Building Mashups by Example

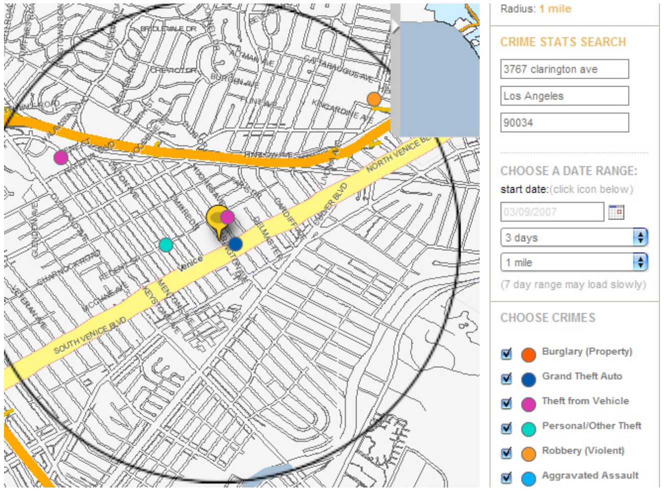
Rattapoom Tuchinda

Doctoral Defense

July 22, 2008

What’s a Mashup?

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



Radius: 1 mile

CRIME STATS SEARCH

3767 clarington ave

Los Angeles

90034

CHOOSE A DATE RANGE:

start date:(click icon below)

03/09/2007

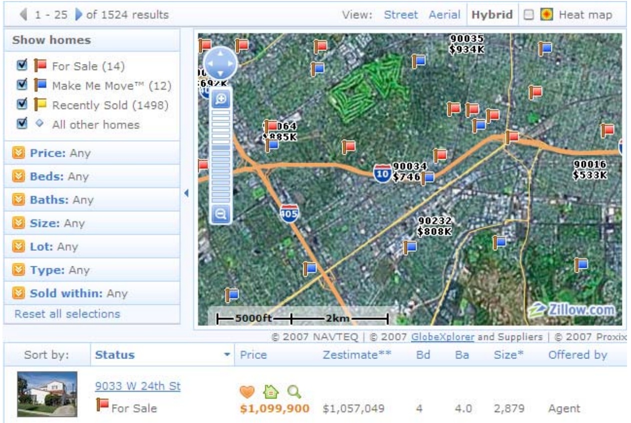
3 days

1 mile

(7 day range may load slowly)

CHOOSE CRIMES

☑ ● Burglary (Property)
☑ ● Grand Theft Auto
☑ ● Theft from Vehicle
☑ ● Personal/Other Theft
☑ ● Robbery (Violent)
☑ ● Aggravated Assault



1 - 25 of 1524 results

View: Street Aerial Hybrid □ Heat map

Show homes

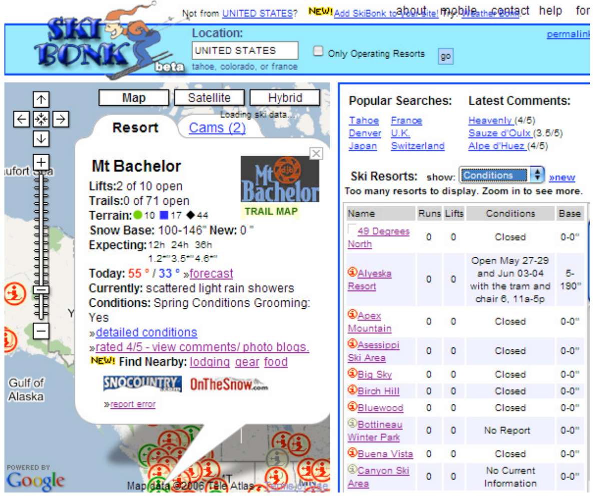
☑ For Sale (14)
☑ Make Me Move™ (12)
☑ Recently Sold (1498)
☑ All other homes

Price: Any
Beds: Any
Baths: Any
Size: Any
Lot: Any
Type: Any
Sold within: Any

Reset all selections

© 2007 NAVTEQ | © 2007 GlobeXplorer and Suppliers | © 2007 Proxix

Sort by:	Status	Price	Zestimate**	Bd	Ba	Size*	Offered by
	9033 W 24th St For Sale	$1,099,900	$1,057,049	4	4.0	2,879	Agent



Not from UNITED STATES? NEW! Add SkiBonk to your site! Try: Weather Bonk. about mobile contact help for

SKI BONK beta

Location: UNITED STATES

tahoe, colorado, or france

□ Only Operating Resorts go

permalink

Mt Bachelor

Lifts:2 of 10 open
Trails:0 of 71 open
Terrain: ● 10 ■ 17 ◆ 44
Snow Base: 100-146" New: 0 "
Expecting: 12h 24h 36h
1.2"" 3.5"" 4.6""
Today: 55 ° / 33 ° »forecast
Currently: scattered light rain showers
Conditions: Spring Conditions Grooming: Yes
»detailed conditions
»rated 4/5 - view comments/ photo blogs.
NEW! Find Nearby: lodging gear food
SNOCOUNTRY OnTheSnow.com
»report error

Popular Searches: Tahoe France Denver U.K. Japan Switzerland

Latest Comments: Heavenly (4/5) Sauze d'Oulx (3.5/5) Alpe d'Huez (4/5)

Ski Resorts: show: Conditions »new

Too many resorts to display. Zoom in to see more.

Name	Runs	Lifts	Conditions	Base
49 Degrees North	0	0	Closed	0-0"
Alyeska Resort	0	0	Open May 27-29 and Jun 03-04 with the tram and chair 6, 11a-5p	5-190"
Apex Mountain	0	0	Closed	0-0"
Asessippi Ski Area	0	0	Closed	0-0"
Big Sky	0	0	Closed	0-0"
Birch Hill	0	0	Closed	0-0"
Bluewood	0	0	Closed	0-0"
Bottineau Winter Park	0	0	No Report	0-0"
Buena Vista	0	0	Closed	0-0"
Canyon Ski Area	0	0	No Current Information	0-0"

a) LA crime map

-Crime Report from different counties

-Map

b) zillow.com

-Real Estate Listing

-Property Tax

c) Ski bonk

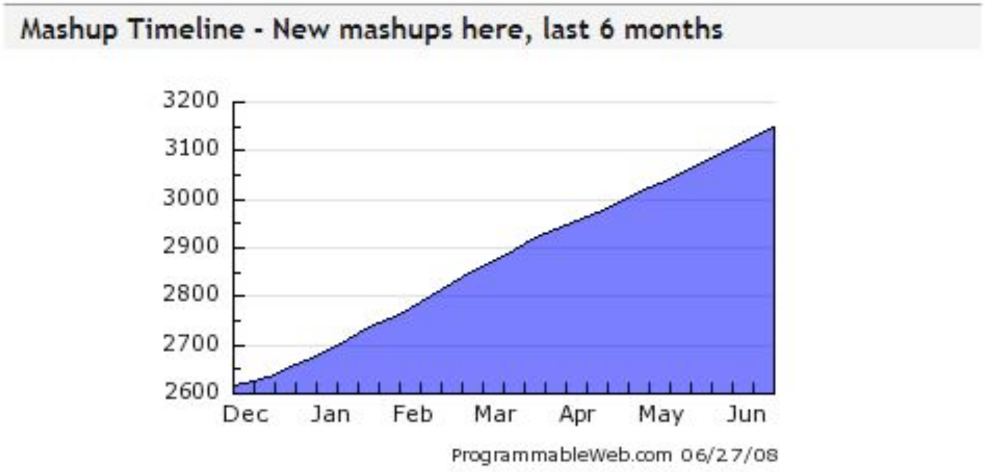
-Weather

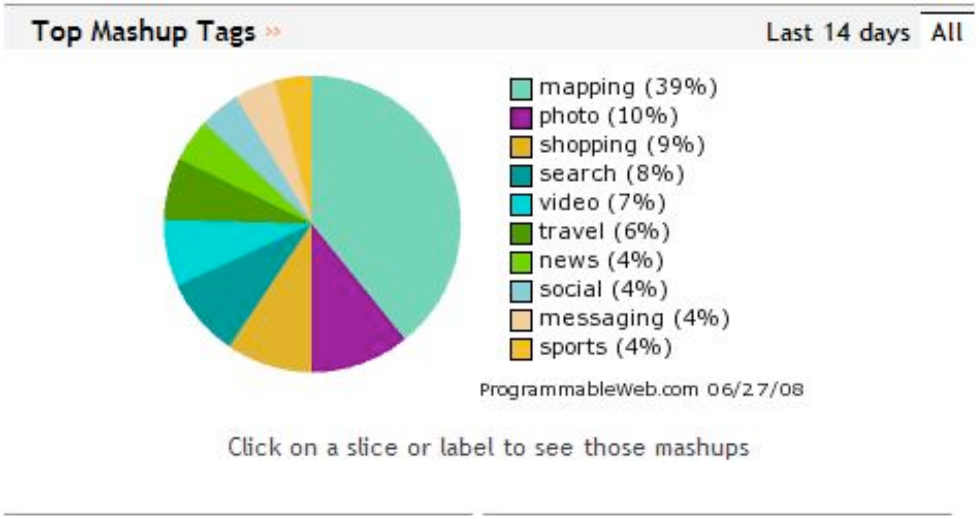
-Snow Report

-Snow Resorts

Combined Data gives new insight / provides new services

Statistics and Trends





Survey of top 50 Mashups

- Divide into five categories based on programming structures
- Focus of this thesis is on the first four categories which account for 47% of the most popular Mashups

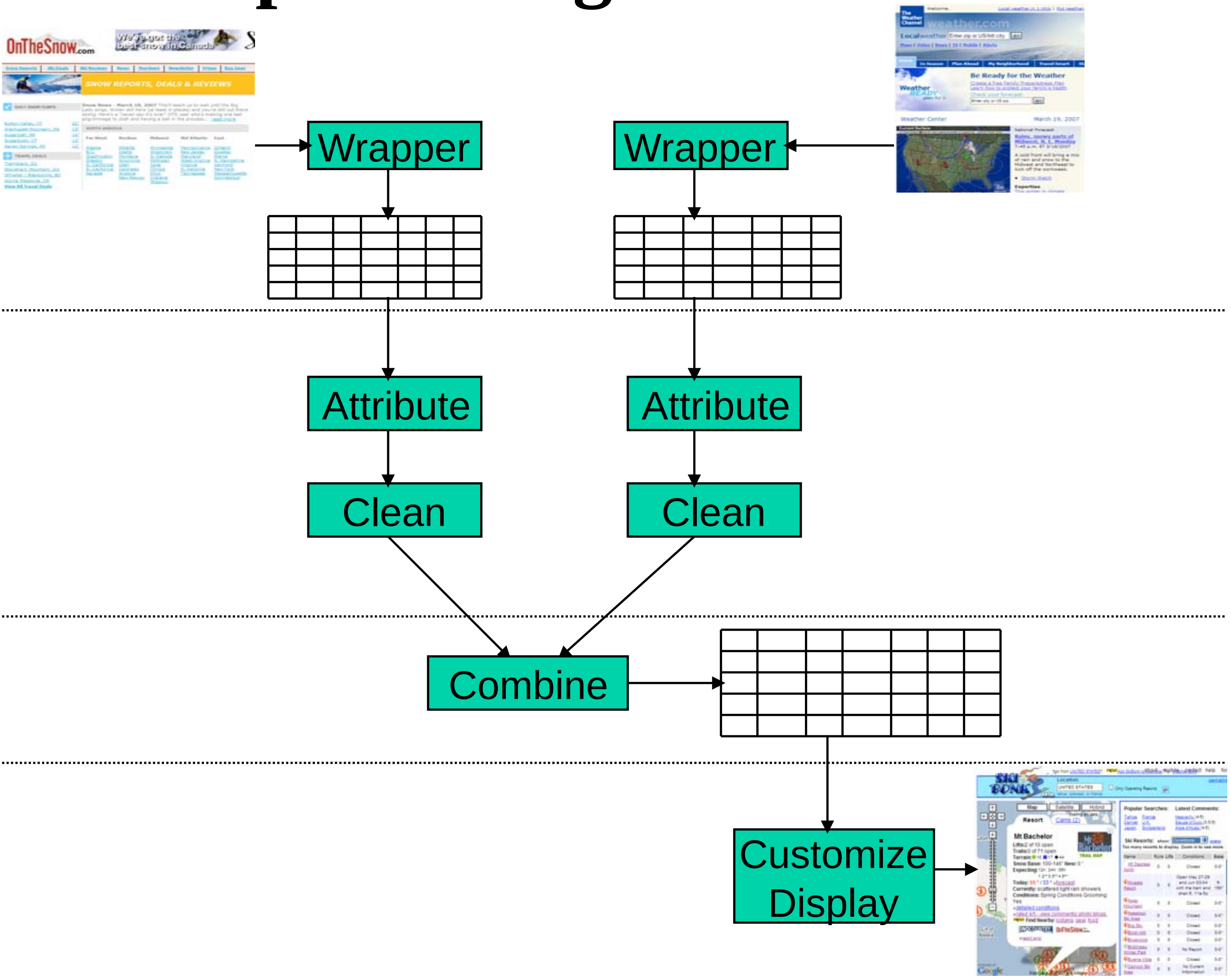
Mashup Building Issues

Data Retrieval

Calibration
-source modeling
-cleaning

Integration

Display



Wrapper

Wrapper

Attribute

Attribute

Clean

Clean

Combine

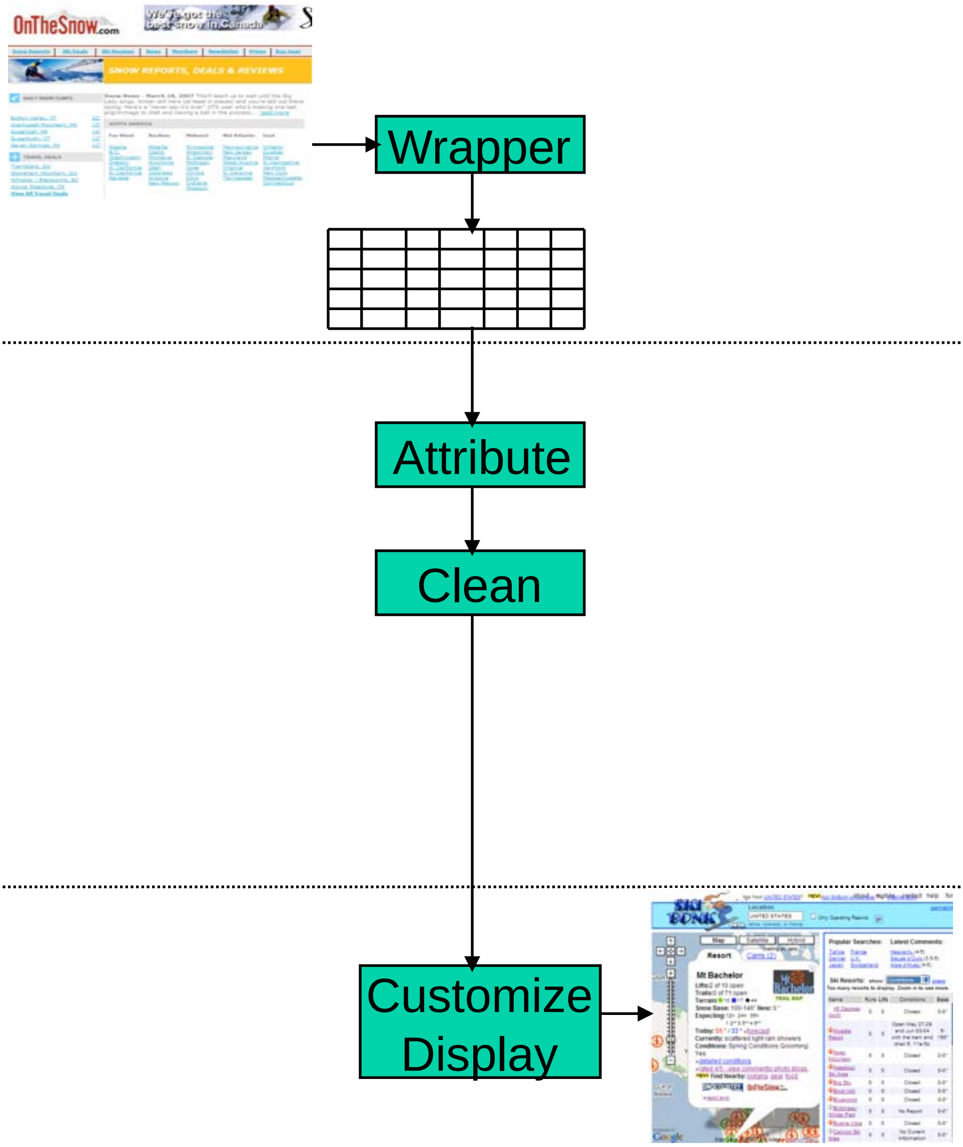
Customize Display

Type 1: One Simple Source

Data Retrieval

Calibration
-source modeling
-cleaning

Display



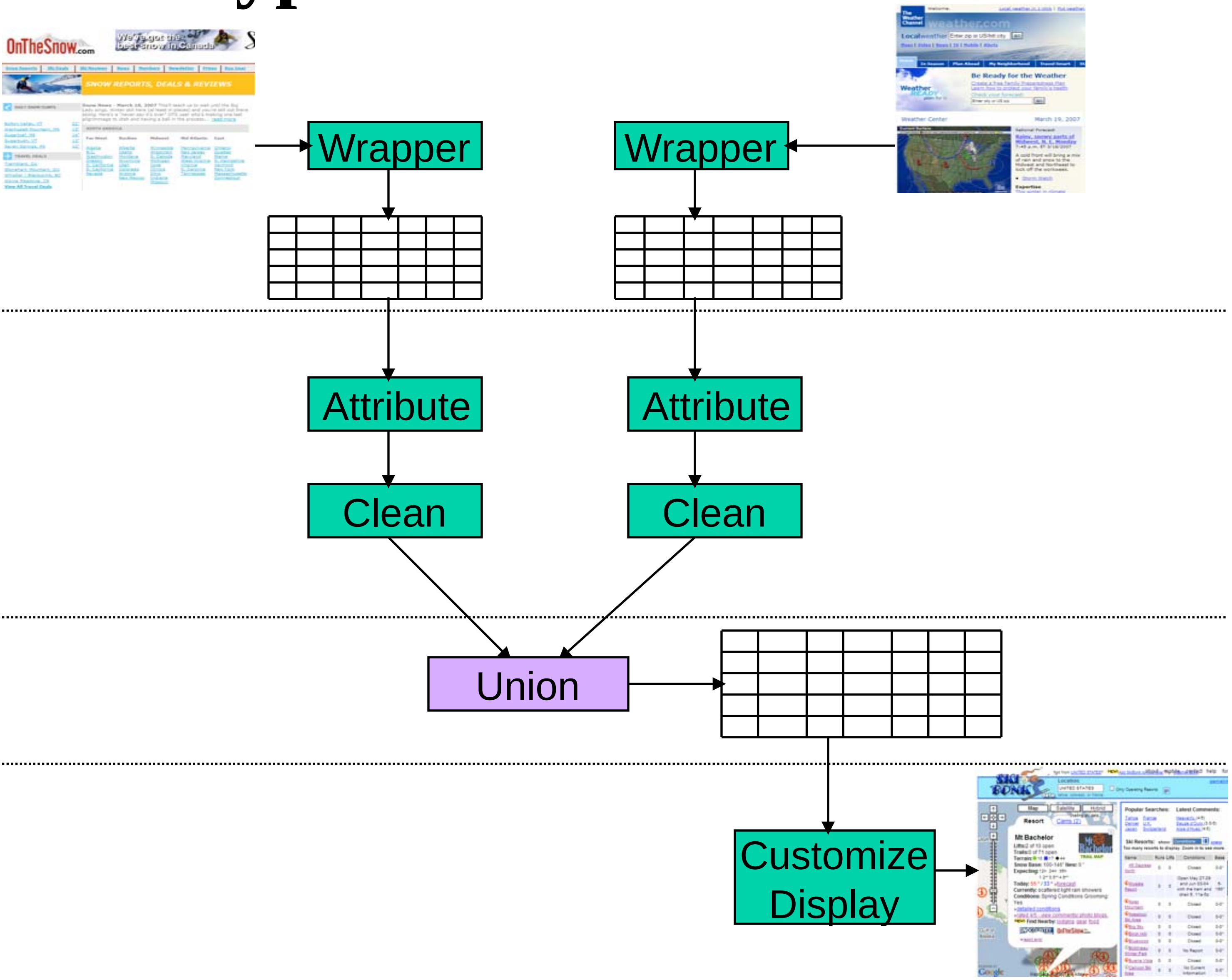
Type 2: Union

Data Retrieval

Calibration
-source modeling
-cleaning

Integration

Display



Wrapper

Wrapper

Attribute

Attribute

Clean

Clean

Union

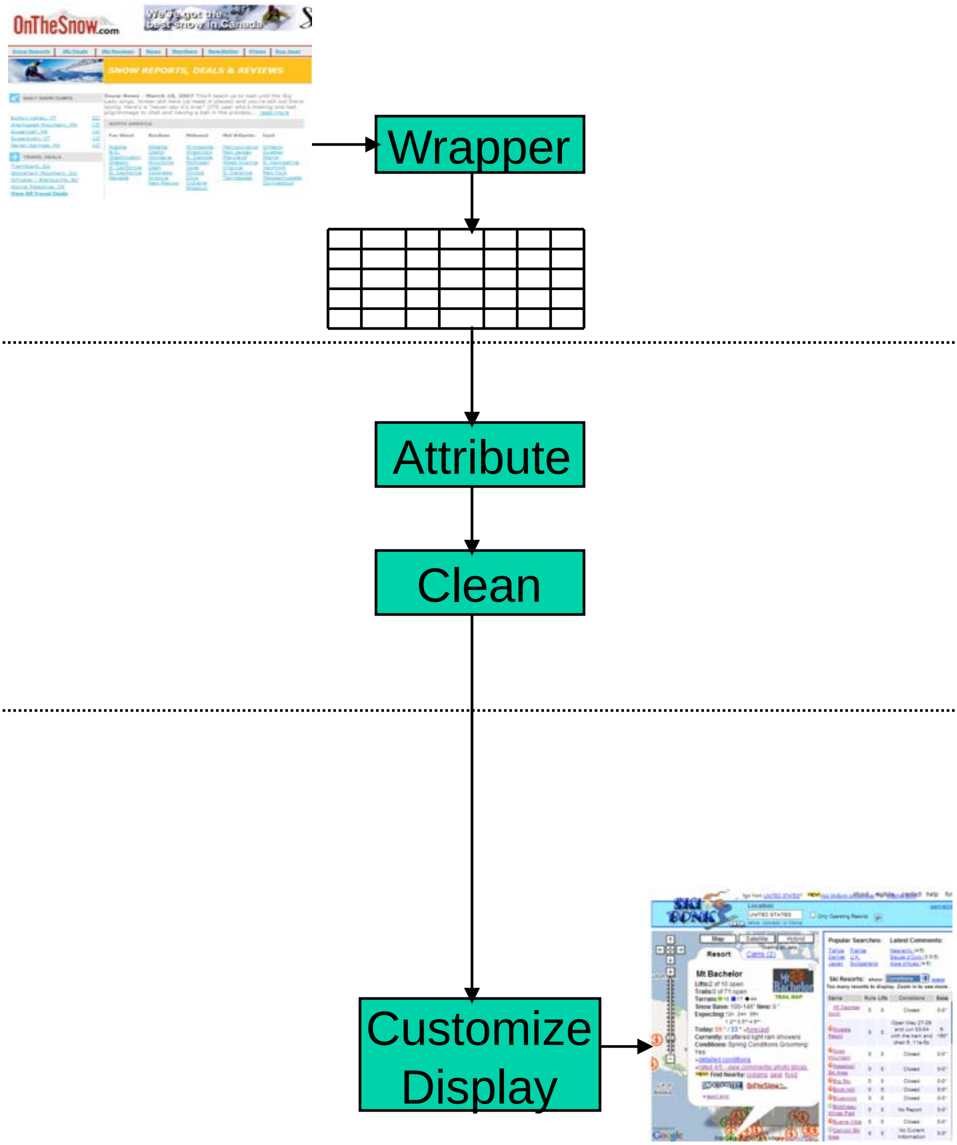
Customize Display

Type 3: One Source with Form

Data Retrieval

Calibration
-source modeling
-cleaning

Display


OnTheSnow.com
We've got the best snow in Canada
SNOW REPORTS, DEALS & REVIEWS
SKI BONK
Mt Bachelor

