra La
 9777 S Santa Monica Blvd , Beverly Hills , CA , 90211
 Intimate and charming Japanese restaurant offers wide range of hand-selected sushi and sashimi.

select one			
Japon Bistro			
Sushi Dokor..			
Hokusai			
Sushi Sasab..			
Sushi Roku			
Hide Sushi			
Fat Fish			
Sushi Katsu-ya			
Gindi Thai /..			
Katana			
Echigo			



Tbody/tr[1]/td[2]/a

Tbody/tr*/td*/a

Data Retrieval: Navigation

- Japon Bistro**
 927 E Colorado Blvd , Pasadena , CA , 91106
 Upscale yet affordable Japanese eatery offers the city's largest sake selection.
- 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.
- 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.
- Sushi Roku**
 8445 W 3rd St , Los Angeles , CA , 90048
 High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

Sushi Dokoro Ki Ra La
 9777 S Santa Monica Blvd , Beverly Hills , CA , 90211
 Intimate and charming Japanese restaurant offers wide range of hand-selected sushi and sashimi.

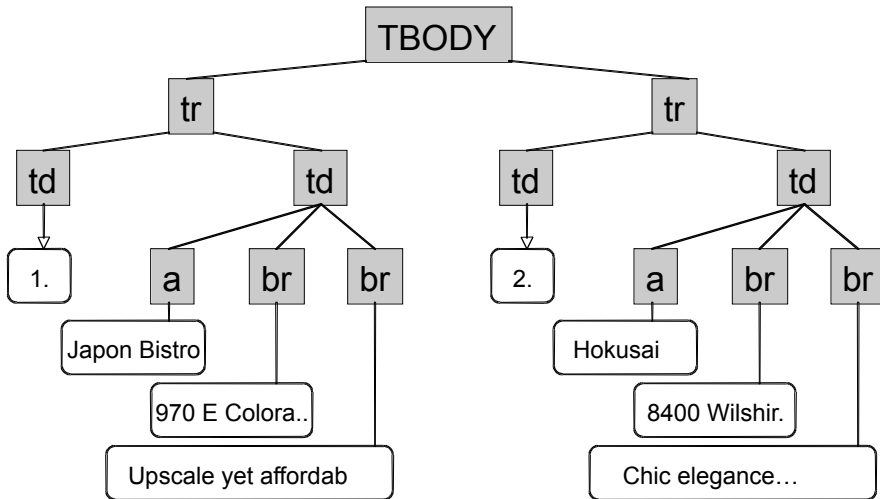
Japon Bistro

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

★★★★★
 31 Reviews | Write a Review



- Map
- Print
- Website
- Email to a Friend
- SMS/Text
- Email the Business
- Promote this business
- Link to this page
- Have Updated Info?
- Suggest a correction



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

Sushi Dokoro Ki Ra La
 9777 S Santa Monica Blvd , Beverly Hills , CA , 90211
 Intimate and charming Japanese restaurant offers wide range of hand-selected sushi and sashimi.