Wrapper

Attribute

Clean

Customize Display

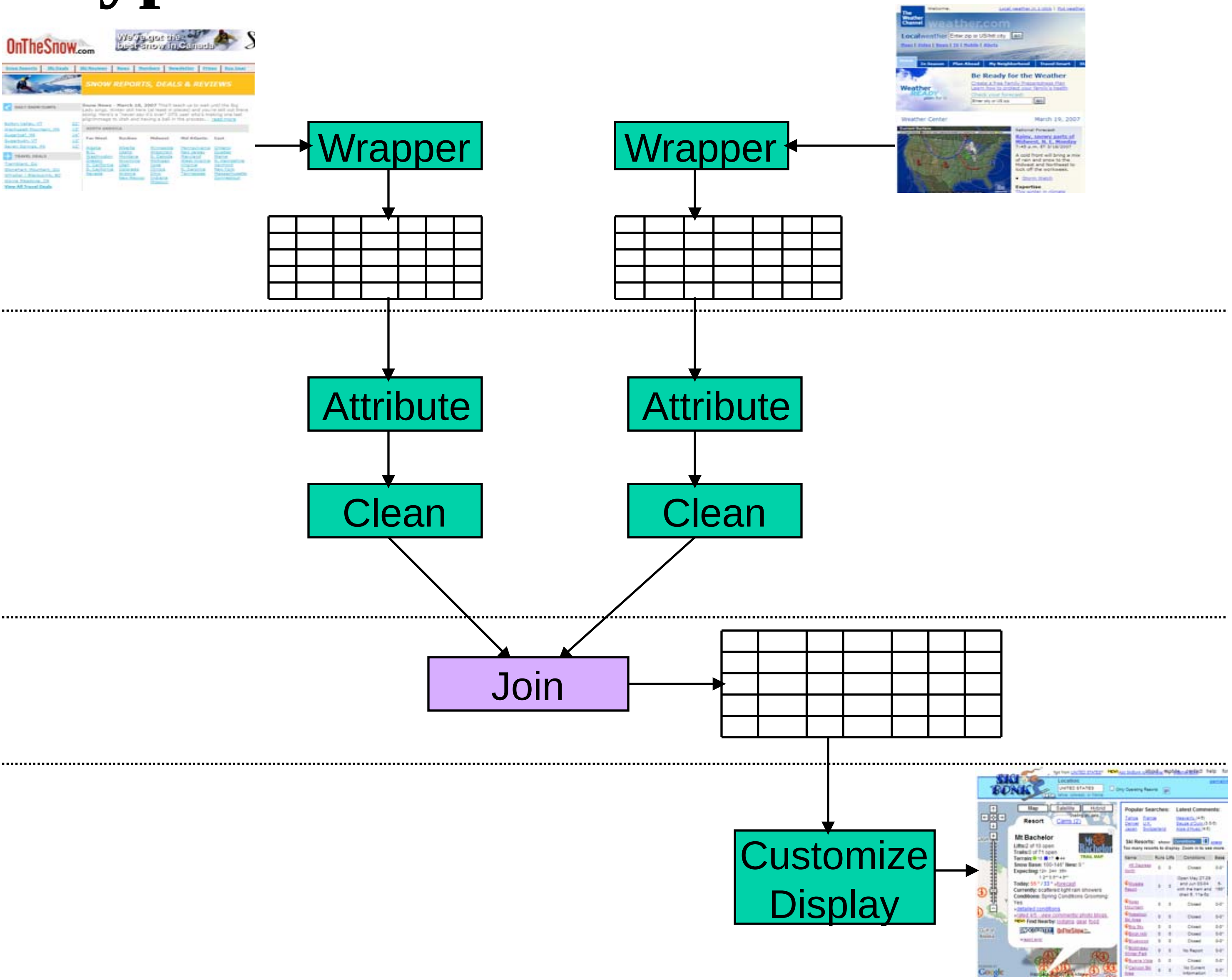
Type 4: Database Join

Data Retrieval

Calibration
-source modeling
-cleaning

Integration

Display



Wrapper

Wrapper

Attribute

Attribute

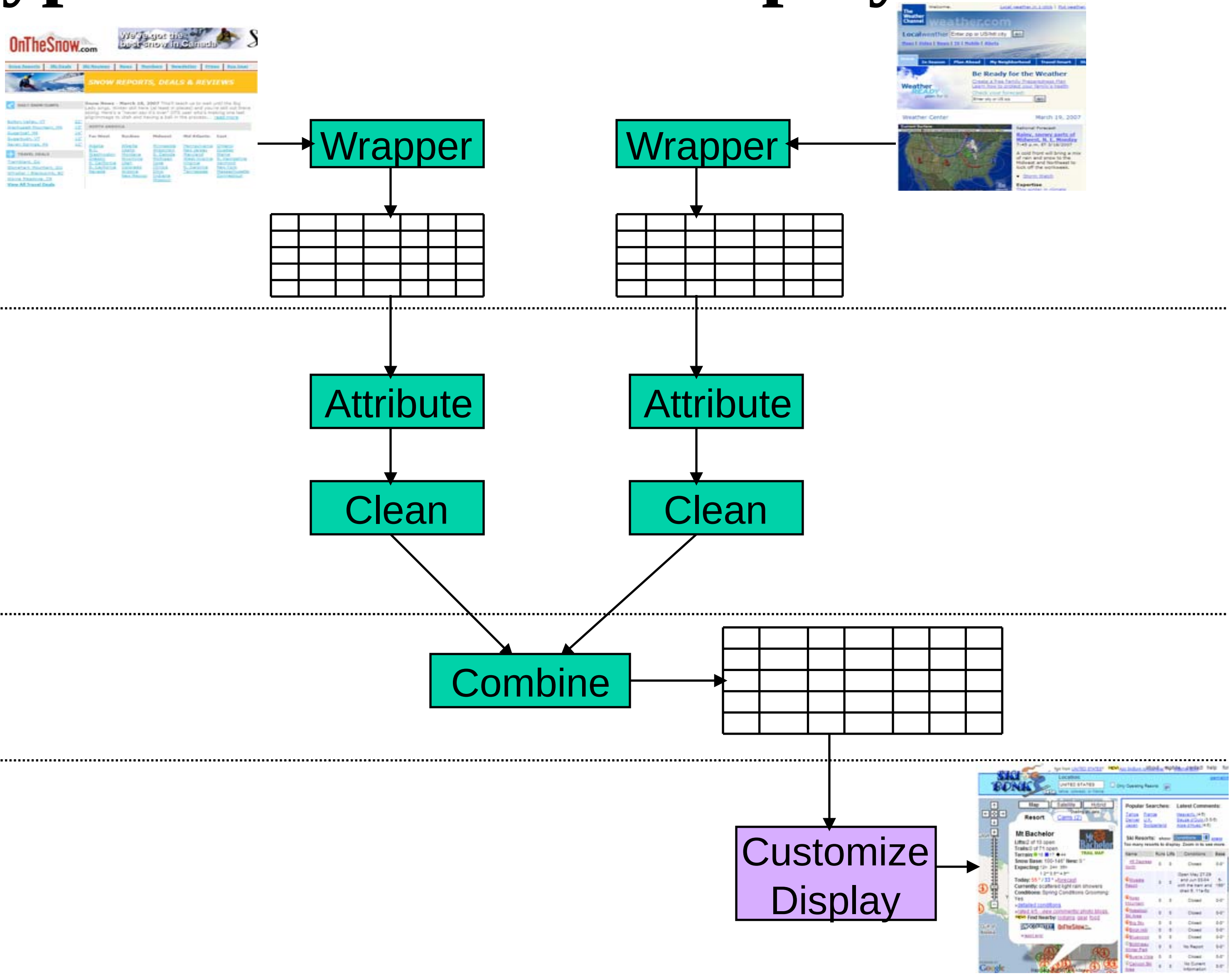
Clean

Clean

Join

Customize Display

Type 5: Customized Display



Data Retrieval

Wrapper

Wrapper

Calibration
-source modeling
-cleaning

Attribute

Attribute

Clean

Clean

Integration

Combine

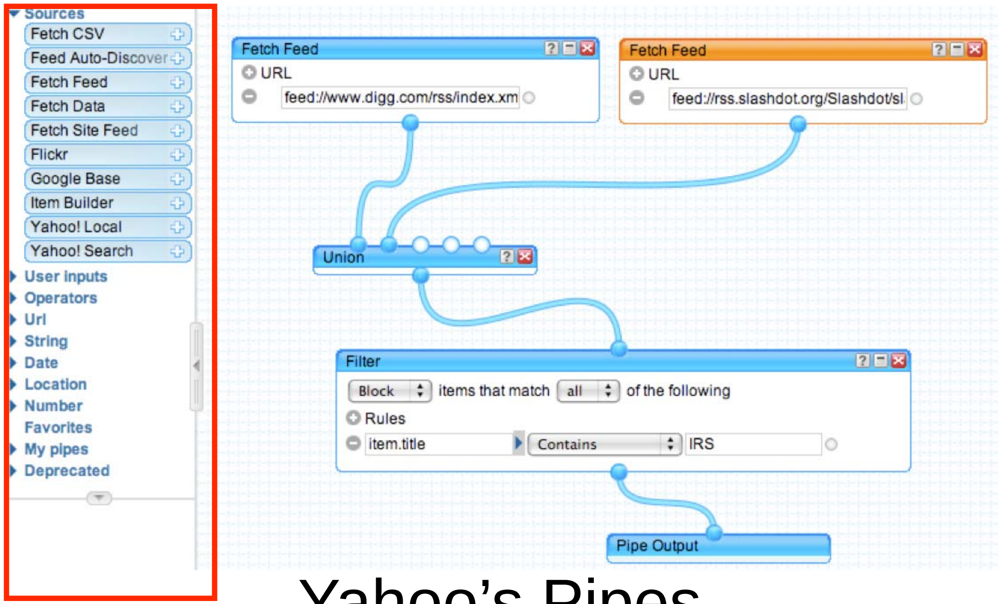
Display

Customize Display

Existing Approaches

Goal: Create Mashups without Programming

- Doesn't translate to not having to understand programming.



Fetch Feed — URL: feed://www.digg.com/rss/index.xm

Fetch Feed — URL: feed://rss.slashdot.org/Slashdot/sl

Filter: Block items that match all of the following — Rules: item.title Contains IRS

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?

Thesis Statement

Web users can build Mashups effectively using an integrated framework that lets them solve the problems of data extraction, source modeling, data cleaning, and data integration by specifying examples instead of programming operations.

Contributions

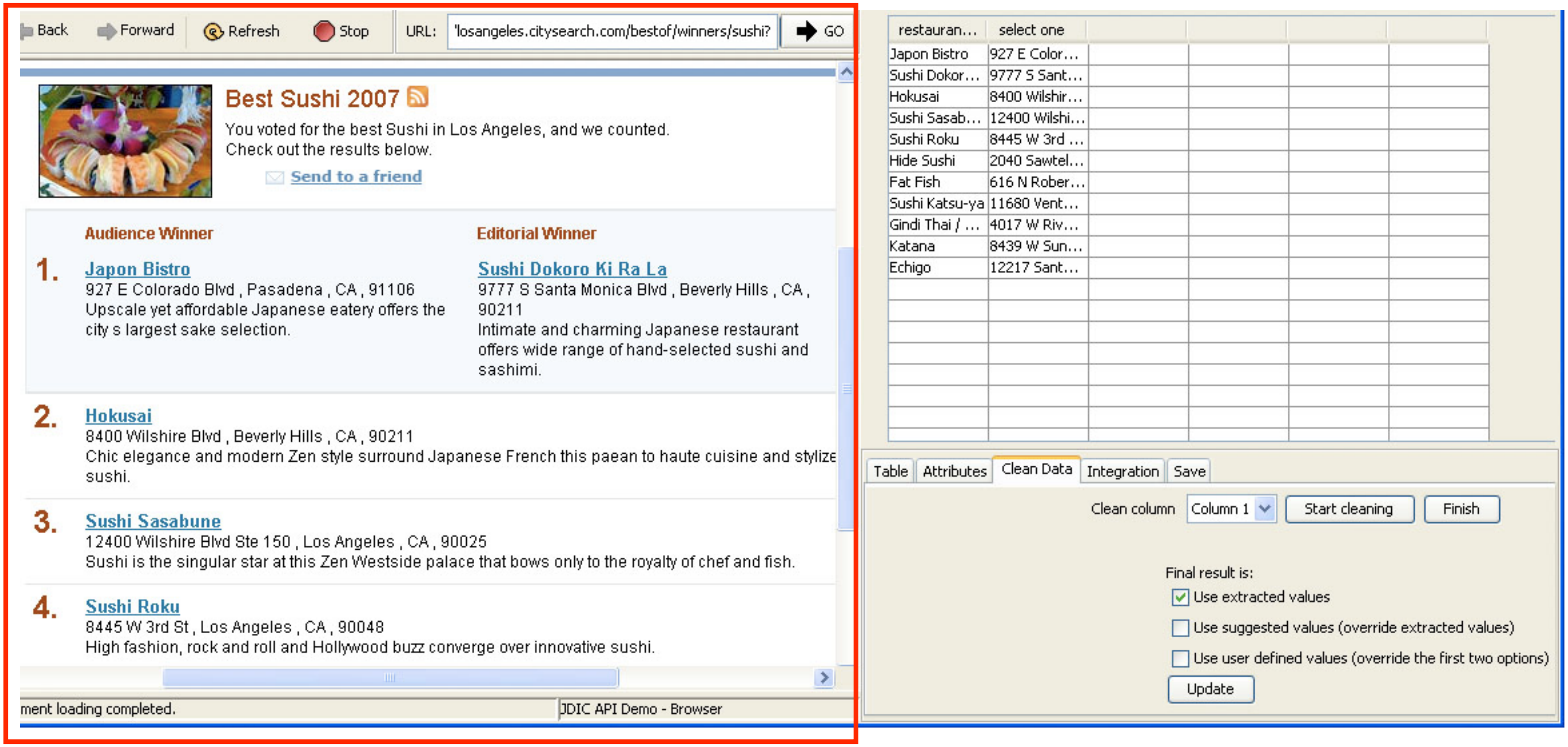
- A programming by demonstration approach that uses a single table for building a Mashup
- An integrated approach that links data extraction, source modeling, data cleaning, and data integration together.
- A query formulation technique that allows users to specify examples to build complicated queries.

Key Ideas

- Focus on data, not operations
  - Users are more familiar with data.
- Leverage existing database
  - Help source modeling, cleaning, and data integration.
- Consolidate as opposed to Divide-And-Conquer
  - Solving a problem in one issue can help solve another issue.
  - Interacting within a single spreadsheet platform

Our system: Karma

Embedded Browser



URL: 'losangeles.citysearch.com/bestof/winners/sushi?

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

Send to a friend

Audience Winner

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

Editorial Winner

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.

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

ment loading completed. JDIC API Demo - Browser

restauran…	select one
Japon Bistro	927 E Color…
Sushi Dokor…	9777 S Sant…
Hokusai	8400 Wilshir…
Sushi Sasab…	12400 Wilshi…
Sushi Roku	8445 W 3rd …
Hide Sushi	2040 Sawtel…
Fat Fish	616 N Rober…
Sushi Katsu-ya	11680 Vent…
Gindi Thai / …	4017 W Riv…
Katana	8439 W Sun…
Echigo	12217 Sant…

Clean column: Column 1

Final result is:

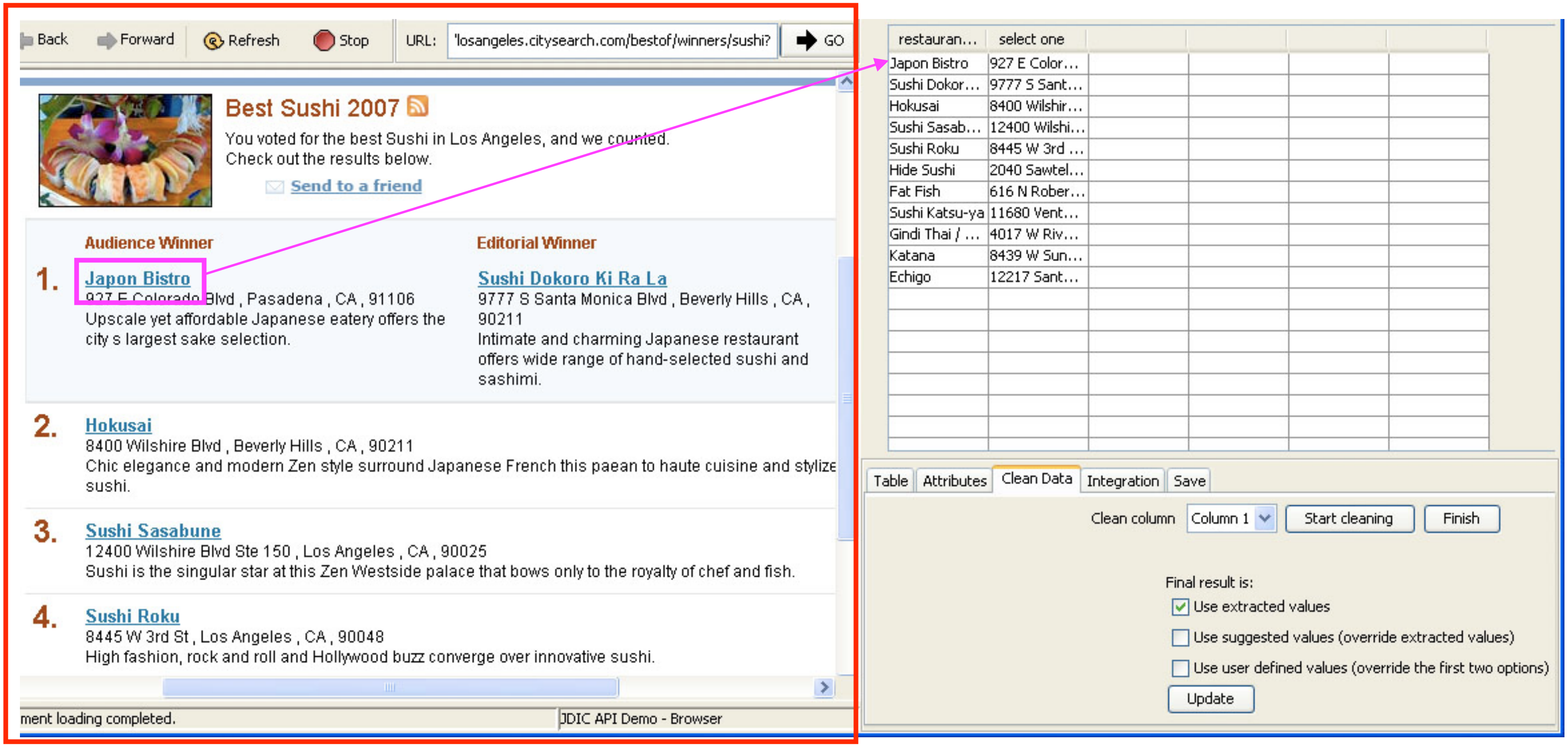
☑ Use extracted values

□ Use suggested values (override extracted values)

□ Use user defined values (override the first two options)

Our system: Karma

Embedded Browser



URL: 'losangeles.citysearch.com/bestof/winners/sushi?

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

Send to a friend

Audience Winner

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

Editorial Winner

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.

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

ment loading completed. JDIC API Demo - Browser

restauran…	select one
Japon Bistro	927 E Color…
Sushi Dokor…	9777 S Sant…
Hokusai	8400 Wilshir…
Sushi Sasab…	12400 Wilshi…
Sushi Roku	8445 W 3rd …
Hide Sushi	2040 Sawtel…
Fat Fish	616 N Rober…
Sushi Katsu-ya	11680 Vent…
Gindi Thai / …	4017 W Riv…
Katana	8439 W Sun…
Echigo	12217 Sant…

Clean column: Column 1

Final result is:

☑ Use extracted values

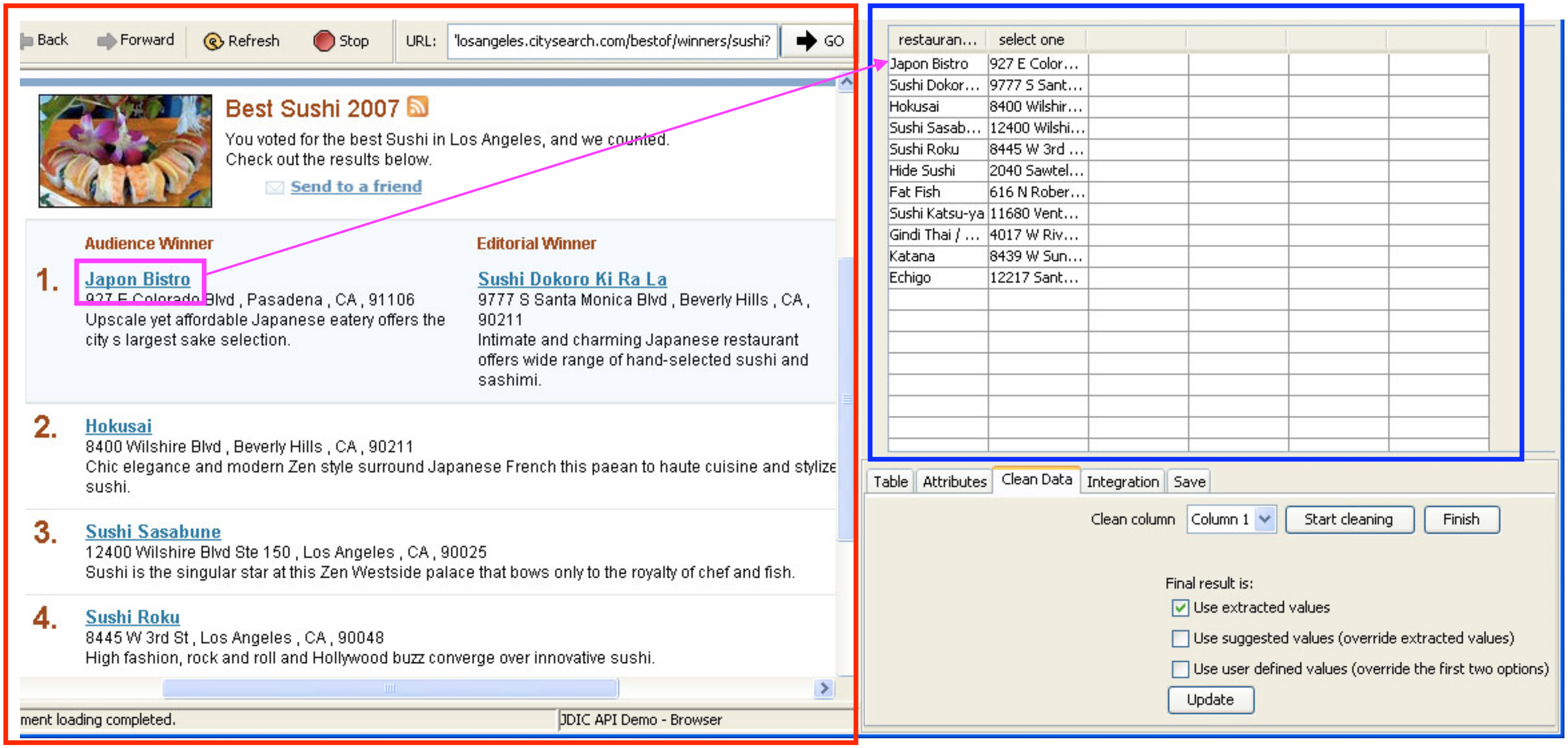
☐ Use suggested values (override extracted values)

☐ Use user defined values (override the first two options)

Our system: Karma

Embedded Browser

Table



'losangeles.citysearch.com/bestof/winners/sushi?

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

Send to a friend

Audience Winner

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

Editorial Winner

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.

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

ment loading completed.

JDIC API Demo - Browser

restauran...	select one
Japon Bistro	927 E Color...
Sushi Dokor...	9777 S Sant...
Hokusai	8400 Wilshir...
Sushi Sasab...	12400 Wilshi...
Sushi Roku	8445 W 3rd ...
Hide Sushi	2040 Sawtel...
Fat Fish	616 N Rober...
Sushi Katsu-ya	11680 Vent...
Gindi Thai / ...	4017 W Riv...
Katana	8439 W Sun...
Echigo	12217 Sant...

Clean column: Column 1

Final result is:

☑ Use extracted values

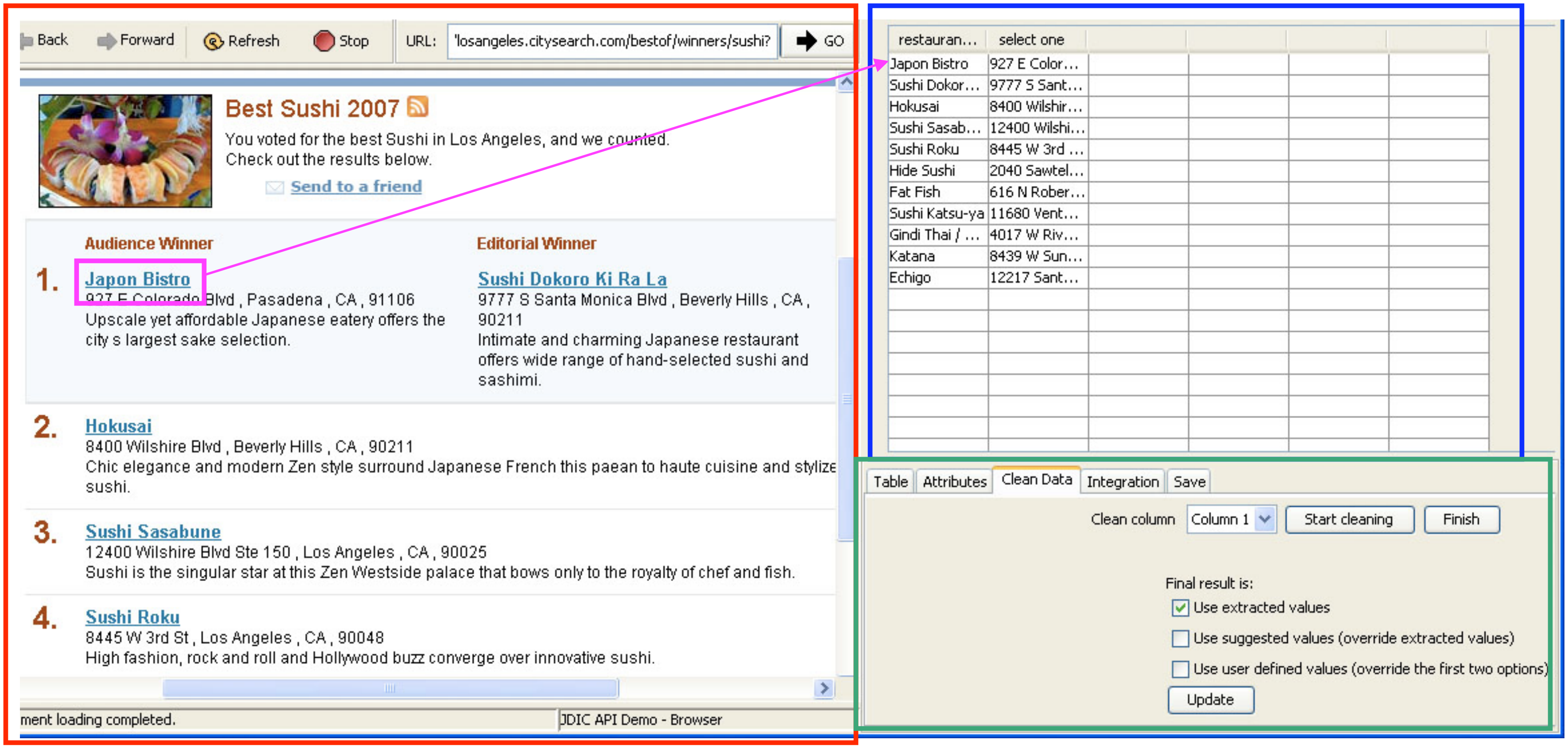
□ Use suggested values (override extracted values)

□ Use user defined values (override the first two options)

Our system: Karma

Embedded Browser

Table



URL: 'losangeles.citysearch.com/bestof/winners/sushi?

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

Send to a friend

Audience Winner

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

Editorial Winner

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.

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

ment loading completed. JDIC API Demo - Browser

restauran...	select one
Japon Bistro	927 E Color...
Sushi Dokor...	9777 S Sant...
Hokusai	8400 Wilshir...
Sushi Sasab...	12400 Wilshi...
Sushi Roku	8445 W 3rd ...
Hide Sushi	2040 Sawtel...
Fat Fish	616 N Rober...
Sushi Katsu-ya	11680 Vent...
Gindi Thai / ...	4017 W Riv...
Katana	8439 W Sun...
Echigo	12217 Sant...

Clean column: Column 1

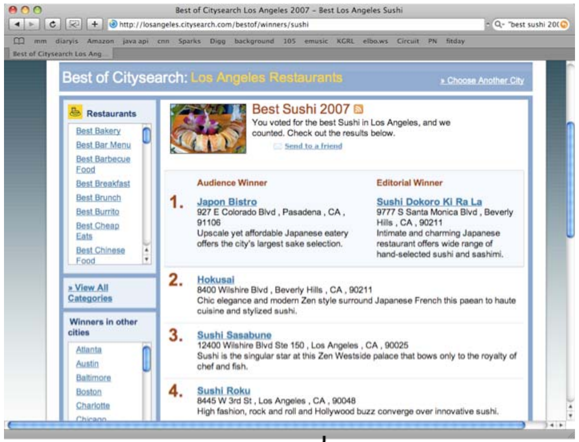
Final result is:

☑ Use extracted values

☐ Use suggested values (override extracted values)

☐ Use user defined values (override the first two options)

Interaction Modes



Best of Citysearch Los Angeles 2007 – Best Los Angeles Sushi

http://losangeles.citysearch.com/bestof/winners/sushi

Best of Citysearch: Los Angeles Restaurants

Best Sushi 2007

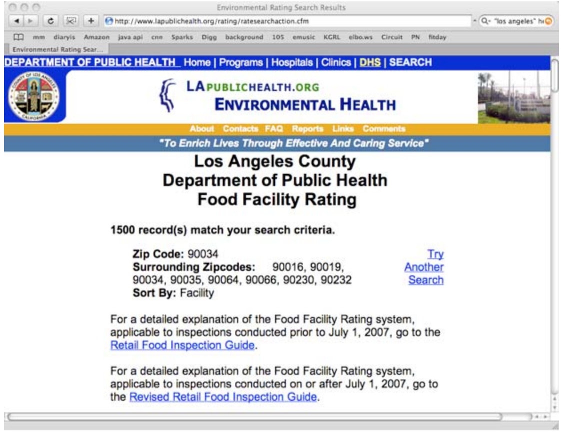
You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

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

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.


mm diaryis Amazon java api cnn Sparks Digg background 105 emusic KGRL elbo.ws Circuit PN fitday
"los angeles" h
Environmental Rating Sear...
DEPARTMENT OF PUBLIC HEALTH Home | Programs | Hospitals | Clinics | DHS | SEARCH
LAPUBLICHEALTH.ORG
ENVIRONMENTAL HEALTH
About Contacts FAQ Reports Links Comments
Try Another Search

Environmental Rating Search Results

http://www.lapublichealth.org/rating/ratesearchaction.cfm

"To Enrich Lives Through Effective And Caring Service"

Los Angeles County
Department of Public Health
Food Facility Rating

1500 record(s) match your search criteria.

Zip Code: 90034
Surrounding Zipcodes: 90016, 90019, 90034, 90035, 90064, 90066, 90230, 90232
Sort By: Facility

For a detailed explanation of the Food Facility Rating system, applicable to inspections conducted prior to July 1, 2007, go to the Retail Food Inspection Guide.

For a detailed explanation of the Food Facility Rating system, applicable to inspections conducted on or after July 1, 2007, go to the Revised Retail Food Inspection Guide.

Extract

{Restaurant name, address, phone, Review}

Clean

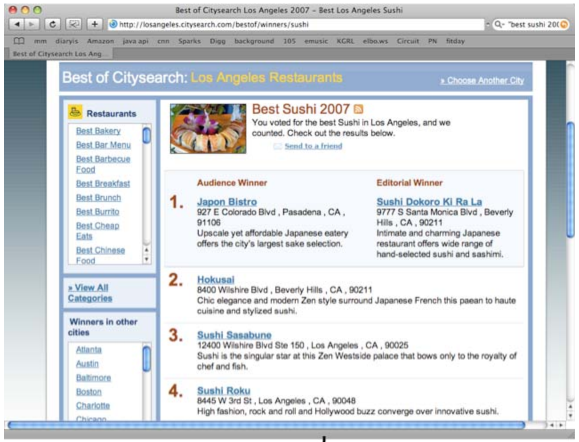
Extract

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

Clean

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

Map


Best of Citysearch Los Angeles 2007 – Best Los Angeles Sushi
http://losangeles.citysearch.com/bestof/winners/sushi
"best sushi 20(
mm diaryis Amazon java api cnn Sparks Digg background 105 emusic KGRL elbo.ws Circuit PN fitday
Best of Citysearch Los Ang...

Best of Citysearch: Los Angeles Restaurants

» Choose Another City

Restaurants

Best Bakery
Best Bar Menu
Best Barbecue Food
Best Breakfast
Best Brunch
Best Burrito
Best Cheap Eats
Best Chinese Food

» View All Categories

Winners in other cities

Atlanta
Austin
Baltimore
Boston
Charlotte
Chicago

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

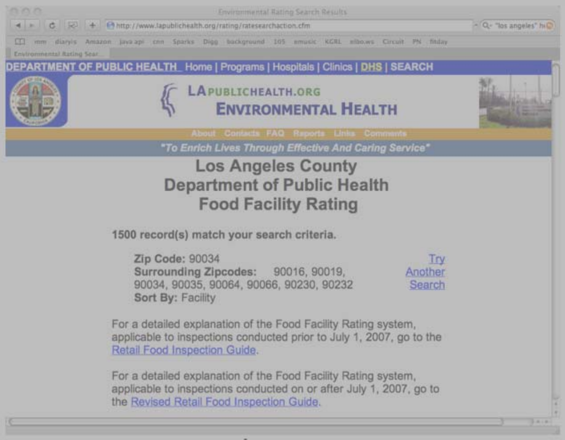
Send to a friend

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

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.


Environmental Rating Search Results
http://www.lapublichealth.org/rating/ratesearchaction.cfm
"los angeles" h
mm diaryis Amazon java api cnn Sparks Digg background 105 emusic KGRL elbo.ws Circuit PN fitday
Environmental Rating Sear...
DEPARTMENT OF PUBLIC HEALTH Home | Programs | Hospitals | Clinics | DHS | SEARCH
LAPUBLICHEALTH.ORG
ENVIRONMENTAL HEALTH
About Contacts FAQ Reports Links Comments
"To Enrich Lives Through Effective And Caring Service"

Los Angeles County
Department of Public Health
Food Facility Rating

1500 record(s) match your search criteria.

Zip Code: 90034
Surrounding Zipcodes: 90016, 90019, 90034, 90035, 90064, 90066, 90230, 90232
Sort By: Facility

Try Another Search

For a detailed explanation of the Food Facility Rating system, applicable to inspections conducted prior to July 1, 2007, go to the Retail Food Inspection Guide.

For a detailed explanation of the Food Facility Rating system, applicable to inspections conducted on or after July 1, 2007, go to the Revised Retail Food Inspection Guide.

Extract

{Restaurant name, address, phone, Review}

Clean

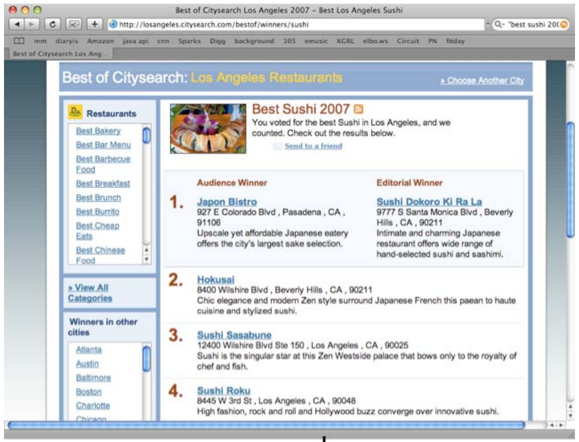
Extract

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

Clean

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

Map


Best of Citysearch Los Angeles 2007 – Best Los Angeles Sushi
http://losangeles.citysearch.com/bestof/winners/sushi
"best sushi 20(
mm diaryis Amazon java api cnn Sparks Digg background 105 emusic KGRL elbo.ws Circuit PN fitday
Best of Citysearch Los Ang…

Best of Citysearch: Los Angeles Restaurants

» Choose Another City

Restaurants

- Best Bakery
- Best Bar Menu
- Best Barbecue Food
- Best Breakfast
- Best Brunch
- Best Burrito
- Best Cheap Eats
- Best Chinese Food

» View All Categories

Winners in other cities

- Atlanta
- Austin
- Baltimore
- Boston
- Charlotte
- Chicago

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

Send to a friend

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

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.

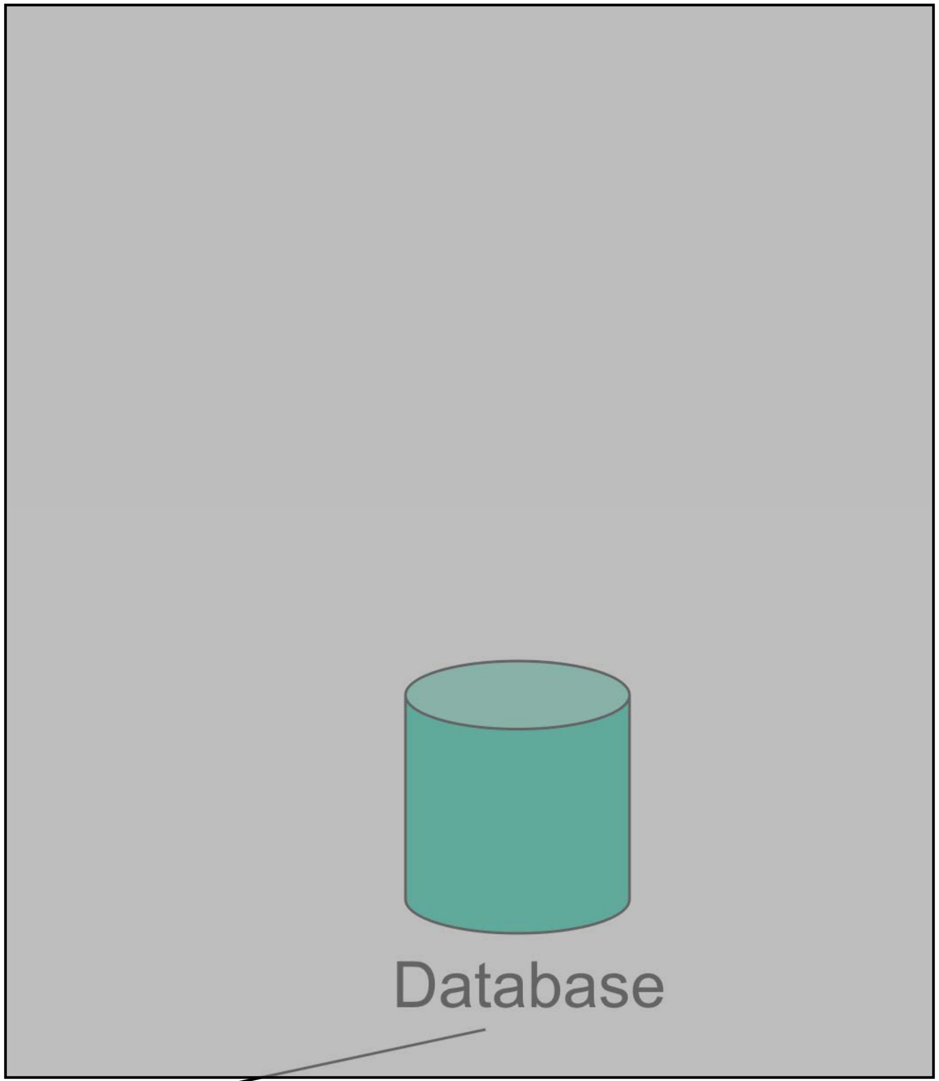
3. Sushi Sasabune
12400 Wilshire Blvd Ste 150 , Los Angeles , CA , 90025
Sushi is the singular star at this Zen Westside palace that bows only to the royalty of chef and fish.

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

Extract

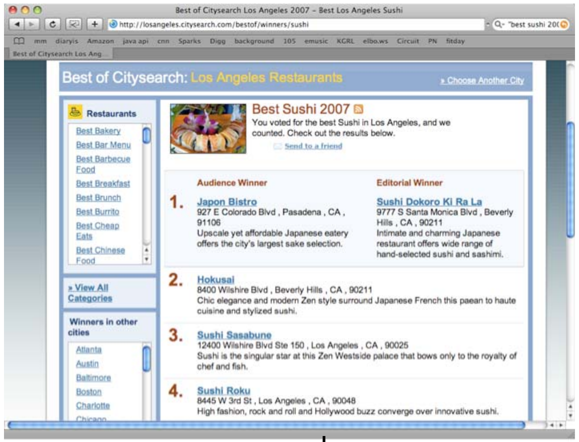
{Restaurant name, address, phone, Review}

Clean



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

Map



Best of Citysearch: Los Angeles Restaurants

Best Sushi 2007

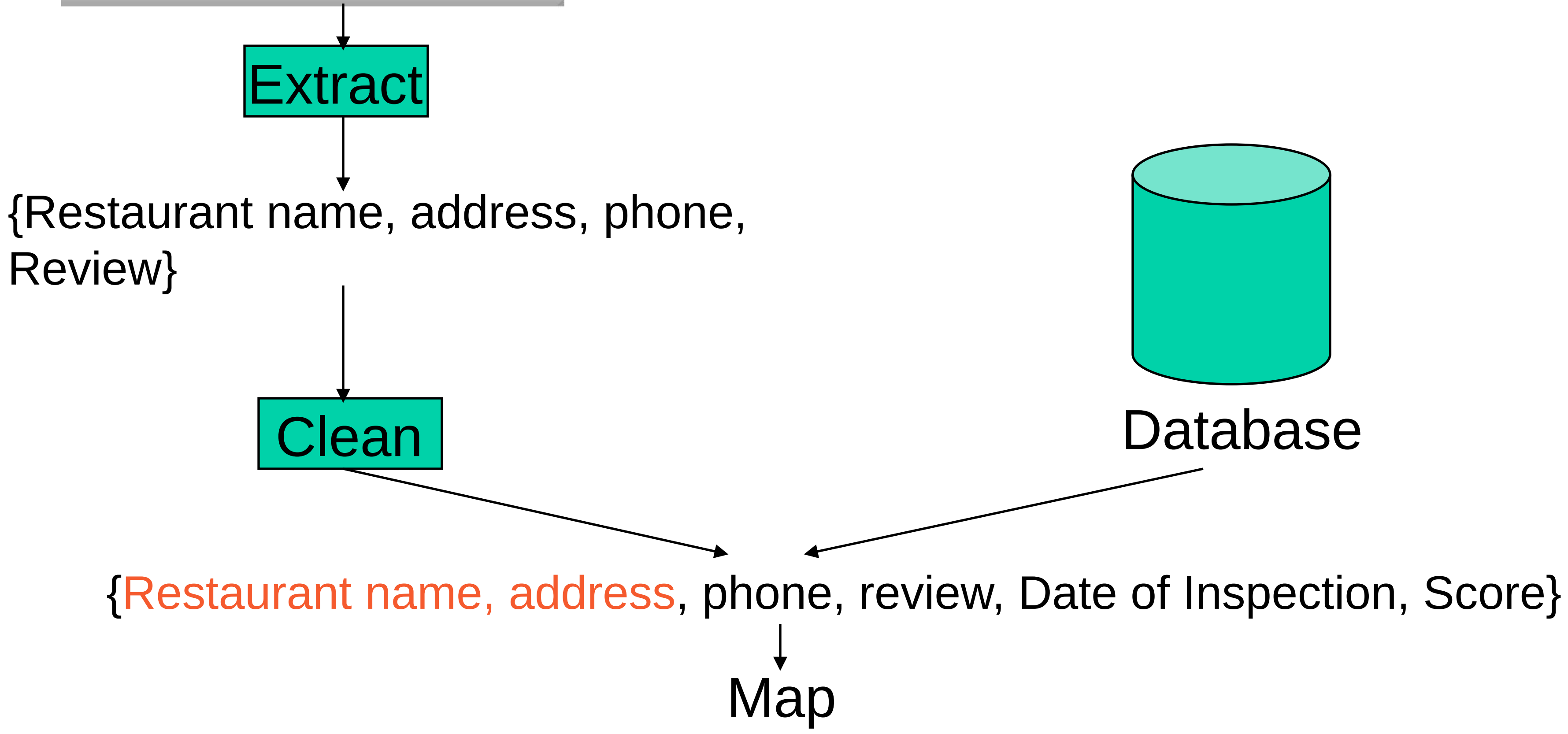
You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

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

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.



Extract

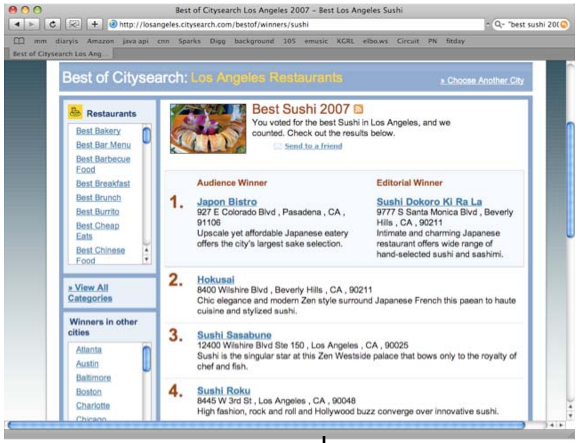
{Restaurant name, address, phone, Review}

Clean

Database

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

Map


Best of Citysearch Los Angeles 2007 – Best Los Angeles Sushi
http://losangeles.citysearch.com/bestof/winners/sushi
"best sushi 20(
mm diaryis Amazon java api cnn Sparks Digg background 105 emusic KGRL elbo.ws Circuit PN fitday
Best of Citysearch Los Ang...

Best of Citysearch: Los Angeles Restaurants

» Choose Another City

Restaurants

Best Bakery
Best Bar Menu
Best Barbecue Food
Best Breakfast
Best Brunch
Best Burrito
Best Cheap Eats
Best Chinese Food

» View All Categories

Winners in other cities

Atlanta
Austin
Baltimore
Boston
Charlotte
Chicago

Best Sushi 2007

You voted for the best Sushi in Los Angeles, and we counted. Check out the results below.

Send to a friend

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

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.

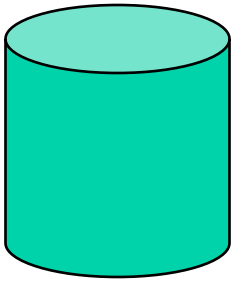
4. Sushi Roku
8445 W 3rd St , Los Angeles , CA , 90048
High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

Extract

{Restaurant name, address, phone, Review}

Clean

Database contains past Mashups and data tables

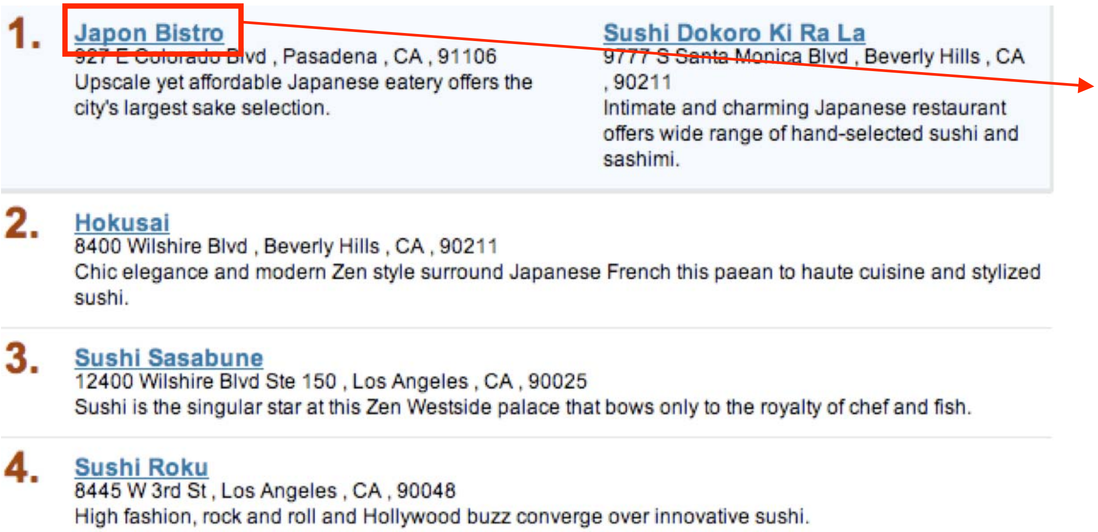


Database

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

Map

Data Retrieval: Extraction



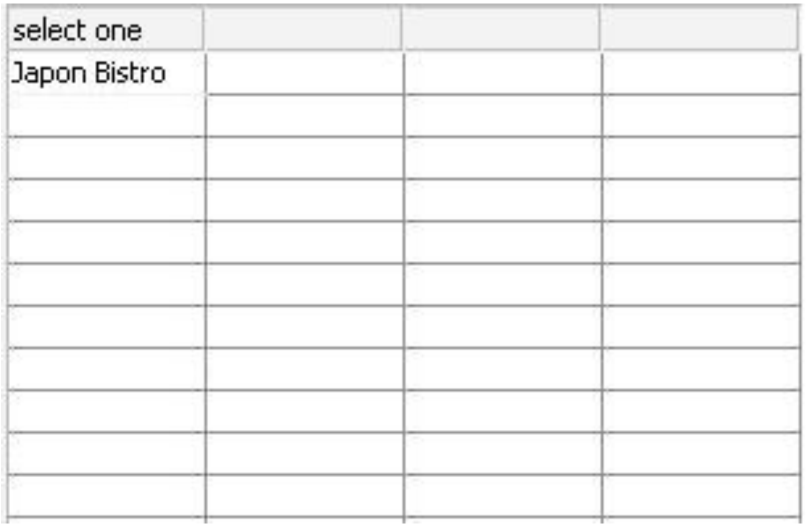
1. Japon Bistro
927 E Colorado Blvd , Pasadena , CA , 91106
Upscale yet affordable Japanese eatery offers the city's largest sake selection.

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.