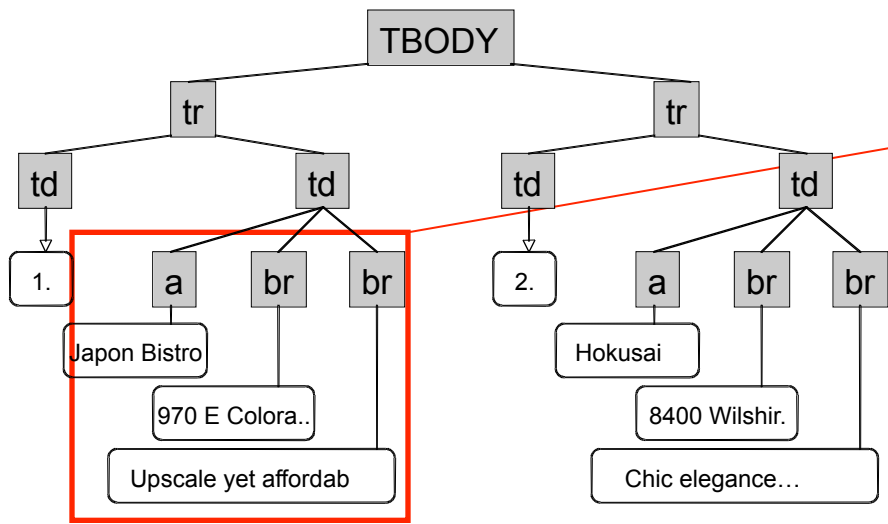
Japon Bistro

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

★★★★★
 31 Reviews | Write a Review



- Map
- Print
- Website
- Email to a Friend
- SMS/Text
- Email the Business
- Promote this business
- Link to this page
- Have Updated Info?
- Suggest a correction



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

Sushi Dokoro Ki Ra La
9777 S Santa Monica Blvd , Beverly Hills , CA , 90211
Intimate and charming Japanese restaurant offers wide range of hand-selected sushi and sashimi.