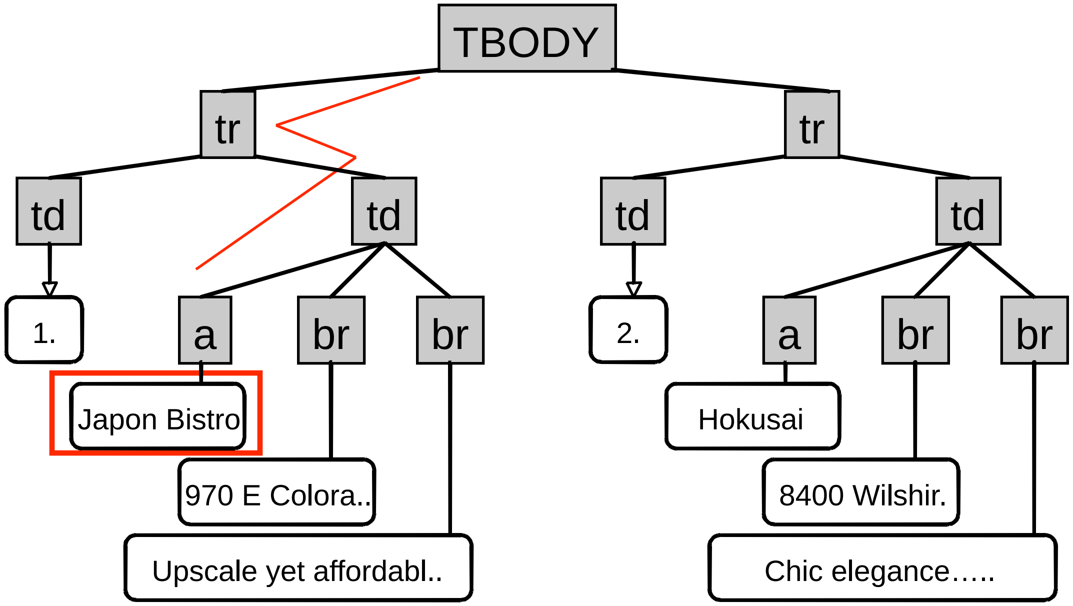
2. Hokusai
8400 Wilshire Blvd , Beverly Hills , CA , 90211
Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.

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.

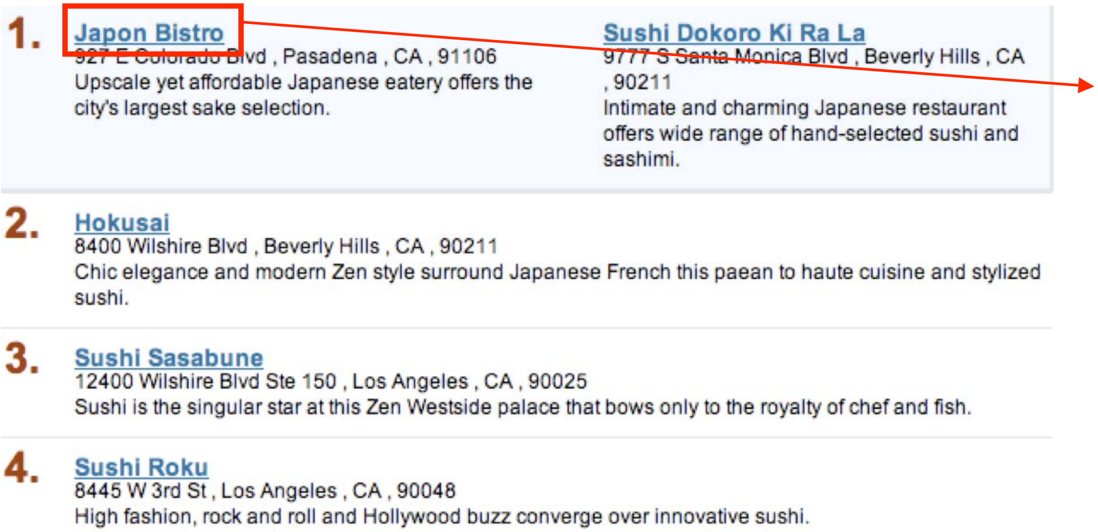


select one			
Japon Bistro			



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

Data Retrieval: Extraction



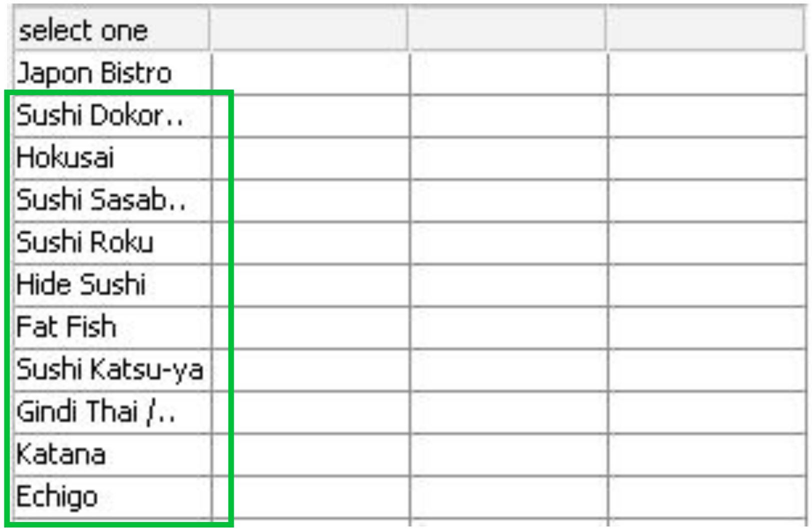
1. Japon Bistro
927 E Colorado Blvd , Pasadena , CA , 91106
Upscale yet affordable Japanese eatery offers the city's largest sake selection.

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.

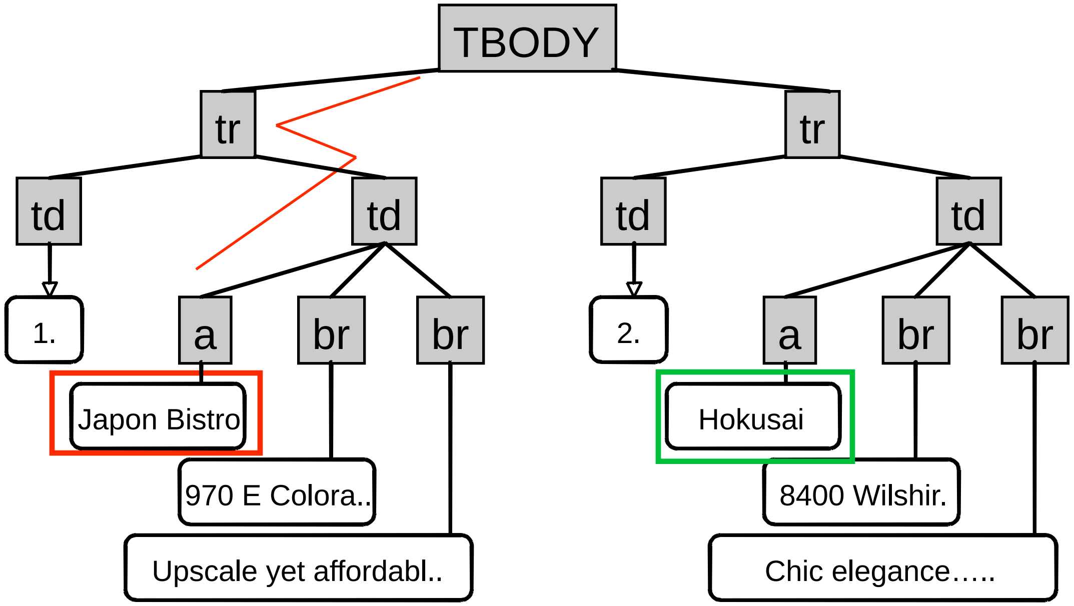
2. Hokusai
8400 Wilshire Blvd , Beverly Hills , CA , 90211
Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.

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.



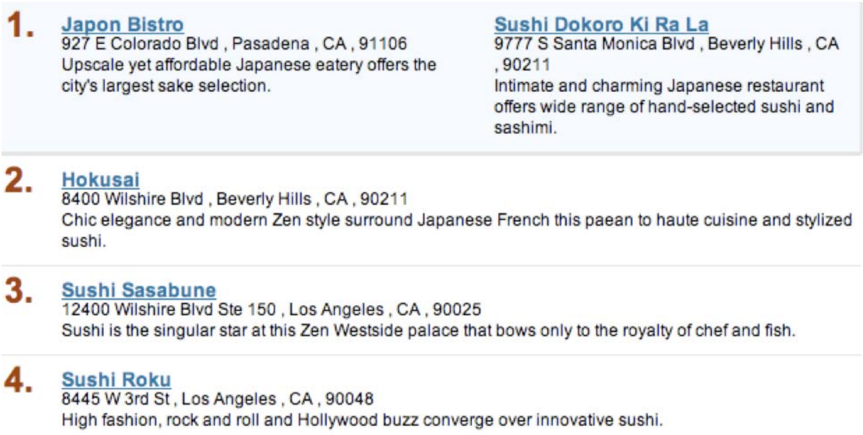
select one			
Japon Bistro			
Sushi Dokor..			
Hokusai			
Sushi Sasab..			
Sushi Roku			
Hide Sushi			
Fat Fish			
Sushi Katsu-ya			
Gindi Thai /..			
Katana			
Echigo			



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

Tbody/tr*/td*/a

Data Retrieval: Navigation



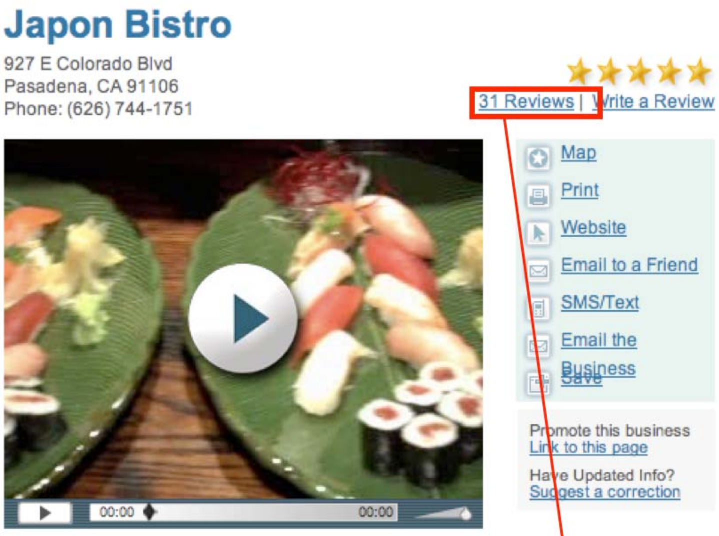
1. Japon Bistro
927 E Colorado Blvd , Pasadena , CA , 91106
Upscale yet affordable Japanese eatery offers the city's largest sake selection.

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.

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.

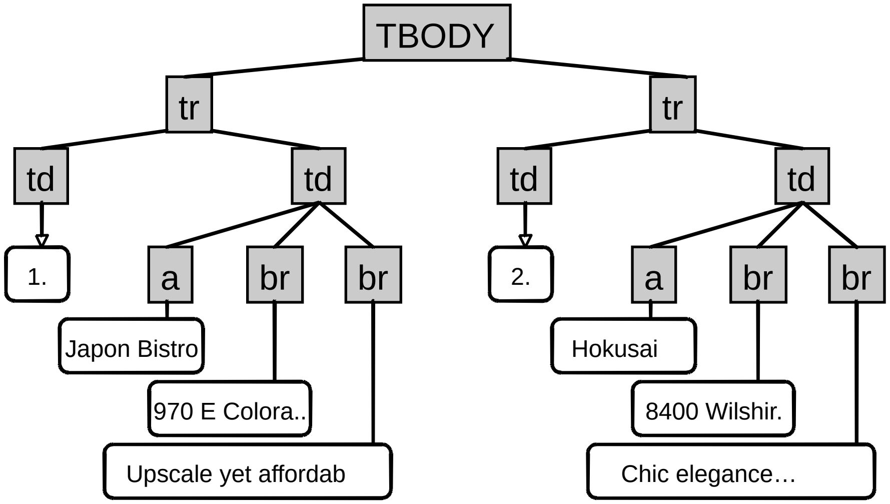
4. Sushi Roku
8445 W 3rd St , Los Angeles , CA , 90048
High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

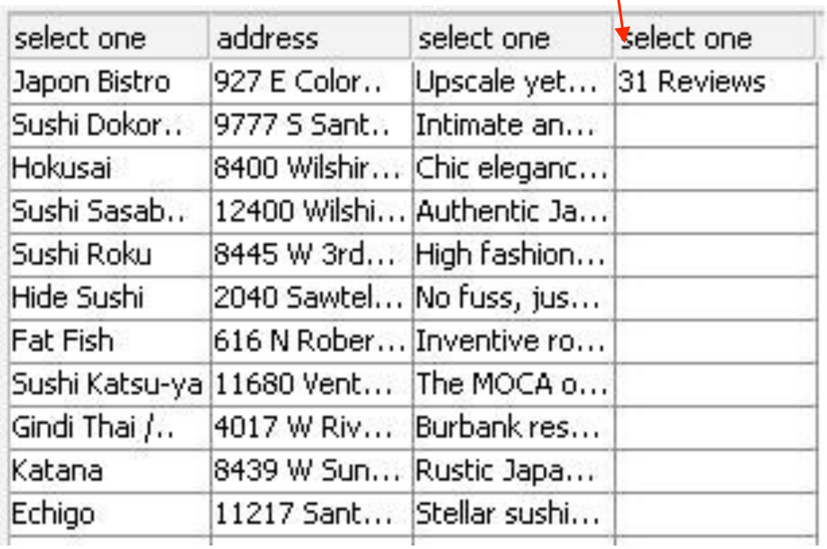


Japon Bistro

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

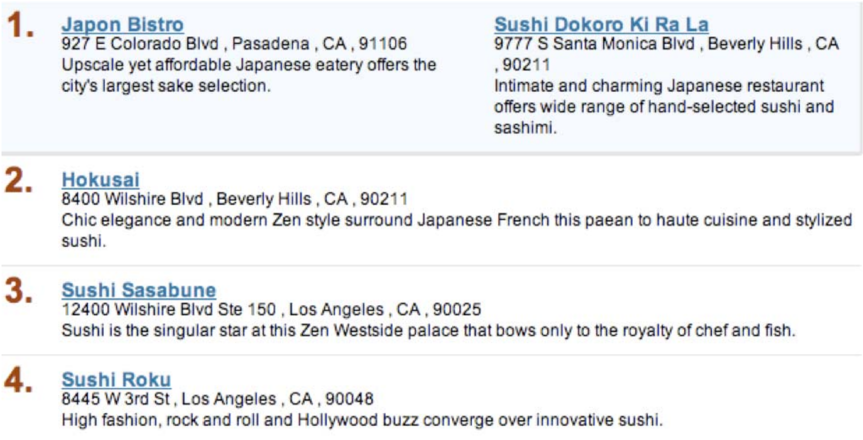
31 Reviews | Write a Review





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



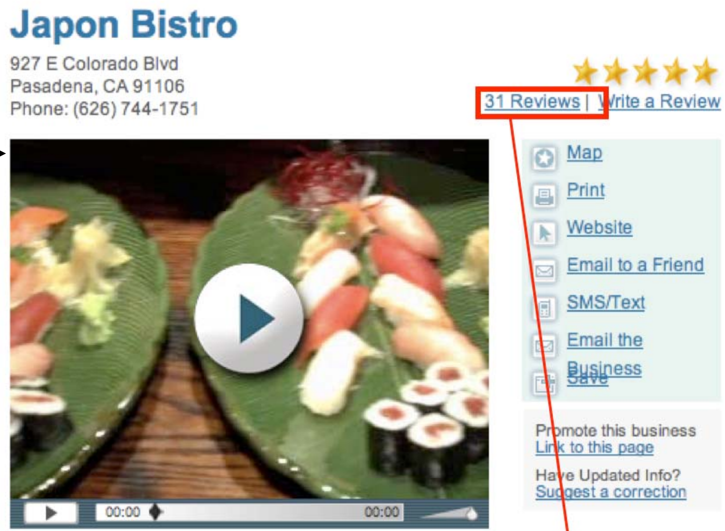
1. **Japon Bistro**
927 E Colorado Blvd , Pasadena , CA , 91106
Upscale yet affordable Japanese eatery offers the city's largest sake selection.

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.

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.



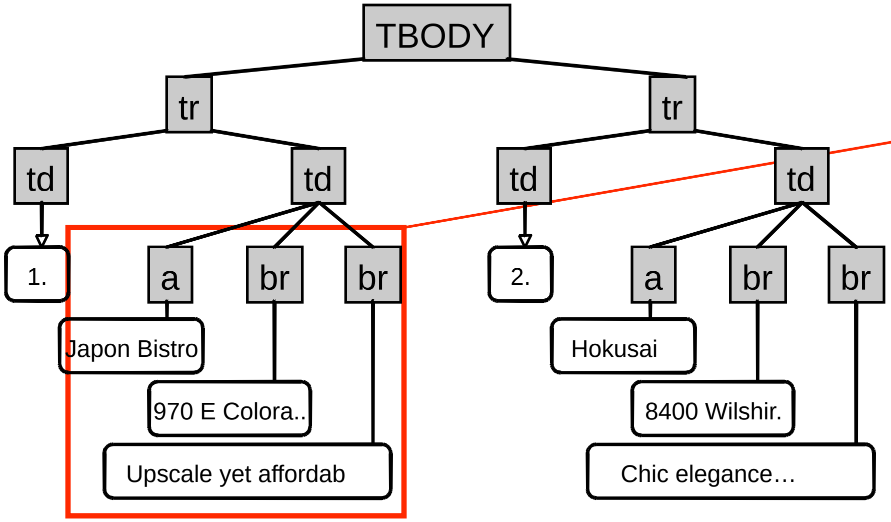
Japon Bistro

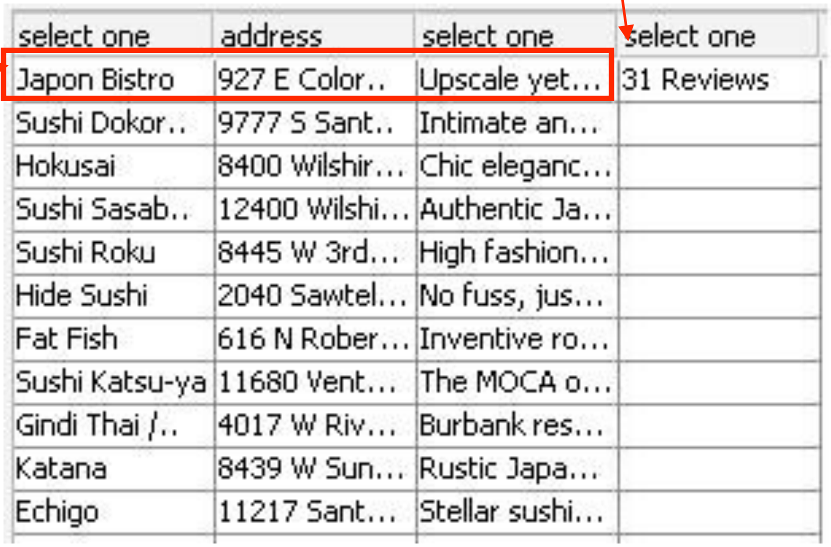
927 E Colorado Blvd
Pasadena, CA 91106
Phone: (626) 744-1751

★★★★★
31 Reviews | Write a Review

Promote this business
Link to this page

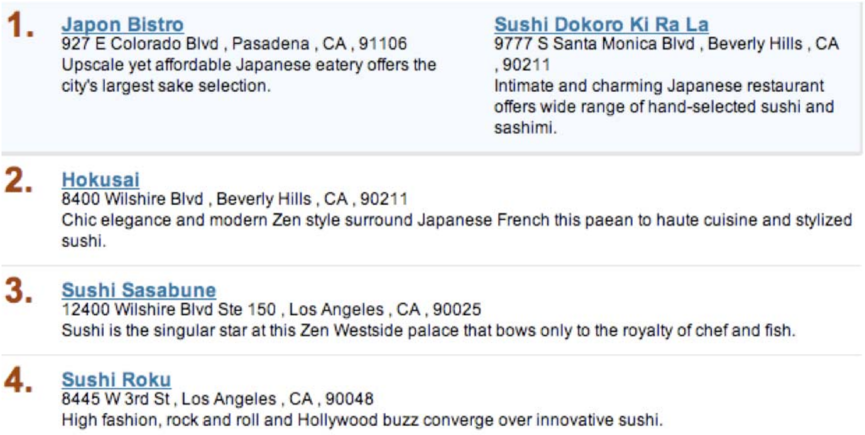
Have Updated Info?
Suggest a correction





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



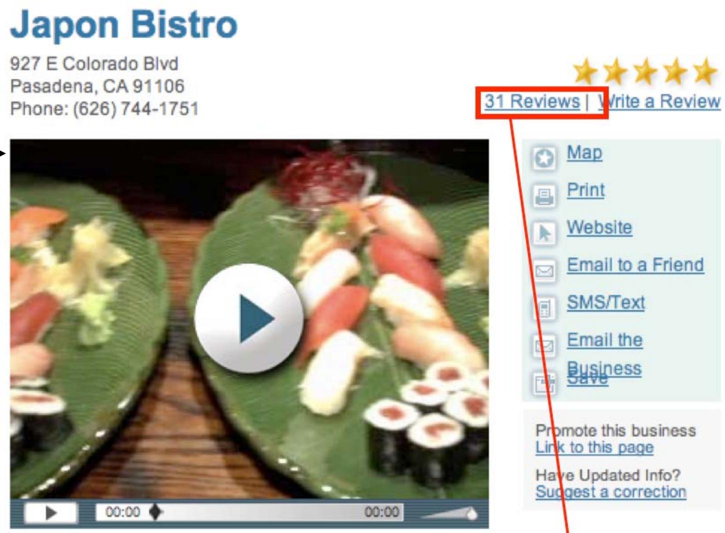
1. **Japon Bistro**
927 E Colorado Blvd , Pasadena , CA , 91106
Upscale yet affordable Japanese eatery offers the city's largest sake selection.

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.

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.

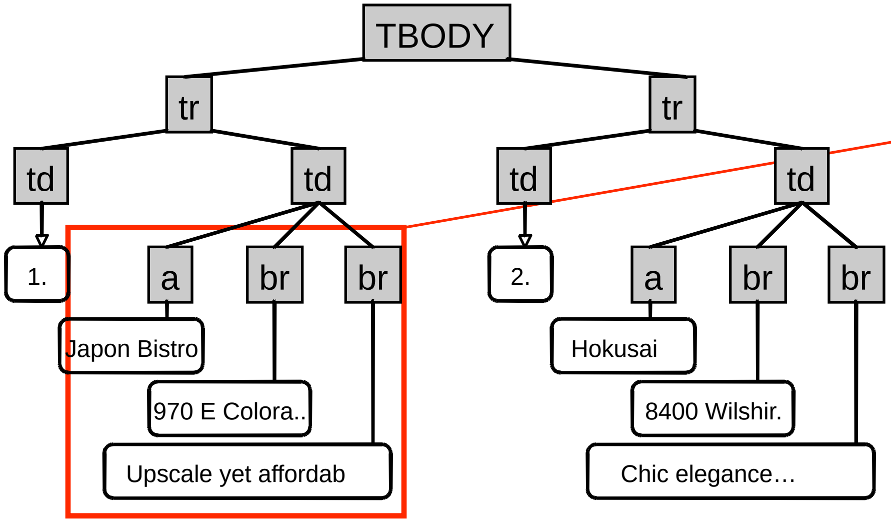
4. **Sushi Roku**
8445 W 3rd St , Los Angeles , CA , 90048
High fashion, rock and roll and Hollywood buzz converge over innovative sushi.

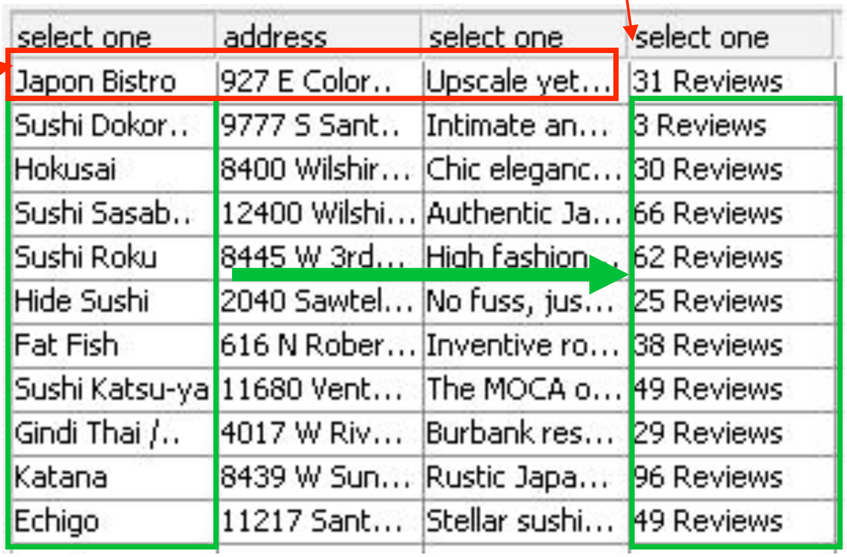


Japon Bistro
927 E Colorado Blvd
Pasadena, CA 91106
Phone: (626) 744-1751

31 Reviews | Write a Review

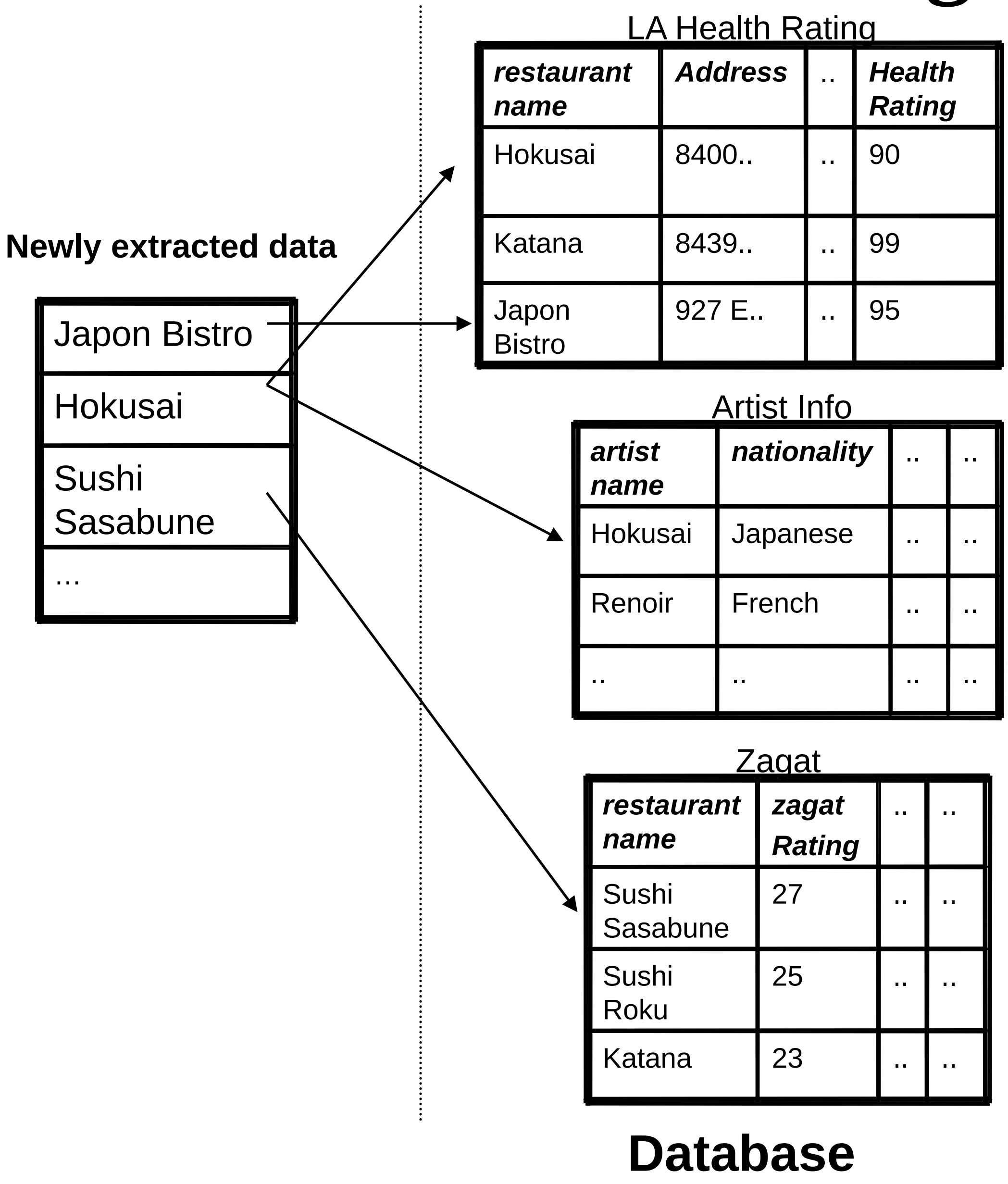
Promote this business
Link to this page
Have Updated Info?
Suggest a correction





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)



Newly extracted data

Japon Bistro
Hokusai
Sushi Sasabune
…

LA Health Rating

restaurant name	***Address***	..	***Health Rating***
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Artist Info

artist name	***nationality***	..	..
Hokusai	Japanese	..	..
Renoir	French	..	..
..	..	..	..

Zagat

restaurant name	***zagat Rating***	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

Database

Possible Attribute

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

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

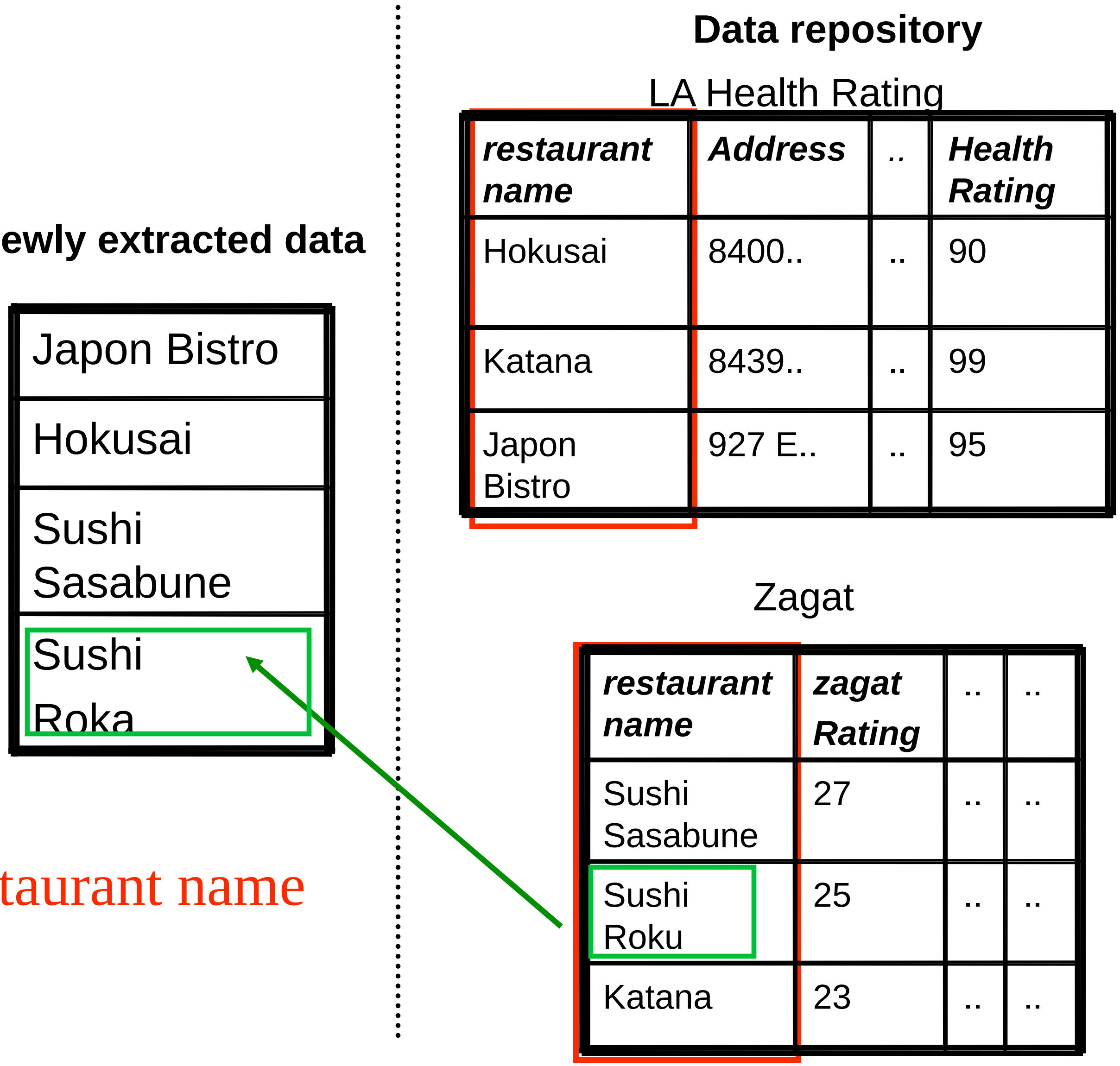
LA Health Rating

restaurant name	*Address*	..	*Health Rating*
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Zagat

restaurant name	*zagat Rating*	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

Data Cleaning: using existing values



Data repository

LA Health Rating

restaurant name	***Address***	..	***Health Rating***
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Newly extracted data

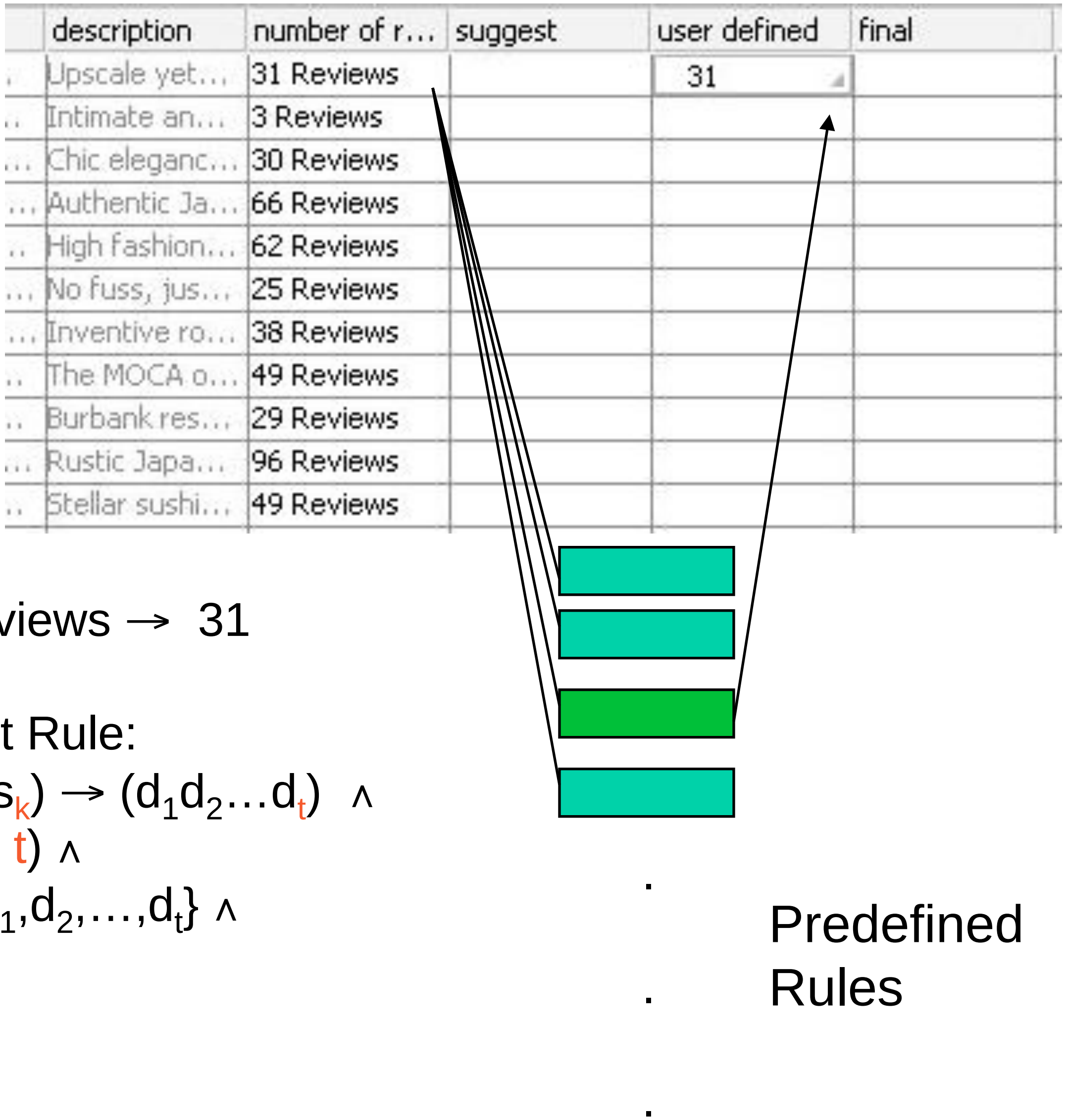
Japon Bistro
Hokusai
Sushi Sasabune
Sushi Roka

Zagat

restaurant name	***zagat Rating***	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

Restaurant name

Data Cleaning: using predefined rules



	description	number of r...	suggest	user defined	final
.	Upscale yet...	31 Reviews		31	
..	Intimate an...	3 Reviews			
...	Chic eleganc...	30 Reviews			
...	Authentic Ja...	66 Reviews			
..	High fashion...	62 Reviews			
...	No fuss, jus...	25 Reviews			
...	Inventive ro...	38 Reviews			
..	The MOCA o...	49 Reviews			
..	Burbank res...	29 Reviews			
...	Rustic Japa...	96 Reviews			
..	Stellar sushi...	49 Reviews			

Predefined Rules

31 Reviews → 31

Subset Rule:

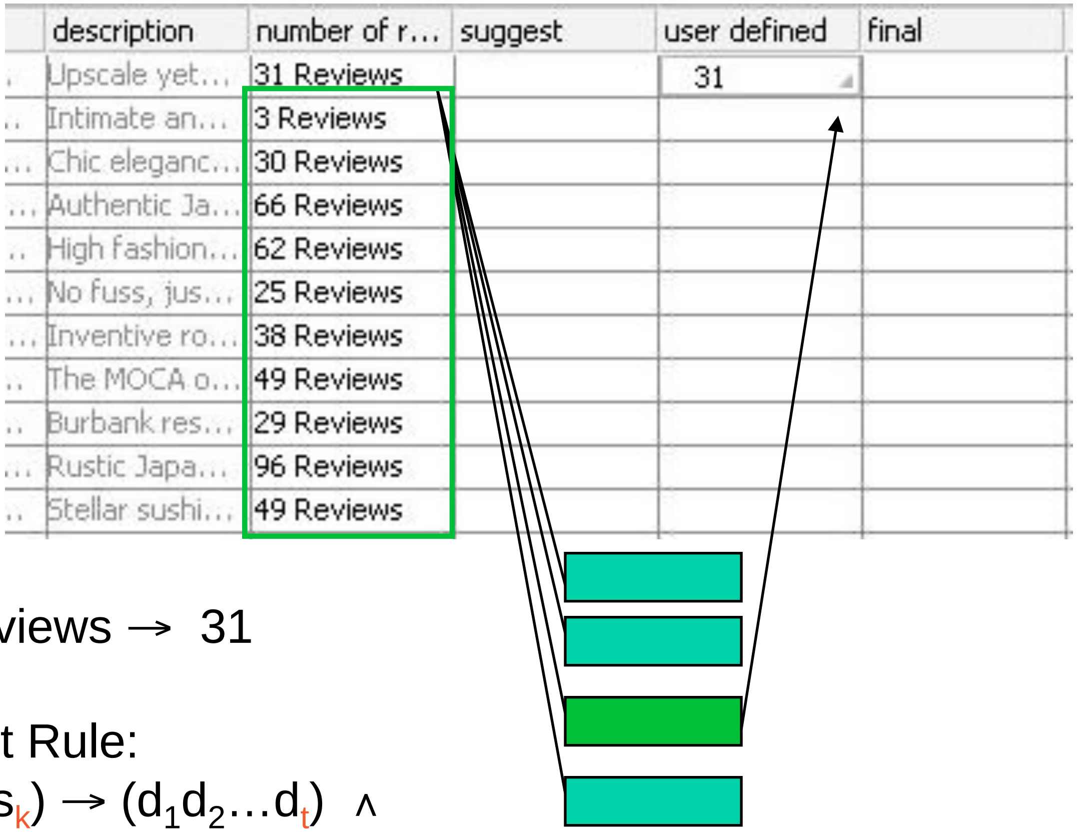
$(s_1 s_2 .. s_k) \rightarrow (d_1 d_2 \ldots d_t) \wedge$

$(k <= t) \wedge$

$s_i \in \{d_1, d_2, \ldots, d_t\} \wedge$

$d_i \neq d_j$

Data Cleaning: using predefined rules



	description	number of r…	suggest	user defined	final
.	Upscale yet…	31 Reviews		31	
..	Intimate an…	3 Reviews			
…	Chic eleganc…	30 Reviews			
…	Authentic Ja…	66 Reviews			
..	High fashion…	62 Reviews			
…	No fuss, jus…	25 Reviews			
…	Inventive ro…	38 Reviews			
..	The MOCA o…	49 Reviews			
..	Burbank res…	29 Reviews			
…	Rustic Japa…	96 Reviews			
..	Stellar sushi…	49 Reviews			

31 Reviews → 31

Subset Rule:

$(s_1 s_2 .. s_k) \rightarrow (d_1 d_2 \ldots d_t) \ \wedge$

$(k \ <= \ t) \wedge$

$s_i \in \{d_1, d_2, \ldots, d_t\} \wedge$

$d_i \neq d_j$

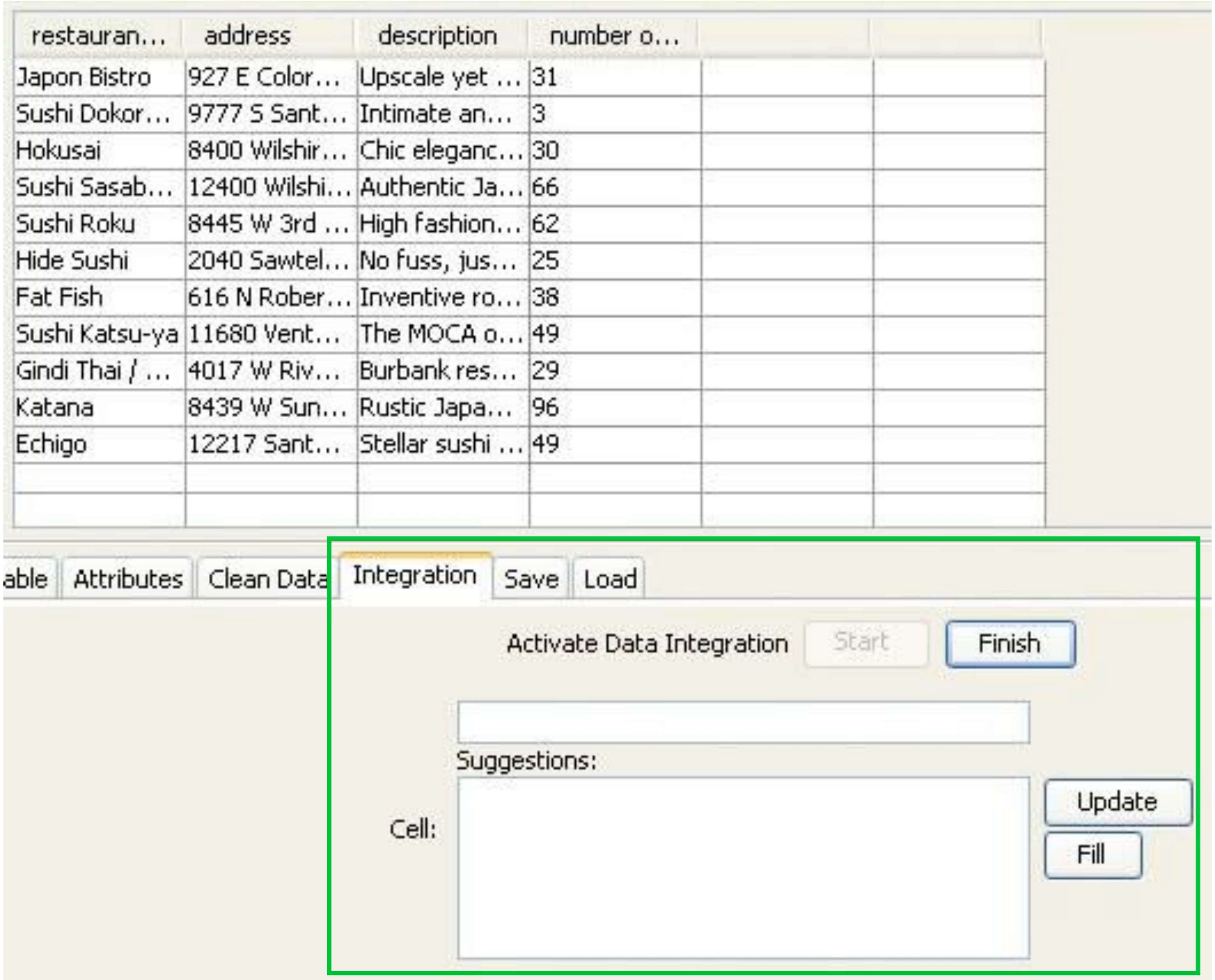
.

.

.

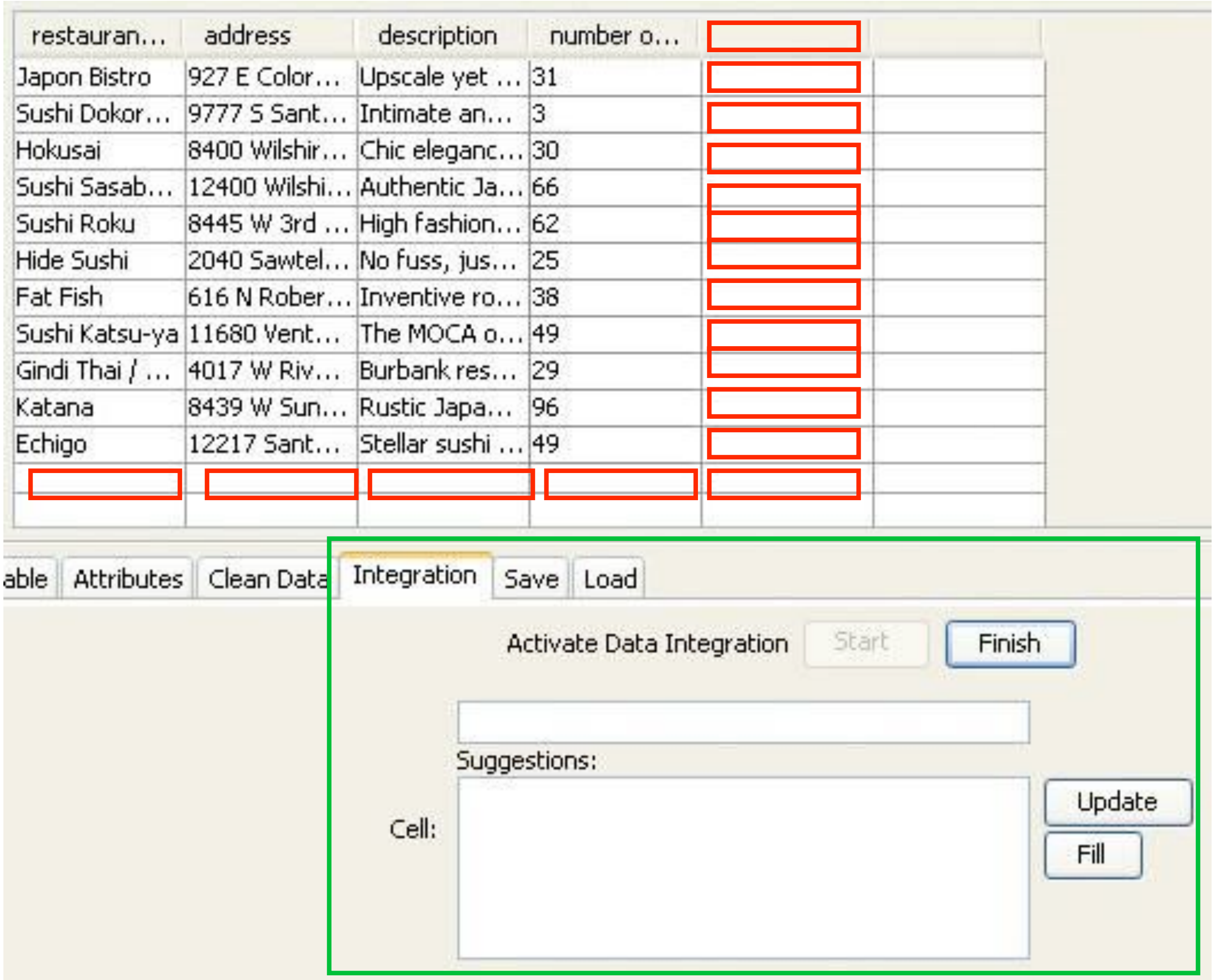
Predefined
Rules

Data Integration [tuchinda 2007]