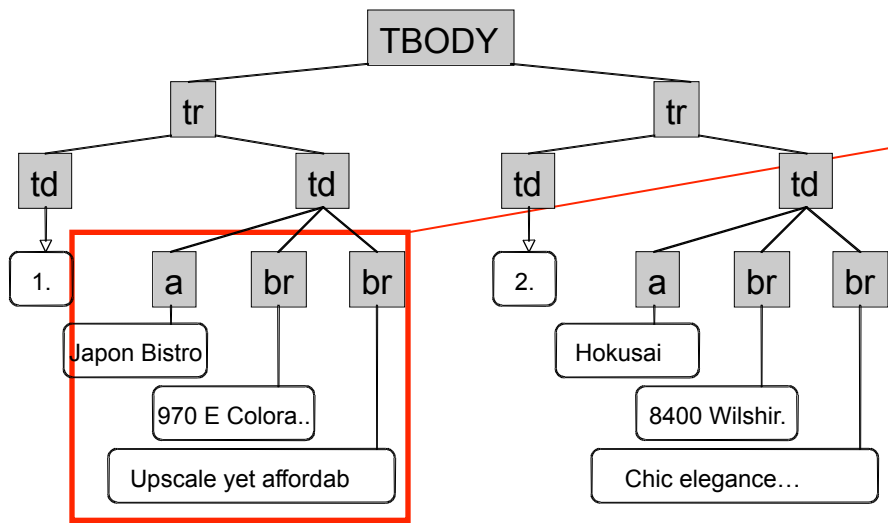
Japon Bistro

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

★★★★★
31 Reviews | Write a Review

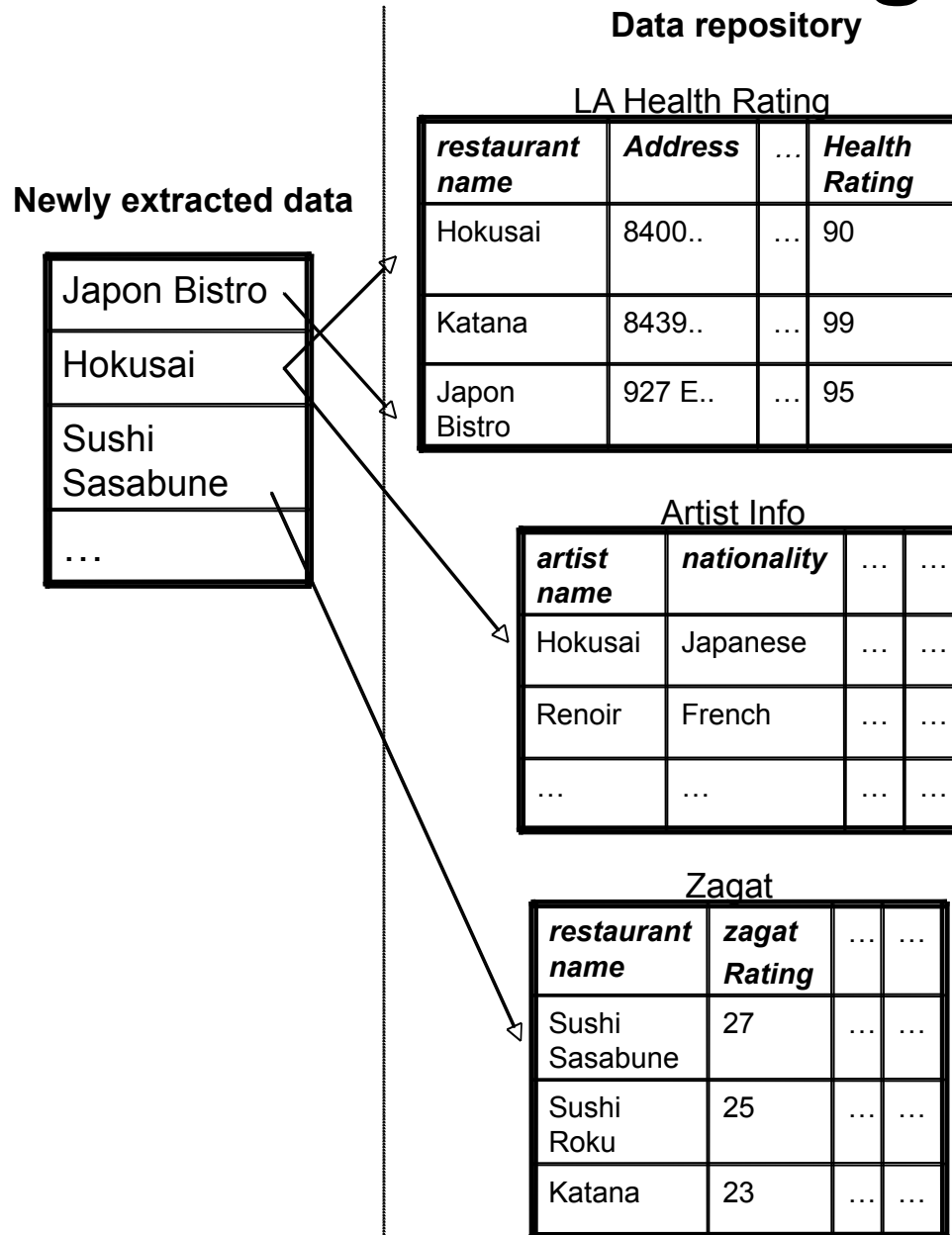


- Map
- Print
- Website
- Email to a Friend
- SMS/Text
- Email the Business
- Promote this business
- Link to this page
- Have Updated Info?
- Suggest a correction



select one	address	select one	select one
Japon Bistro	927 E Color..	Upscale yet...	31 Reviews
Sushi Dokor..	9777 S Sant..	Intimate an...	3 Reviews
Hokusai	8400 Wilshir...	Chic eleganc...	30 Reviews
Sushi Sasab..	12400 Wilshi...	Authentic Ja...	66 Reviews
Sushi Roku	8445 W 3rd...	High fashion ...	62 Reviews
Hide Sushi	2040 Sawtel...	No fuss, jus...	25 Reviews
Fat Fish	616 N Rober...	Inventive ro...	38 Reviews
Sushi Katsu-ya	11680 Vent...	The MOCA o...	49 Reviews
Gindi Thai /..	4017 W Riv...	Burbank res...	29 Reviews
Katana	8439 W Sun...	Rustic Japa...	96 Reviews
Echigo	11217 Sant...	Stellar sushi...	49 Reviews

Source Modeling (Attribute selection)



Possible Attribute

$$\{a \mid a, s: a \in \text{att}(s) \wedge (\text{val}(a, s) \subset V)\}$$

restaurant name (3)

artist name (1)

Data Cleaning: using existing values

Newly extracted data

Japon Bistro
Hokusai
Sushi Sasabune
Sushi Roka

restaurant name

Data repository

LA Health Rating

<i>restaurant name</i>	<i>Address</i>	...	<i>Health Rating</i>
Hokusai	8400..	...	90
Katana	8439..	...	99
Japon Bistro	927 E..	...	95

Zagat

<i>restaurant name</i>	<i>zagat Rating</i>
Sushi Sasabune	27
Sushi Roku	25
Katana	23