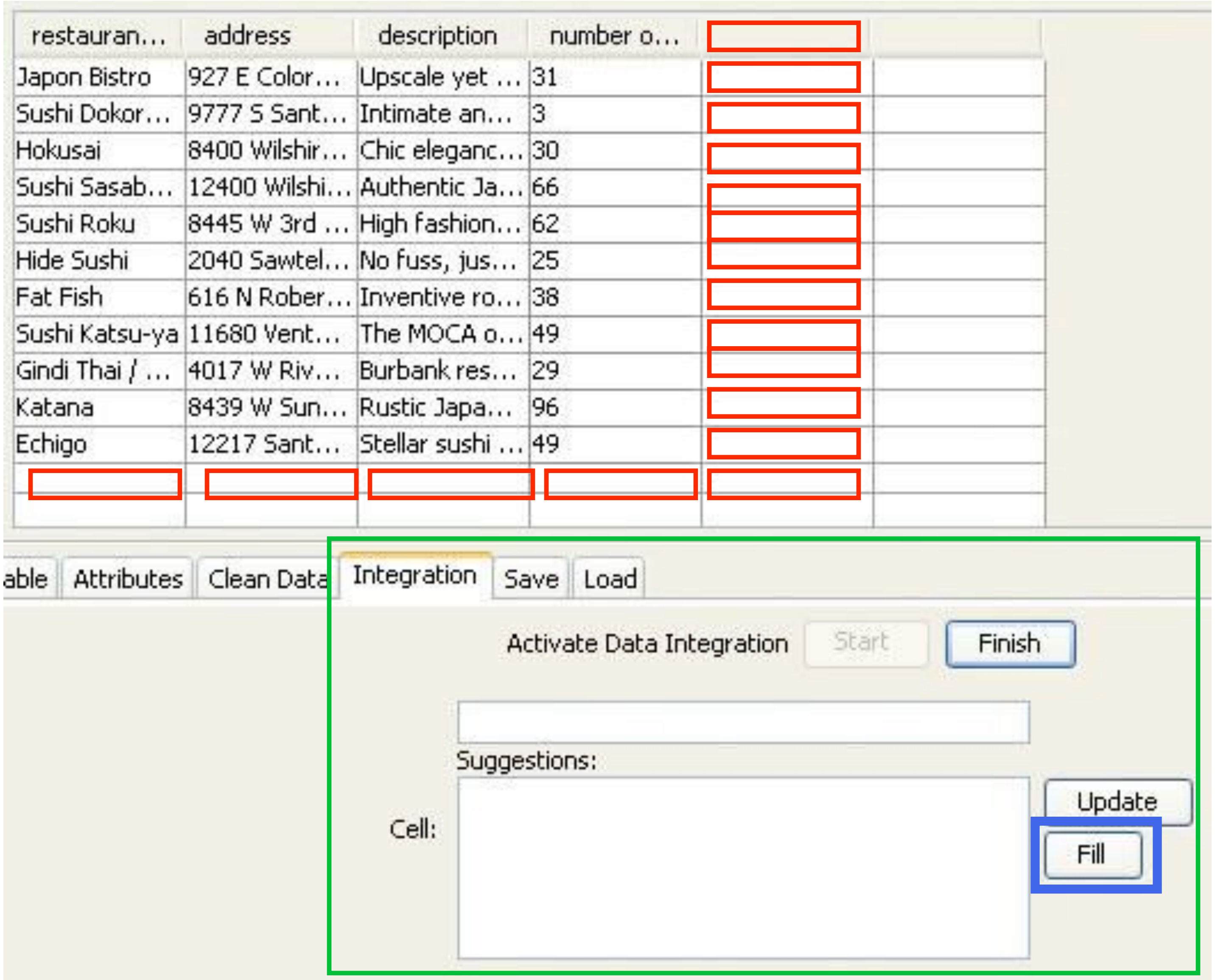
restauran...	address	description	number o...
Japon Bistro	927 E Color...	Upscale yet ...	31
Sushi Dokor...	9777 S Sant...	Intimate an...	3
Hokusai	8400 Wilshir...	Chic eleganc...	30
Sushi Sasab...	12400 Wilshi...	Authentic Ja...	66
Sushi Roku	8445 W 3rd ...	High fashion...	62
Hide Sushi	2040 Sawtel...	No fuss, jus...	25
Fat Fish	616 N Rober...	Inventive ro...	38
Sushi Katsu-ya	11680 Vent...	The MOCA o...	49
Gindi Thai / ...	4017 W Riv...	Burbank res...	29
Katana	8439 W Sun...	Rustic Japa...	96
Echigo	12217 Sant...	Stellar sushi ...	49

Data Integration [tuchinda 2007]



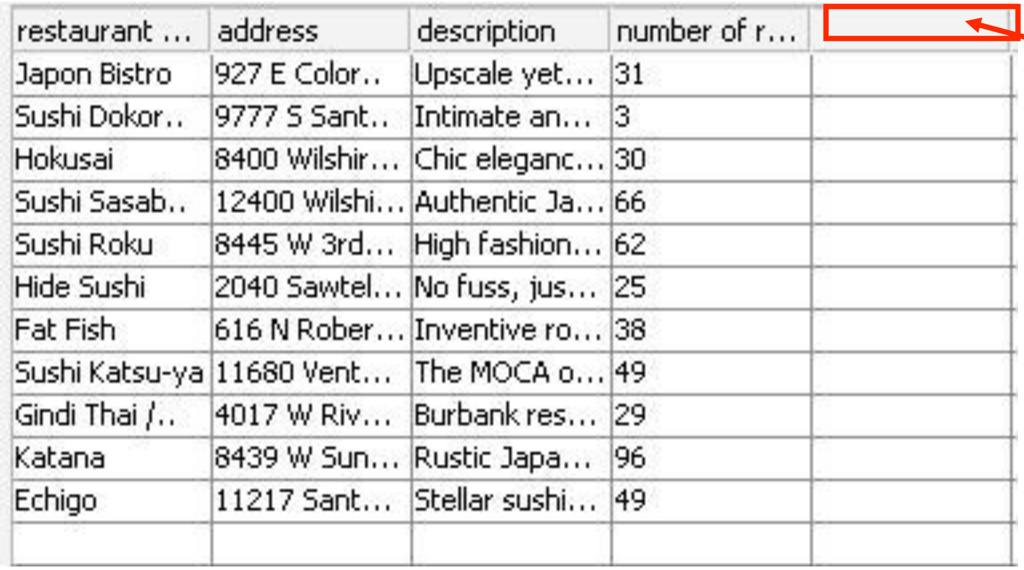
restauran…	address	description	number o…	
Japon Bistro	927 E Color…	Upscale yet …	31	
Sushi Dokor…	9777 S Sant…	Intimate an…	3	
Hokusai	8400 Wilshir…	Chic eleganc…	30	
Sushi Sasab…	12400 Wilshi…	Authentic Ja…	66	
Sushi Roku	8445 W 3rd …	High fashion…	62	
Hide Sushi	2040 Sawtel…	No fuss, jus…	25	
Fat Fish	616 N Rober…	Inventive ro…	38	
Sushi Katsu-ya	11680 Vent…	The MOCA o…	49	
Gindi Thai / …	4017 W Riv…	Burbank res…	29	
Katana	8439 W Sun…	Rustic Japa…	96	
Echigo	12217 Sant…	Stellar sushi …	49	

Data Integration [tuchinda 2007]



restauran…	address	description	number o…	
Japon Bistro	927 E Color…	Upscale yet …	31	
Sushi Dokor…	9777 S Sant…	Intimate an…	3	
Hokusai	8400 Wilshir…	Chic eleganc…	30	
Sushi Sasab…	12400 Wilshi…	Authentic Ja…	66	
Sushi Roku	8445 W 3rd …	High fashion…	62	
Hide Sushi	2040 Sawtel…	No fuss, jus…	25	
Fat Fish	616 N Rober…	Inventive ro…	38	
Sushi Katsu-ya	11680 Vent…	The MOCA o…	49	
Gindi Thai / …	4017 W Riv…	Burbank res…	29	
Katana	8439 W Sun…	Rustic Japa…	96	
Echigo	12217 Sant…	Stellar sushi …	49	

Data Integration (cont.)



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

Data repository

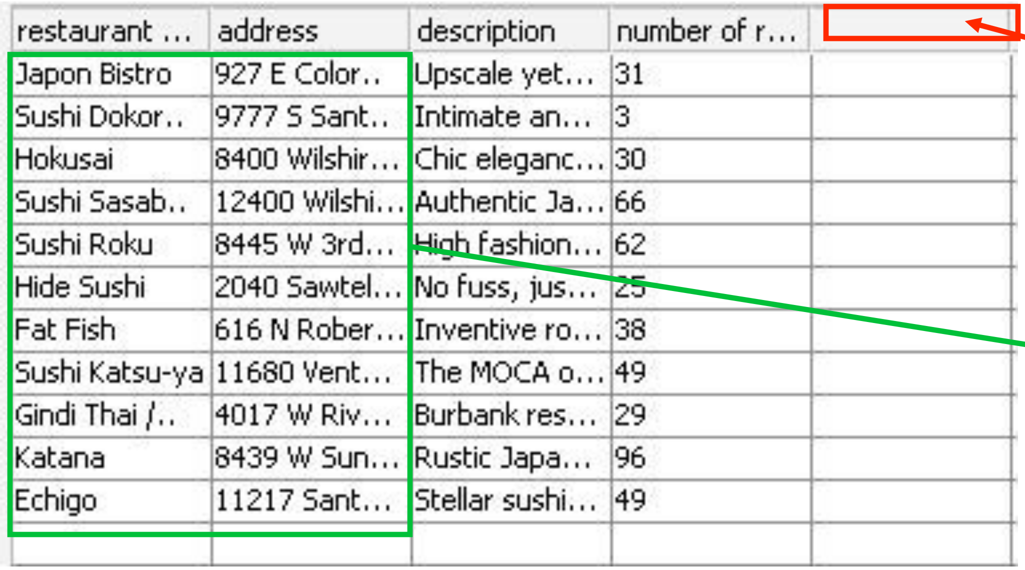
LA Health Rating

restaurant name	***Address***	..	***Health Rating***
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Zagat

restaurant name	***zagat Rating***	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

Data Integration (cont.)



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

Data repository

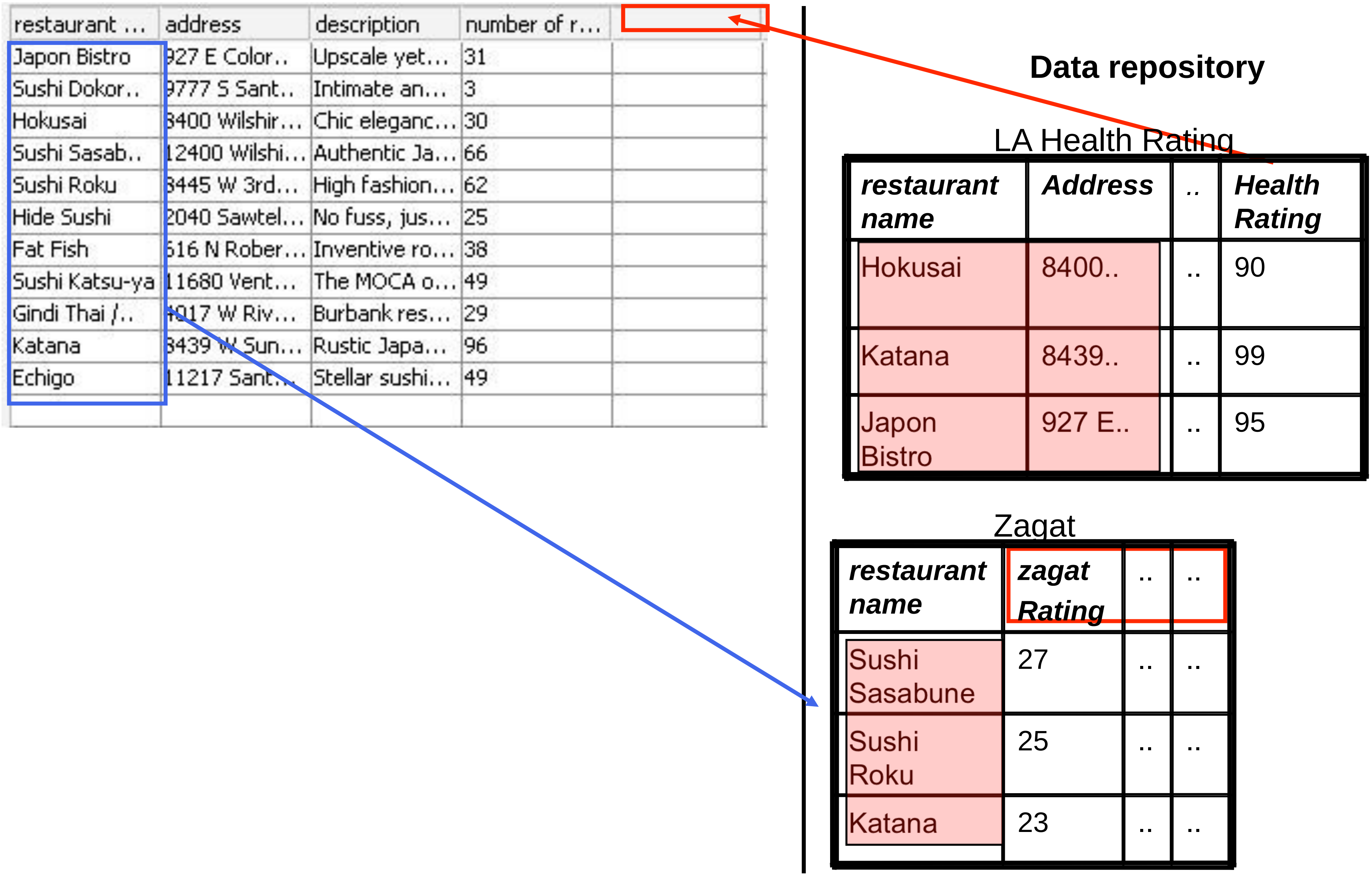
LA Health Rating

restaurant name	*Address*	..	*Health Rating*
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Zagat

restaurant name	*zagat Rating*	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

Data Integration (cont.)



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

Data repository

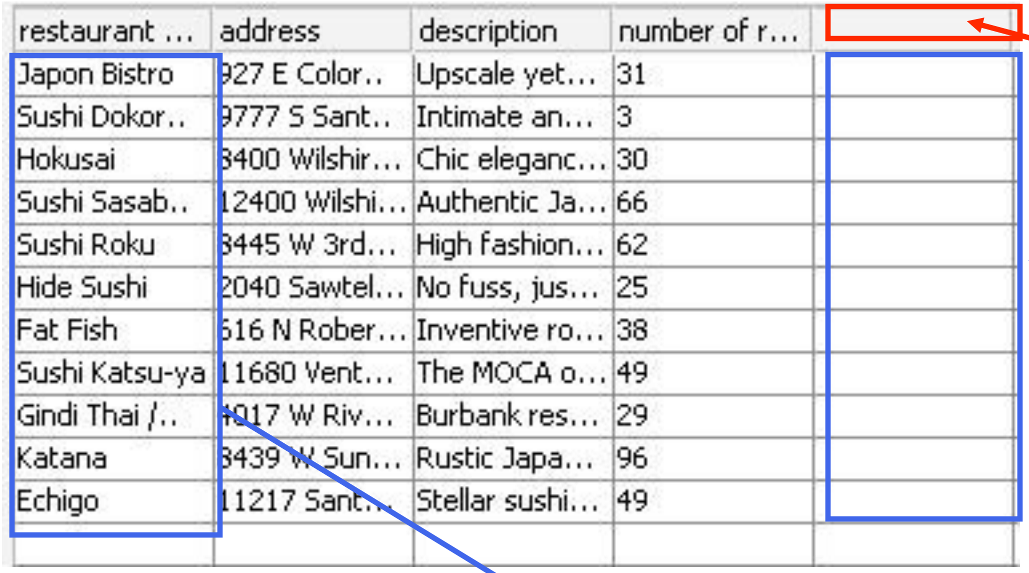
LA Health Rating

restaurant name	***Address***	..	***Health Rating***
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Zagat

restaurant name	***zagat Rating***	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

Data Integration (cont.)



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

Data repository

LA Health Rating

restaurant name	*Address*	..	*Health Rating*
Hokusai	8400..	..	90
Katana	8439..	..	99
Japon Bistro	927 E..	..	95

Zagat

restaurant name	*zagat Rating*	..	..
Sushi Sasabune	27	..	..
Sushi Roku	25	..	..
Katana	23	..	..

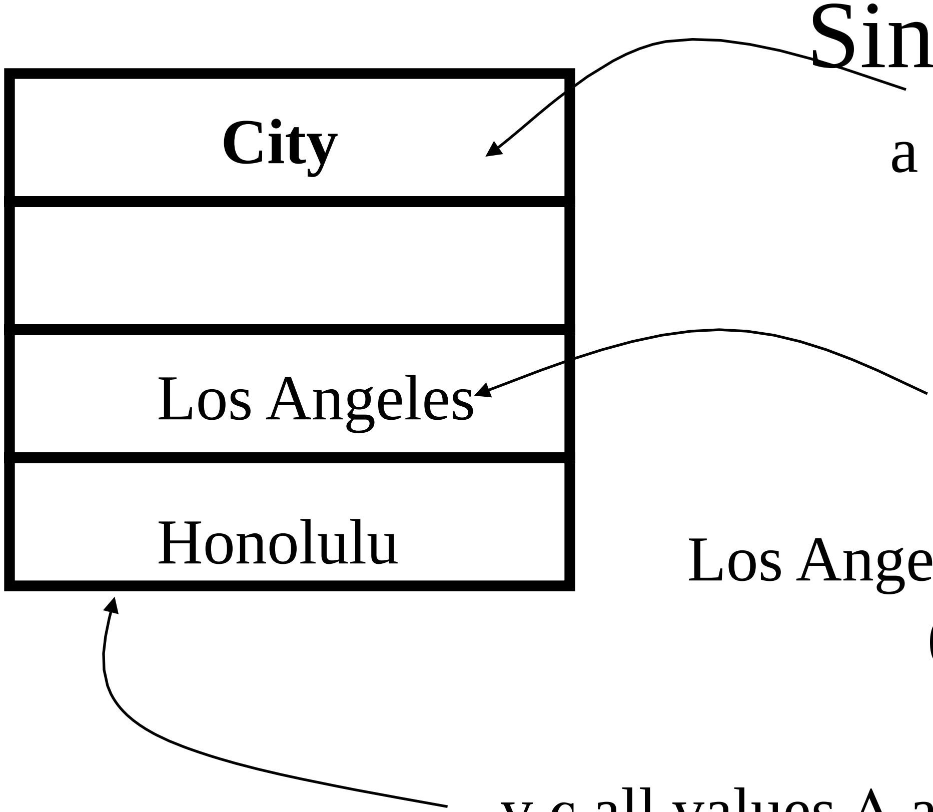
$\{a\}_R$ = possible new attribute selection for row *i*.

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

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

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

Single Column Example



a ϵ all attribute

v ϵ all value

Los Angeles : {(**city**, tax_properties),
~~(**song name**, pop_music~~)}

v ϵ all values Λ attributeOf(v) ϵ {city, song name}

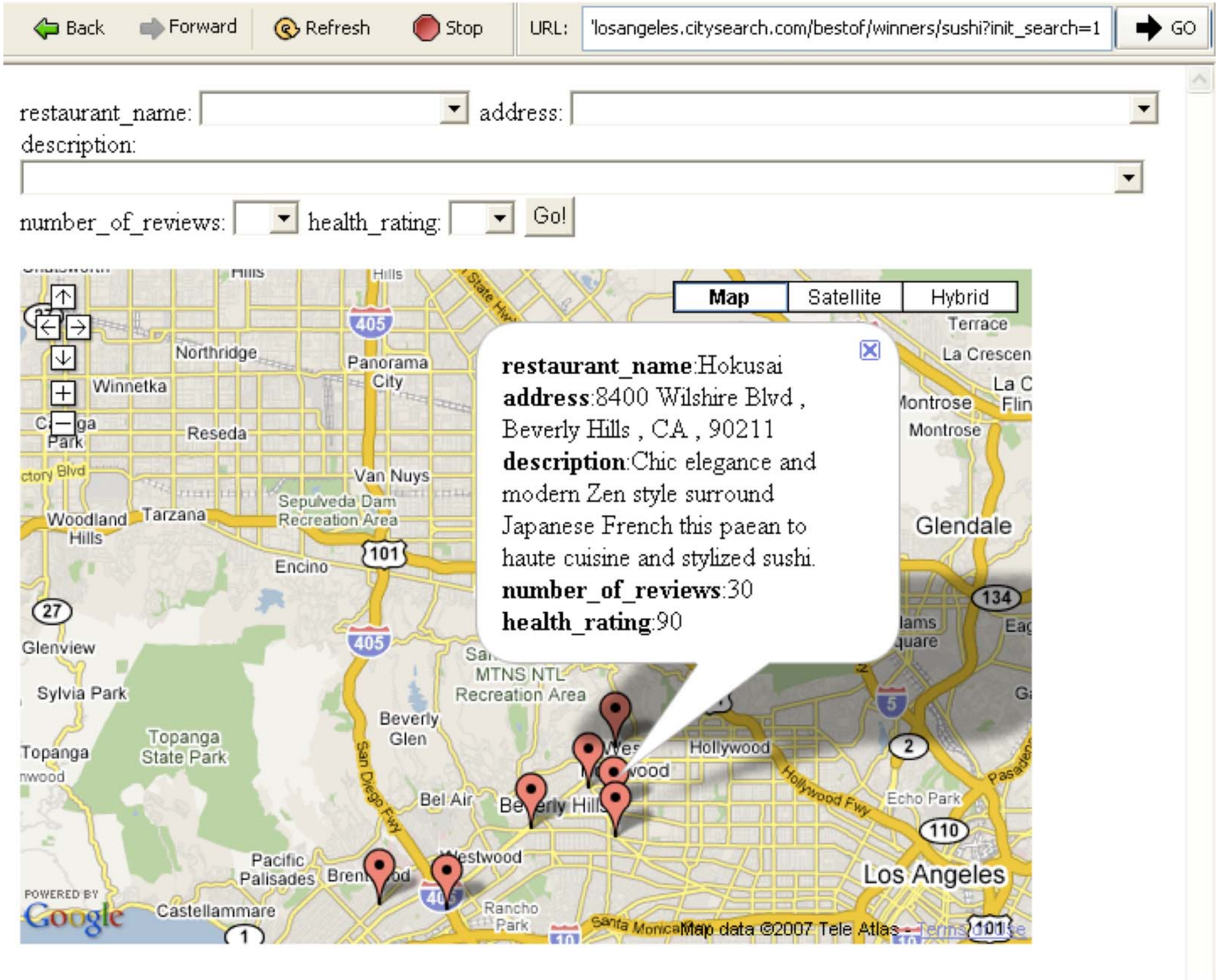
Can we determine the attribute now? Yes

Honolulu : {(**city**, tax_properties),
(**city**, favorite_vacation_spot)}

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

$\{v\} = val(a,s)$ where $a \varepsilon \{x\} \Lambda s$ is any source where $att(s) \cap \{x\} \neq \{\}$

Map Generation



URL: 'losangeles.citysearch.com/bestof/winners/sushi?init_search=1

restaurant_name: address:

description:

number_of_reviews: health_rating:

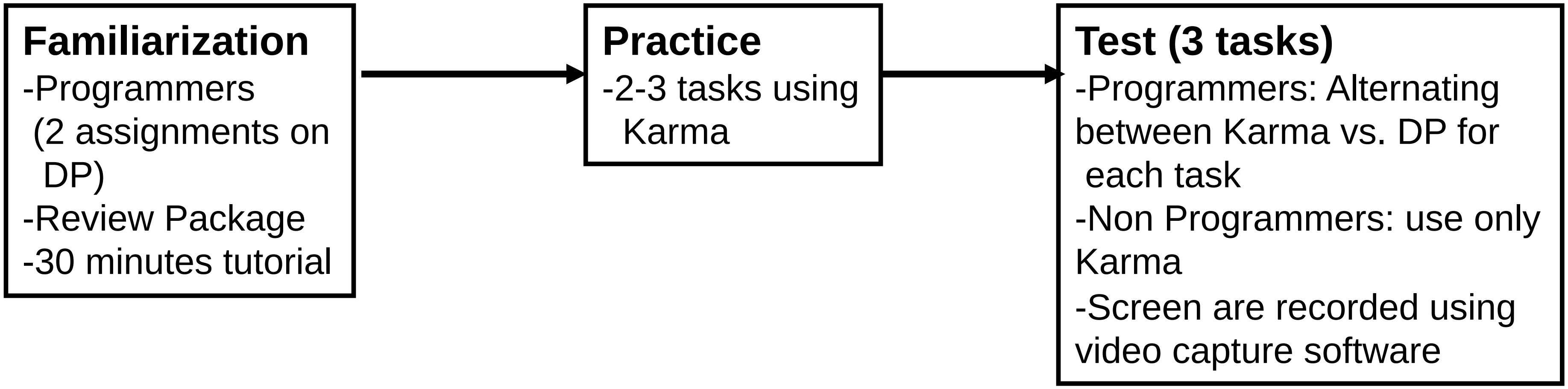
restaurant_name:Hokusai
address:8400 Wilshire Blvd , Beverly Hills , CA , 90211
description:Chic elegance and modern Zen style surround Japanese French this paean to haute cuisine and stylized sushi.
number_of_reviews:30
health_rating:90

Evaluation

- Baseline: A combination of Dapper/Pipes
- Claims:
  1. Users with no programming experiences can build all four Mashup types.
  2. Karma takes less time to complete each subtask.
  3. Overall, the user takes less time to build the same Mashup in Karma compared to Dapper/Pipes
- Users:
  - Programmers (20)
  - Non-programmers (3)

If subjects (programmers) who are familiar with workflow and widgets spend more time on Dapper/Pipes in general, then the non-programmer subjects would spend more time on Dapper/Pipes as well if they were to learn how to use those systems.

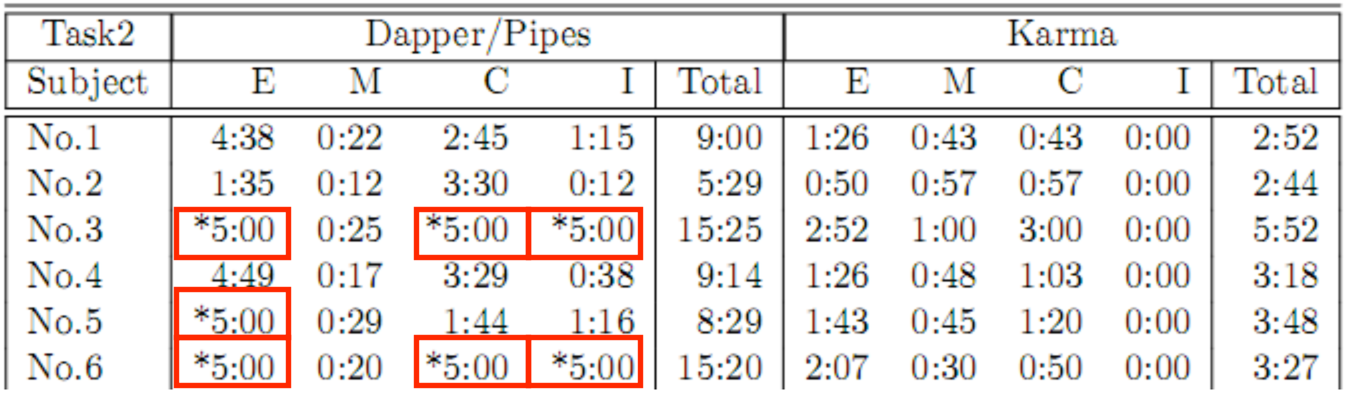
Evaluation: Setup



Familiarization
-Programmers
(2 assignments on DP)
-Review Package
-30 minutes tutorial

Practice
-2-3 tasks using Karma

Test (3 tasks)
-Programmers: Alternating between Karma vs. DP for each task
-Non Programmers: use only Karma
-Screen are recorded using video capture software



Task2	Dapper/Pipes					Karma				
Subject	E	M	C	I	Total	E	M	C	I	Total
No.1	4:38	0:22	2:45	1:15	9:00	1:26	0:43	0:43	0:00	2:52
No.2	1:35	0:12	3:30	0:12	5:29	0:50	0:57	0:57	0:00	2:44
No.3	*5:00	0:25	*5:00	*5:00	15:25	2:52	1:00	3:00	0:00	5:52
No.4	4:49	0:17	3:29	0:38	9:14	1:26	0:48	1:03	0:00	3:18
No.5	*5:00	0:29	1:44	1:16	8:29	1:43	0:45	1:20	0:00	3:48
No.6	*5:00	0:20	*5:00	*5:00	15:20	2:07	0:30	0:50	0:00	3:27

Using 5 minutes cut off time

Evaluation: Tasks

Task No.	Mashup Type	Data Extraction	Source Modeling	Data Cleaning	Data Integration
1	1 (1 source)	Moderate	Simple	Difficult	N/A
2	2,3 (union+form)	Difficult	Simple	Simple	Union (simple)
3	4 (join 2 sources)	Simple	Simple	N/A	Join (difficult)

Claim 1

Users with no programming experiences can build all four Mashup types.

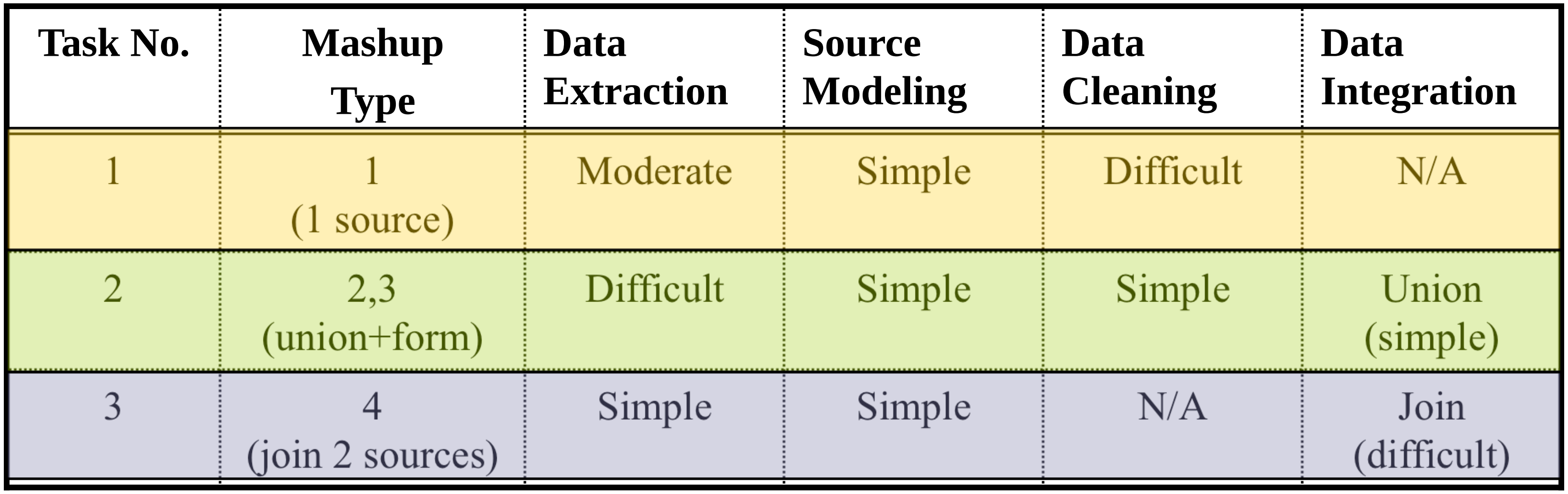
Evaluation: Tasks

Task No.	Mashup Type	Data Extraction	Source Modeling	Data Cleaning	Data Integration
1	1 (1 source)	Moderate	Simple	Difficult	N/A
2	2,3 (union+form)	Difficult	Simple	Simple	Union (simple)
3	4 (join 2 sources)	Simple	Simple	N/A	Join (difficult)

Claim 2

When the Mashup subtask is difficult, Karma takes less time to complete that subtask.

Evaluation: Tasks



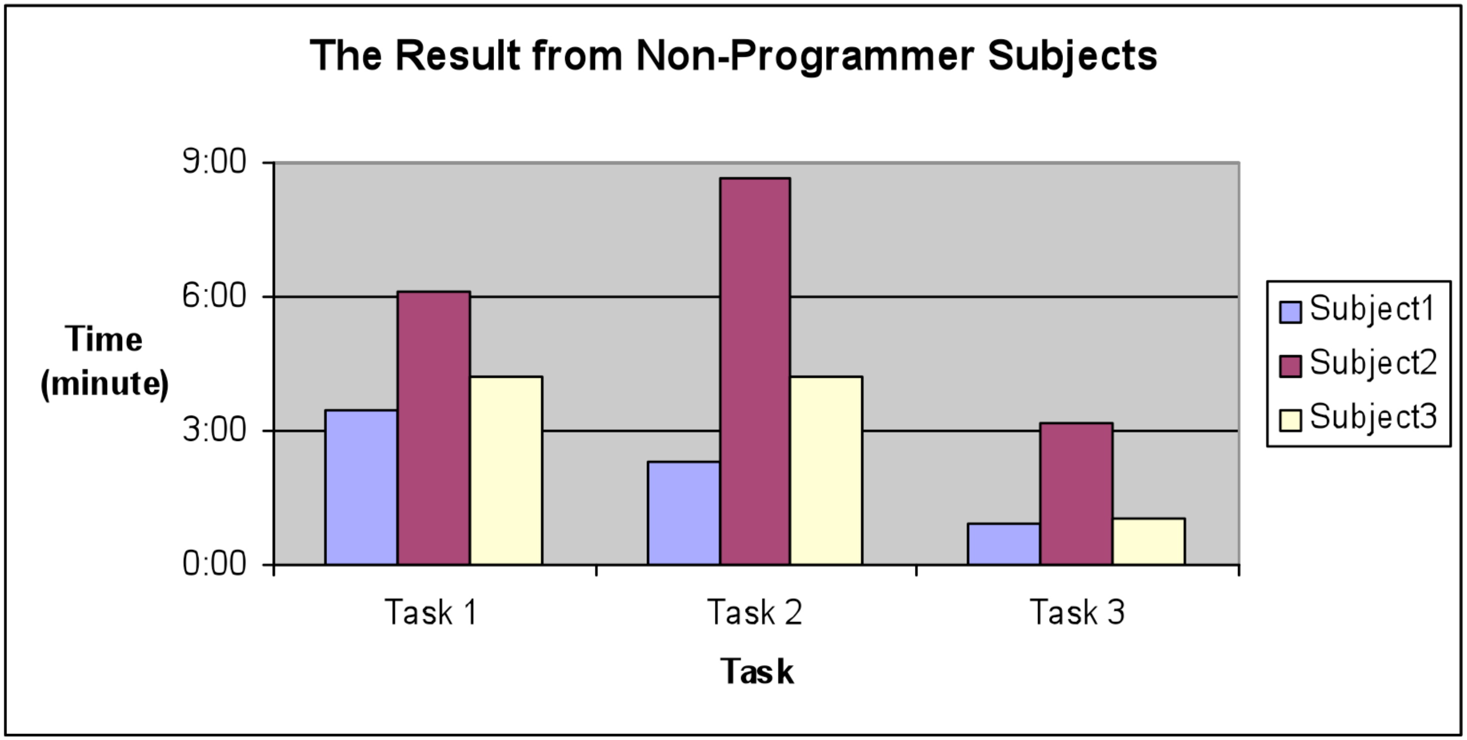
Task No.	Mashup Type	Data Extraction	Source Modeling	Data Cleaning	Data Integration
1	1 (1 source)	Moderate	Simple	Difficult	N/A
2	2,3 (union+form)	Difficult	Simple	Simple	Union (simple)
3	4 (join 2 sources)	Simple	Simple	N/A	Join (difficult)

Claim 3

Overall, the user takes less time to build the same Mashup in Karma compared to Dapper/Pipes

Claim 1: Users with no programming experiences can build all four Mashup types

Evaluation: Non-Programmers

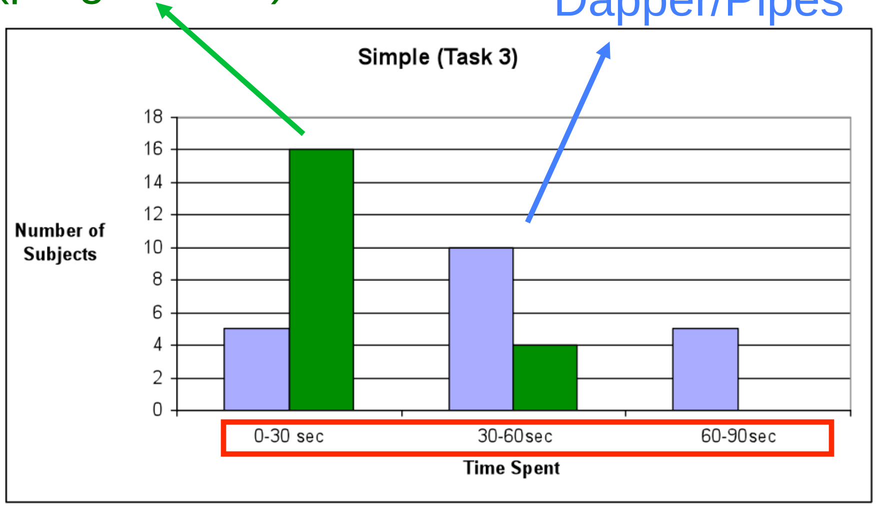


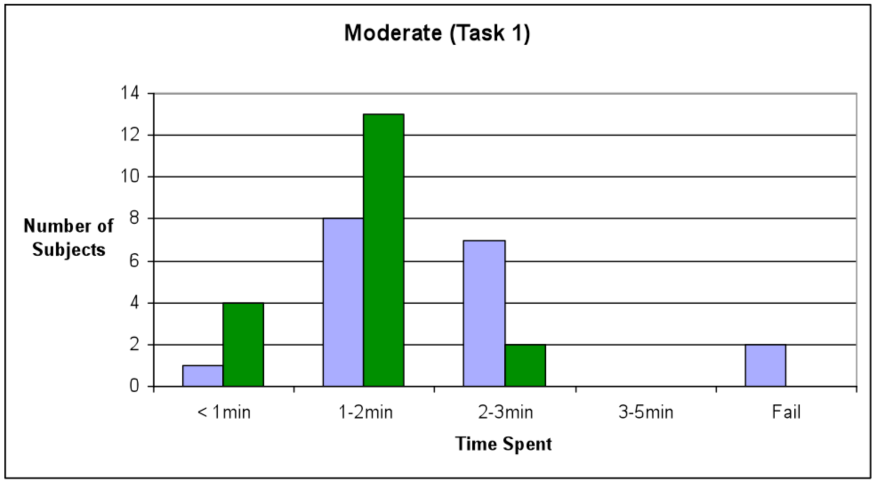
Claim 2: Karma takes less time to complete each subtask

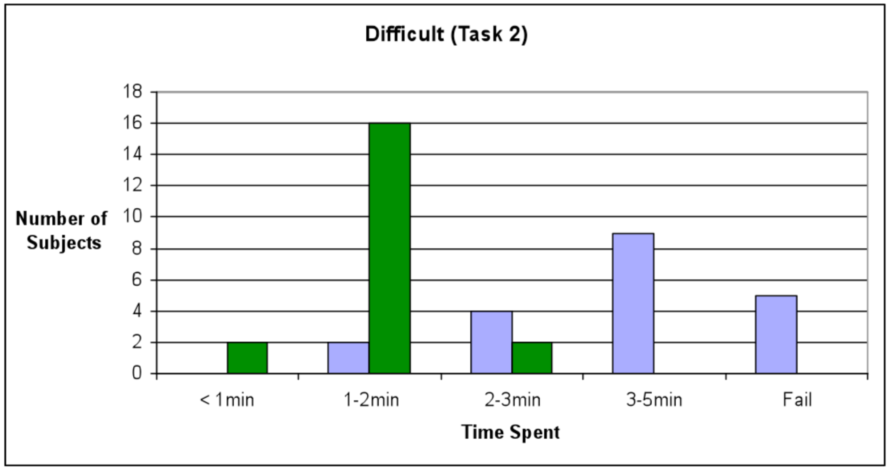
Evaluation: Extraction

Karma (programmer)

Dapper/Pipes

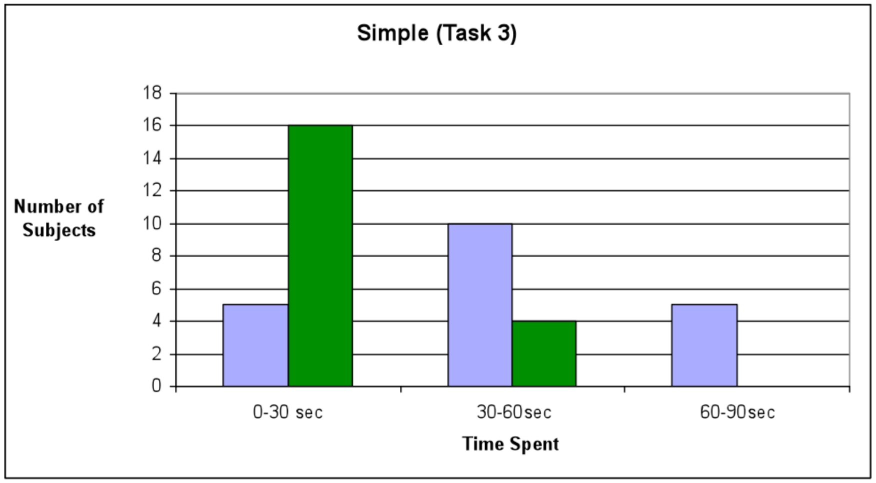


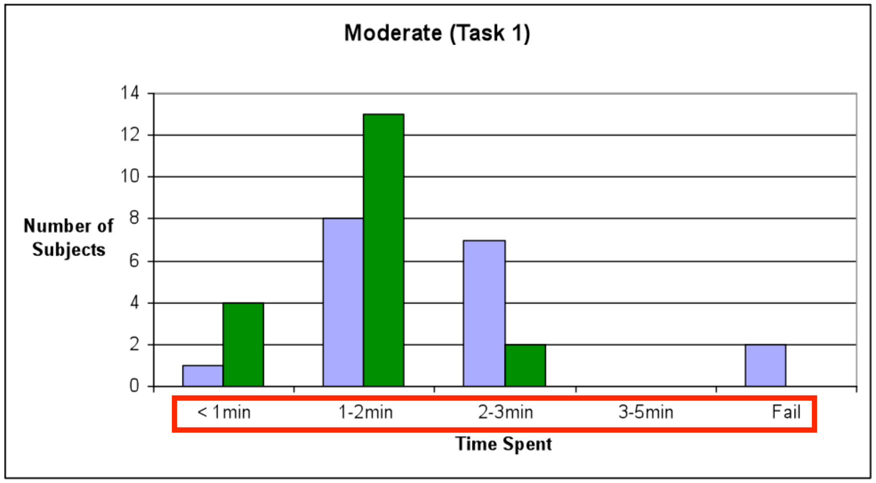




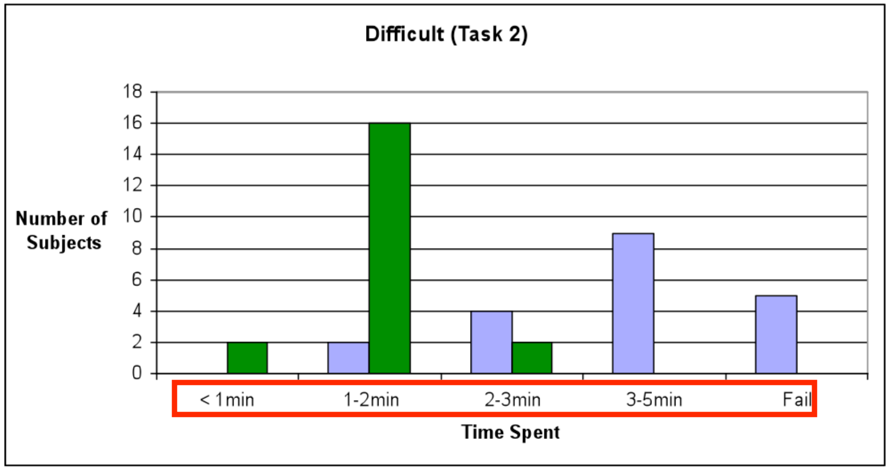
🟦 **Dapper/Pipes** 🟩 **Karma**

Evaluation: Extraction



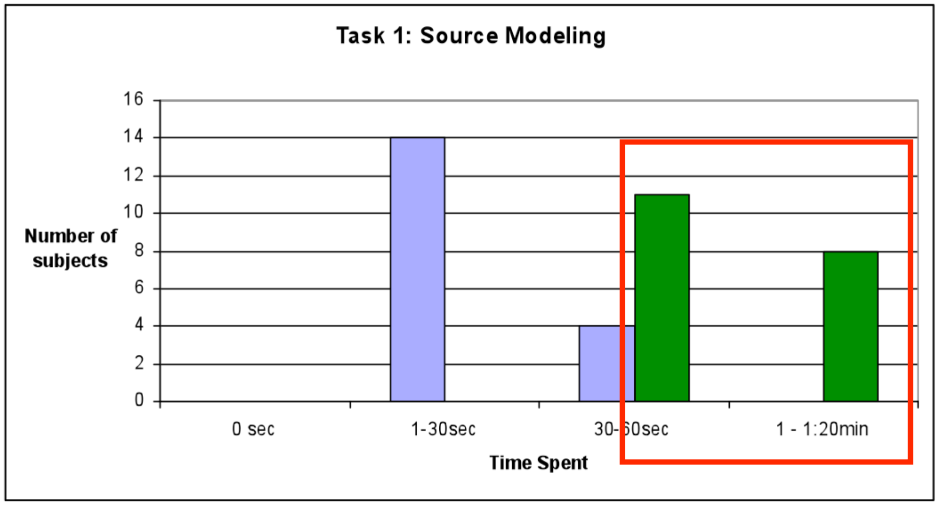


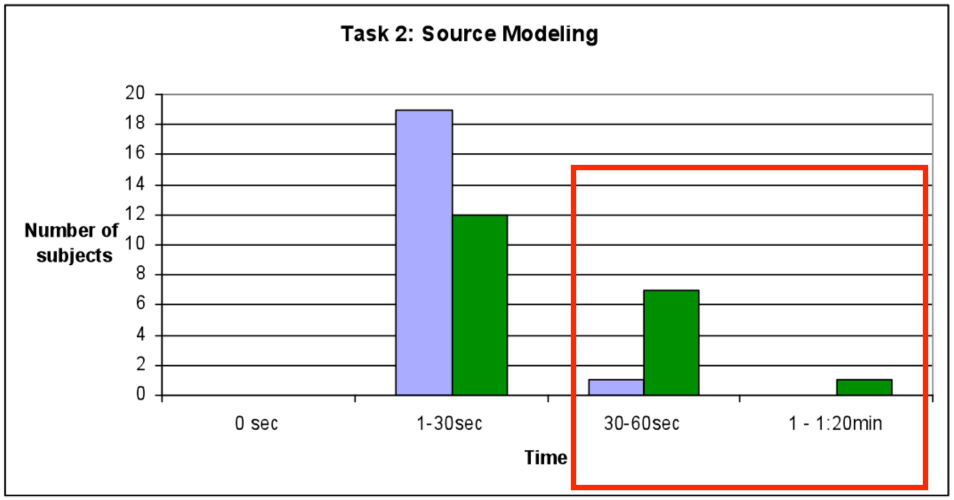
- As the extraction task gets more difficult, Dapper/Pipes takes
  - longer
  - more subjects failing to complete the task (11% for moderate and 25% for difficult)



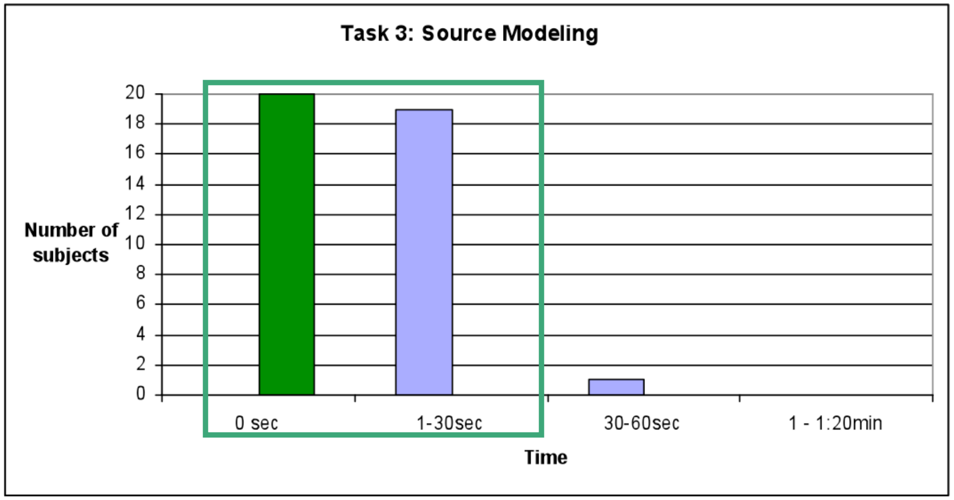
🟦 **Dapper/Pipes** 🟩 **Karma**

Evaluation: Source Modeling


Task 1: Source Modeling

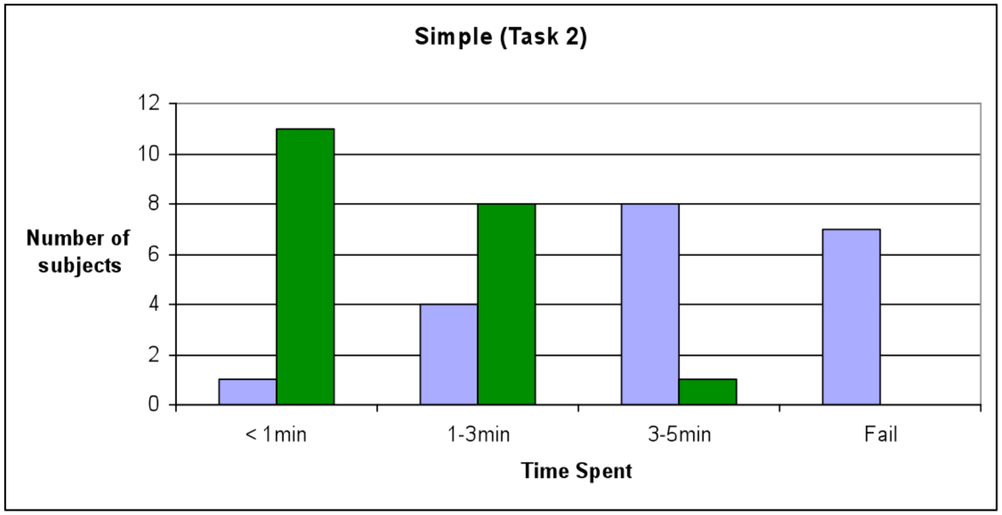

Task 2: Source Modeling

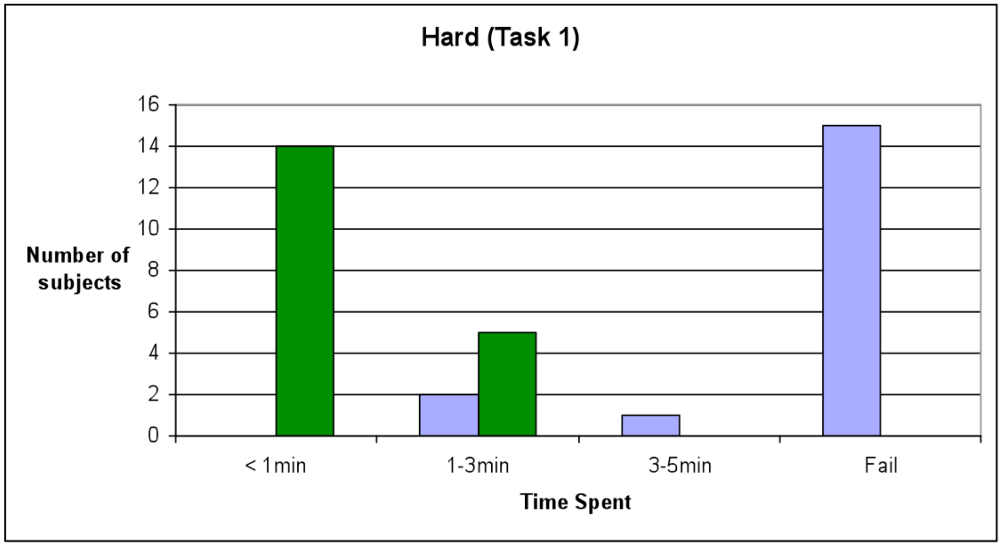
- Karma performed worse in task 1 and tasks 2
  - only 30 sec difference
  -subjects take times selecting attributes
  - the saving will be realized in the data integration step.
- Karma performed better in task 3 because of union