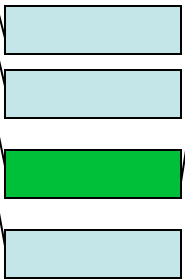
Data Cleaning: using predefined rules

s	description	number o...	suggest	user defi...	final
lora...	Upscale yet ...	28 Reviews		28	
ant...	Intimate and...	3 Reviews			
shir...	Chic eleganc...	30 Reviews			
ilshi...	Authentic Ja...	66 Reviews			
3rd ...	High fashion...	62 Reviews			
vtell...	No fuss, just...	25 Reviews			
iber...	Inventive rol...	38 Reviews			
ntu...	The MOCA o...	49 Reviews			
Rive...	Burbank rest...	29 Reviews			
5un...	Rustic Japan...	96 Reviews			
anta...	Stellar sushi ...	49 Reviews			

28 Reviews → 28

Subset Rule:

$$\begin{aligned}
 &(s_1 s_2 \dots s_k) \rightarrow (d_1 d_2 \dots d_t) \wedge \\
 &(k \leq t) \wedge \\
 &s_i \in \{d_1, d_2, \dots, d_t\} \wedge \\
 &d_i \neq d_j
 \end{aligned}$$



Predefined Rules

Data Integration

Based on [tuchinda 2007]

restaurant...	address	description	number o...		
Japon Bistro	927 E Color...	Upscale yet ...	28 Reviews		
Sushi Dokor...	9777 S Sant...	Intimate an...	3 Reviews		
Hokusai	8400 Wilshir...	Chic eleganc...	30 Reviews		
Sushi Sasab...	12400 Wilshi...	Authentic Ja...	66 Reviews		
Sushi Roku	8445 W 3rd ...	High fashion...	62 Reviews		
Hide Sushi	2040 Sawtel...	No fuss, jus...	25 Reviews		
Fat Fish	616 N Rober...	Inventive ro...	38 Reviews		
Sushi Katsu-ya	11680 Vent...	The MOCA o...	49 Reviews		
Gindi Thai / ...	4017 W Riv...	Burbank res...	29 Reviews		
Katana	8439 W Sun...	Rustic Japa...	96 Reviews		
Echigo	12217 Sant...	Stellar sushi ...	49 Reviews		

able | Attributes | Clean Data | **Integration** | Save | Load

Activate Data Integration

Suggestions:

Cell:

Data Integration (cont.)

restaurant...	address	description	number o...	
Japon Bistro	927 E Color...	Upscale yet ...	28 Reviews	
Sushi Dokor...	9777 S Sant...	Intimate an...	3 Reviews	
Hokusai	8400 Wilshir...	Chic eleganc...	30 Reviews	
Sushi Sasab...	12400 Wilshi...	Authentic Ja...	66 Reviews	
Sushi Roku	8445 W 3rd ..	High fashion...	62 Reviews	
Hide Sushi	2040 Sawtel..	No fuss, jus...	25 Reviews	
Fat Fish	616 N Rober...	Inventive ro...	38 Reviews	
Sushi Katsu-ya	11680 Vent...	The MOCA o...	49 Reviews	
Gindi Thai / ...	4017 W Riv...	Burbank res...	29 Reviews	
Katana	8439 W Sun...	Rustic Japa...	96 Reviews	
Echigo	12217 Sant...	Stellar sushi ...	49 Reviews	

Data repository

LA Health Rating

<i>restaurant name</i>	<i>Address</i>	<i>Health Rating</i>
Hokusai	8400	90
Katana	8439...	99
Japon Bistro	927 E..	95

Zagat

<i>restaurant name</i>	<i>zagat Rating</i>
Sushi Sasabune	27
Sushi Roku	25
Katana	23

$\{a\}_R =$ possible new attribute selection for row i .

$\{x\} =$ Set intersection($\{a\}$) over all the value rows.

$\{v\} = \text{val}(a,s)$ where $a \in \{x\}$

s is any source where $\text{att}(s) \cap \{x\} \neq \{\}$

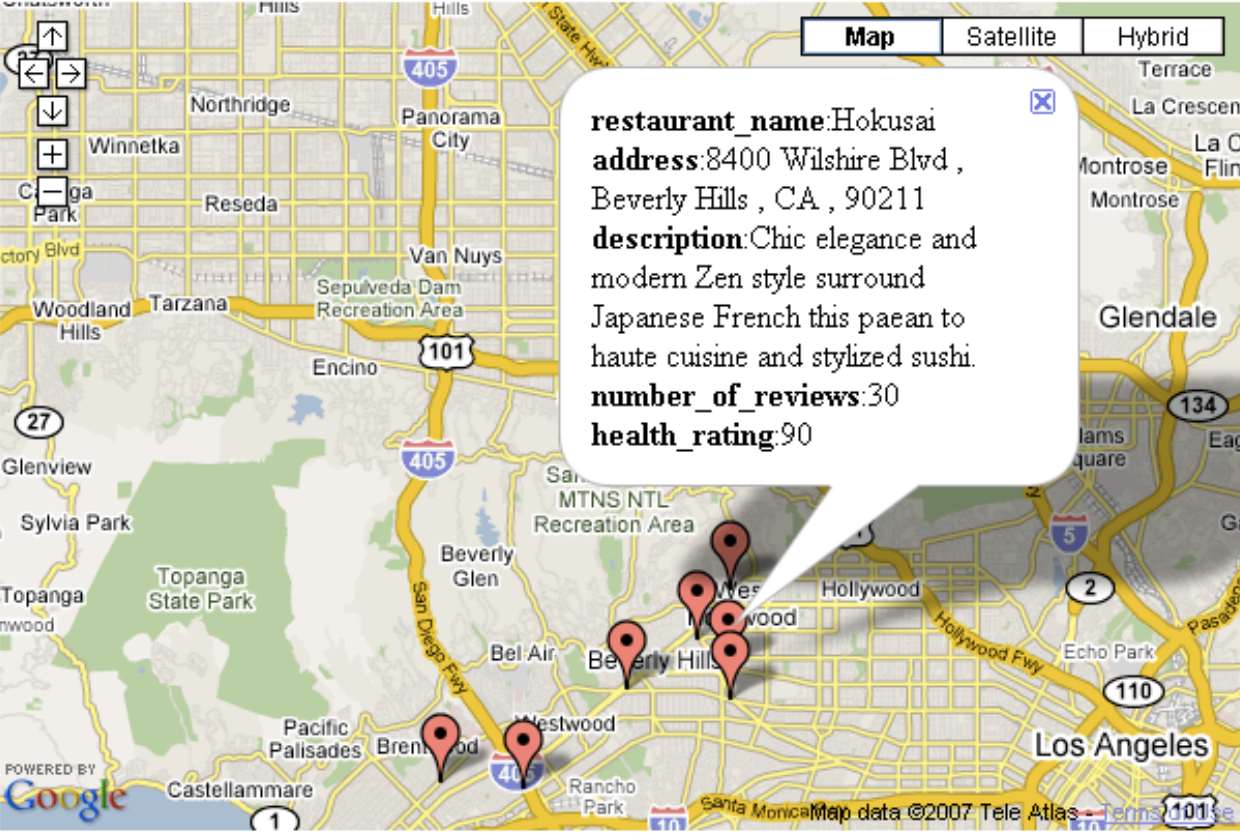
Map Generation

Back Forward Refresh Stop URL: losangeles.citysearch.com/bestof/winners/sushi?init_search=1 GO

restaurant_name: address:

description:

number_of_reviews: health_rating: Go!



The map shows a section of Los Angeles, California, with several red location pins. A popup window is open over one of the pins in the Beverly Hills area. The popup contains the following information:

- restaurant_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

The map interface includes navigation controls (back, forward, refresh, stop) and a search bar. The popup also has a close button (X). The map data is attributed to Tele Atlas and is dated 2007.

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

Preliminary Evaluation

	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Total Steps
Task1 K	3	7	6	0	16
Task1 DP	8	10	21	9	48
Task2 K	9	10	0	0	19
Task2 DP	18	30	0	28	76
Task3 K	5	10	4	5	24
Task3 DP	8	11	16	12	47

- 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