Task 3: Source Modeling

🟦 **Dapper/Pipes** 🟩 **Karma**

Evaluation: Data Cleaning


Simple (Task 2)

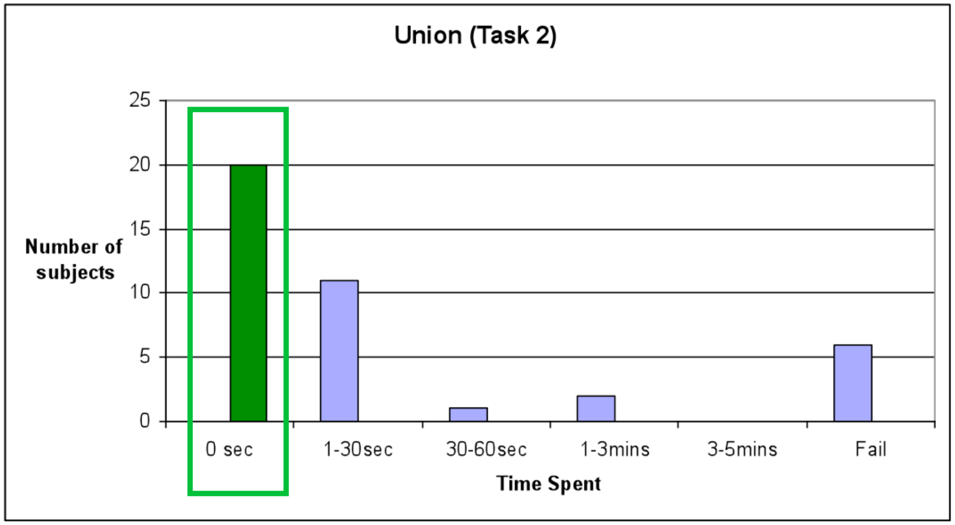

Hard (Task 1)

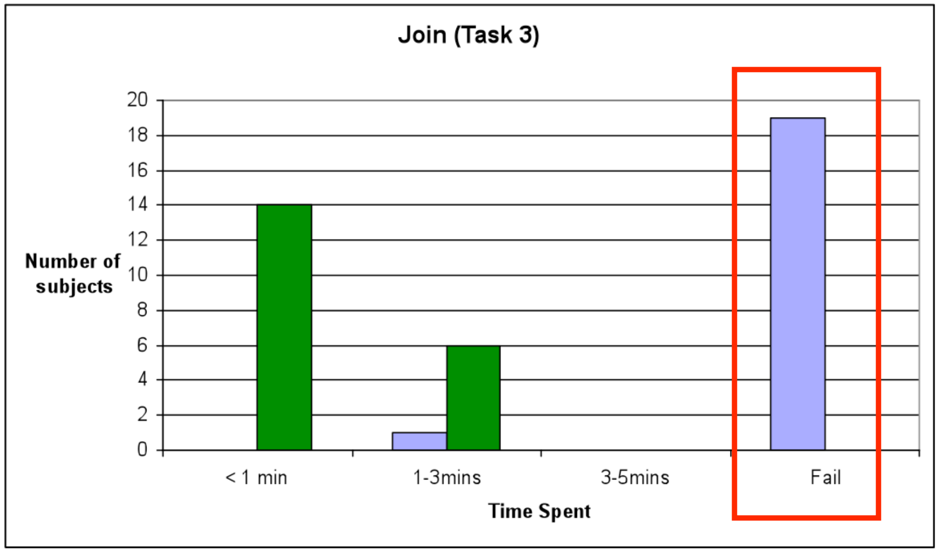
- Karma performed better in both tasks

- When the cleaning task gets harder, more subjects are failing in Dapper/ Pipes (35% for simple and 83% in hard)

🟦 **Dapper/Pipes** 🟩 **Karma**

Evaluation: Data Integration





- Because of the table structure, subjects can specify union indirectly by dropping data into the right cell

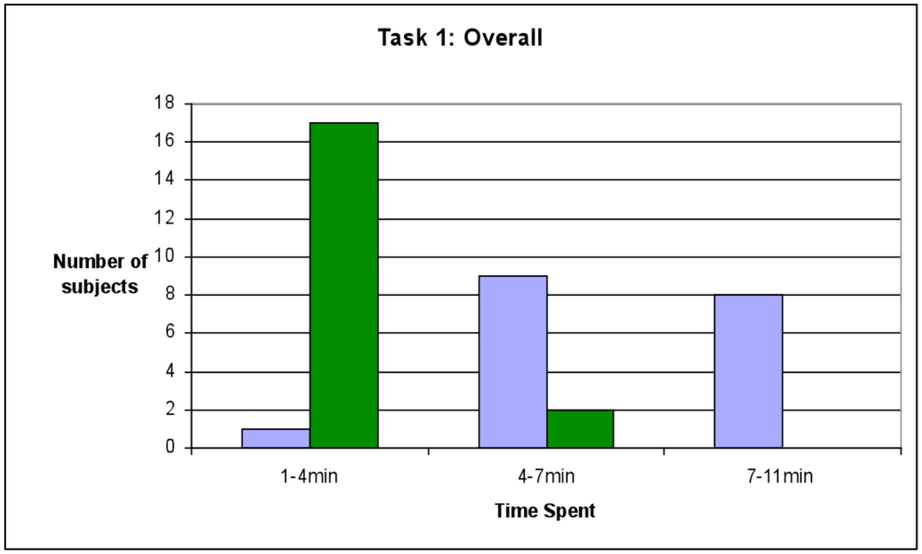
- The time spent in source modeling step allows Karma to suggest the linking source

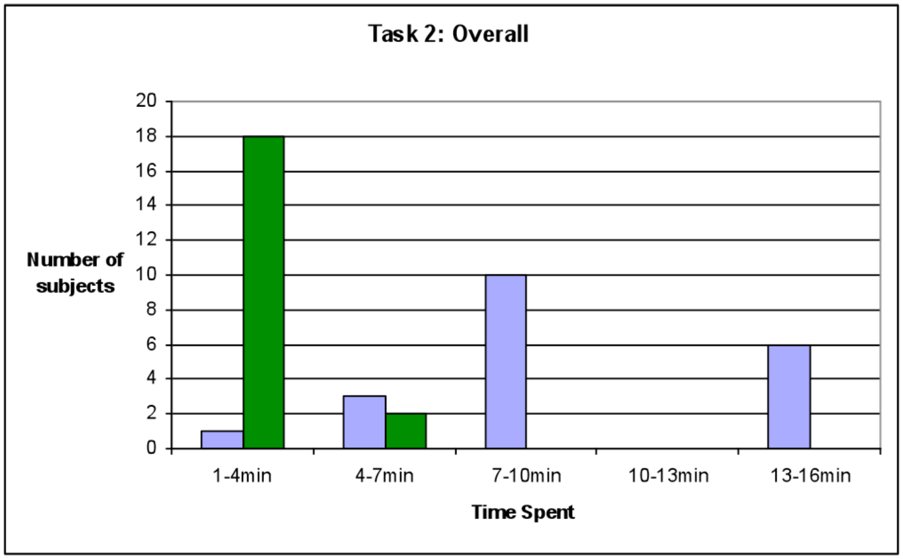
- Dapper/Pipes: 30% fail in the union case and 95% fail in the join case

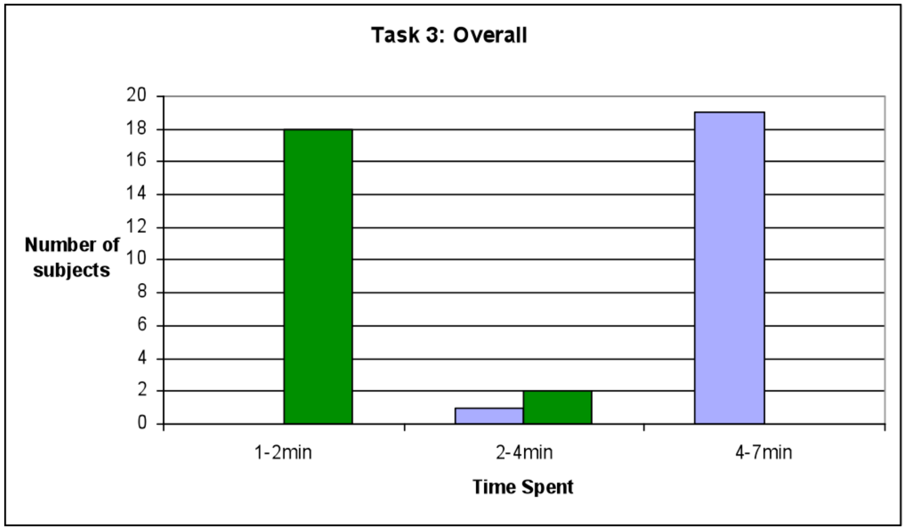
🟦 **Dapper/Pipes** 🟩 **Karma**

Claim 3: Overall, the user takes less time to build the same Mashup in Karma compared to Dapper/Pipes

Evaluation: Overall

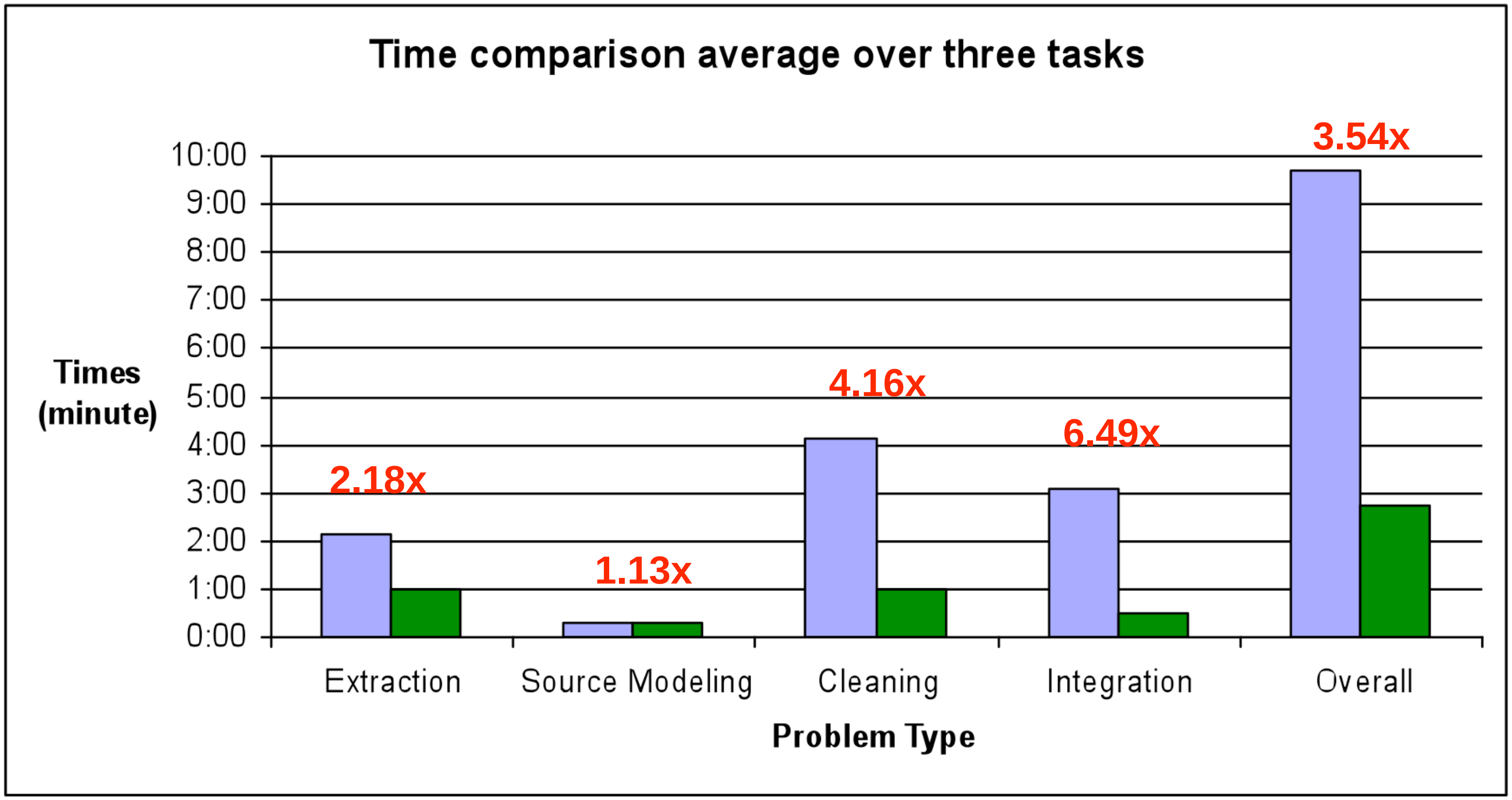

Task 1: Overall


Task 2: Overall


Task 3: Overall

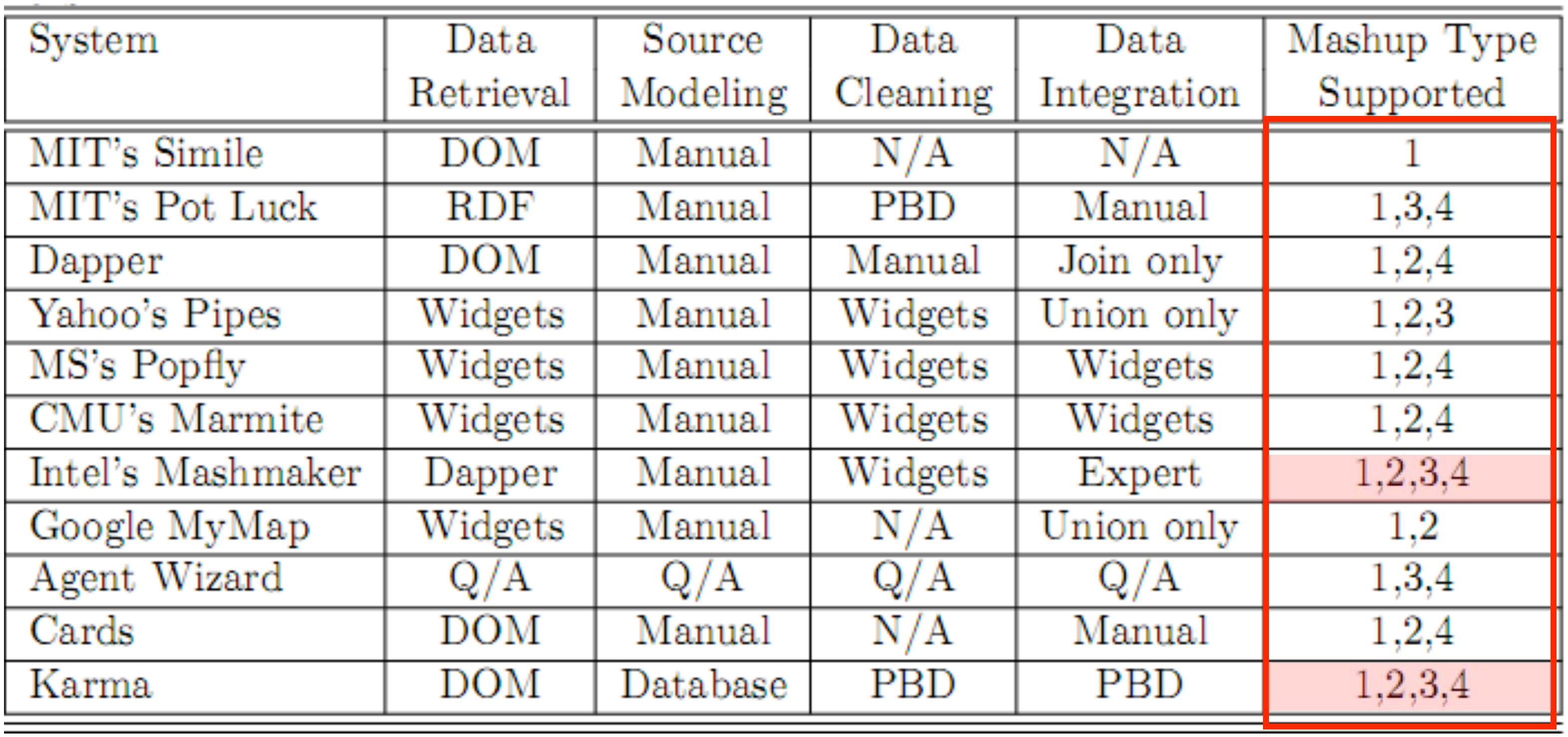
🟦 **Dapper/Pipes** 🟩 **Karma**

Evaluation: Average



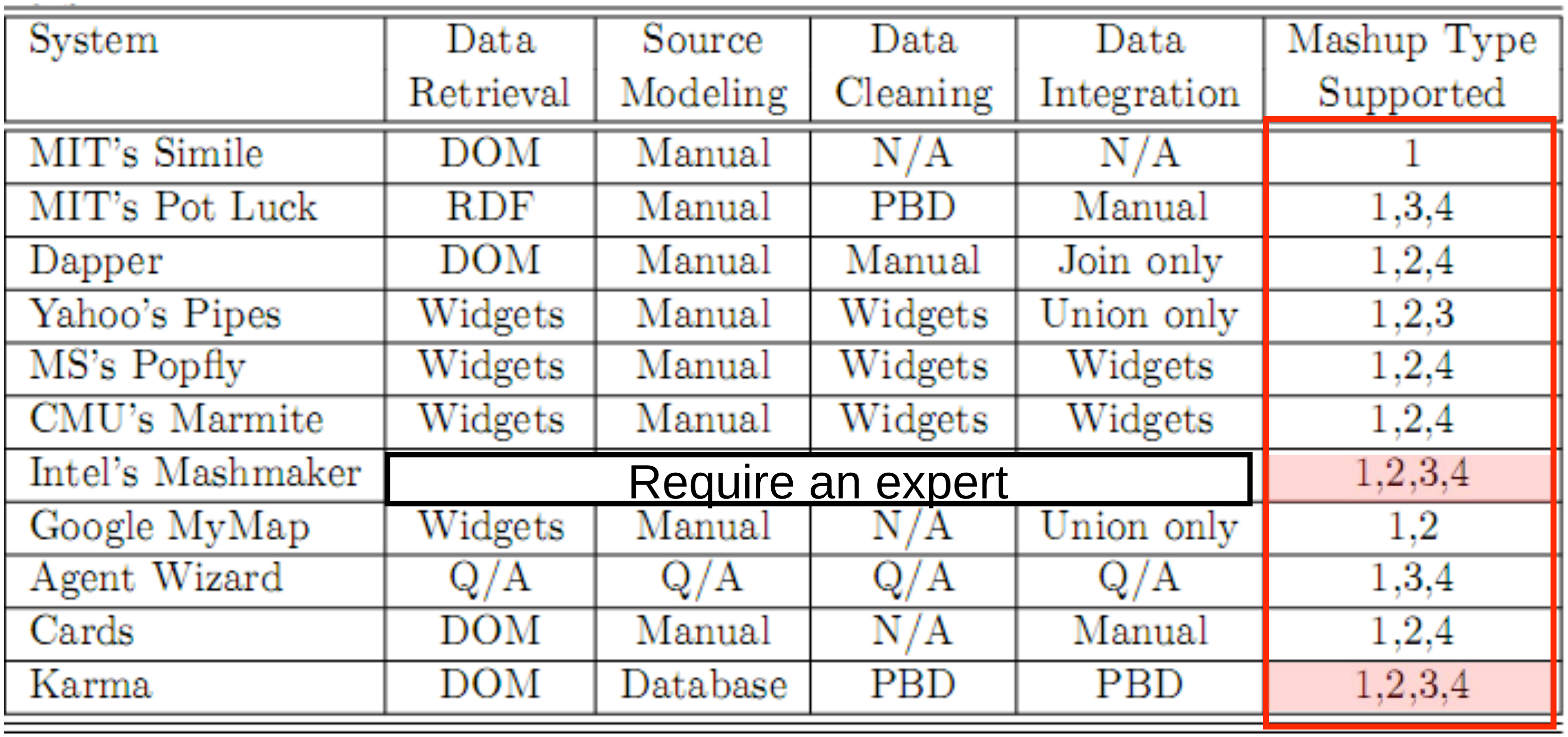
🟦 **Dapper/Pipes** 🟩 **Karma**

Related Work: Mashup Systems



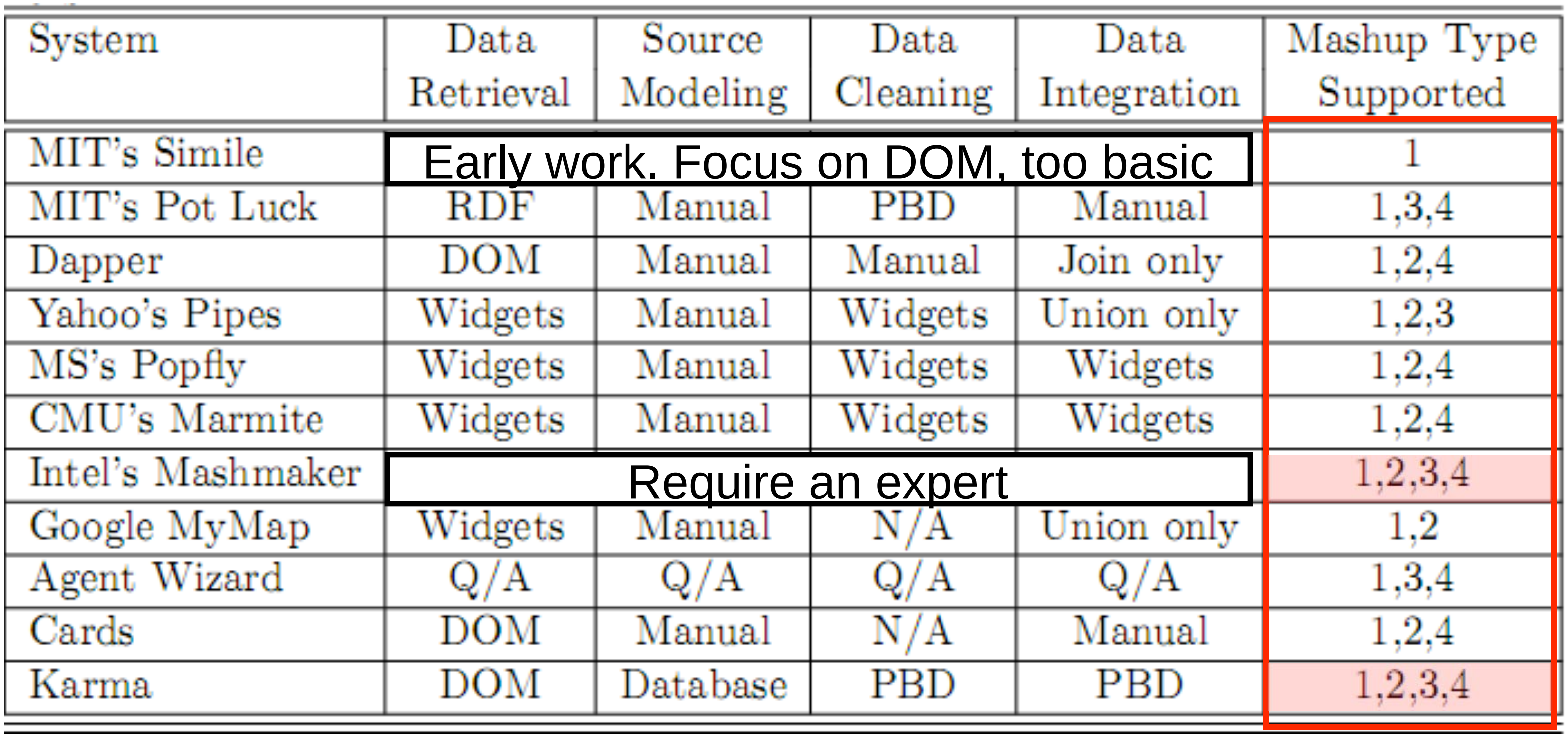
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	DOM	Manual	N/A	N/A	1
MIT's Pot Luck	RDF	Manual	PBD	Manual	1,3,4
Dapper	DOM	Manual	Manual	Join only	1,2,4
Yahoo's Pipes	Widgets	Manual	Widgets	Union only	1,2,3
MS's Popfly	Widgets	Manual	Widgets	Widgets	1,2,4
CMU's Marmite	Widgets	Manual	Widgets	Widgets	1,2,4
Intel's Mashmaker	Dapper	Manual	Widgets	Expert	1,2,3,4
Google MyMap	Widgets	Manual	N/A	Union only	1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



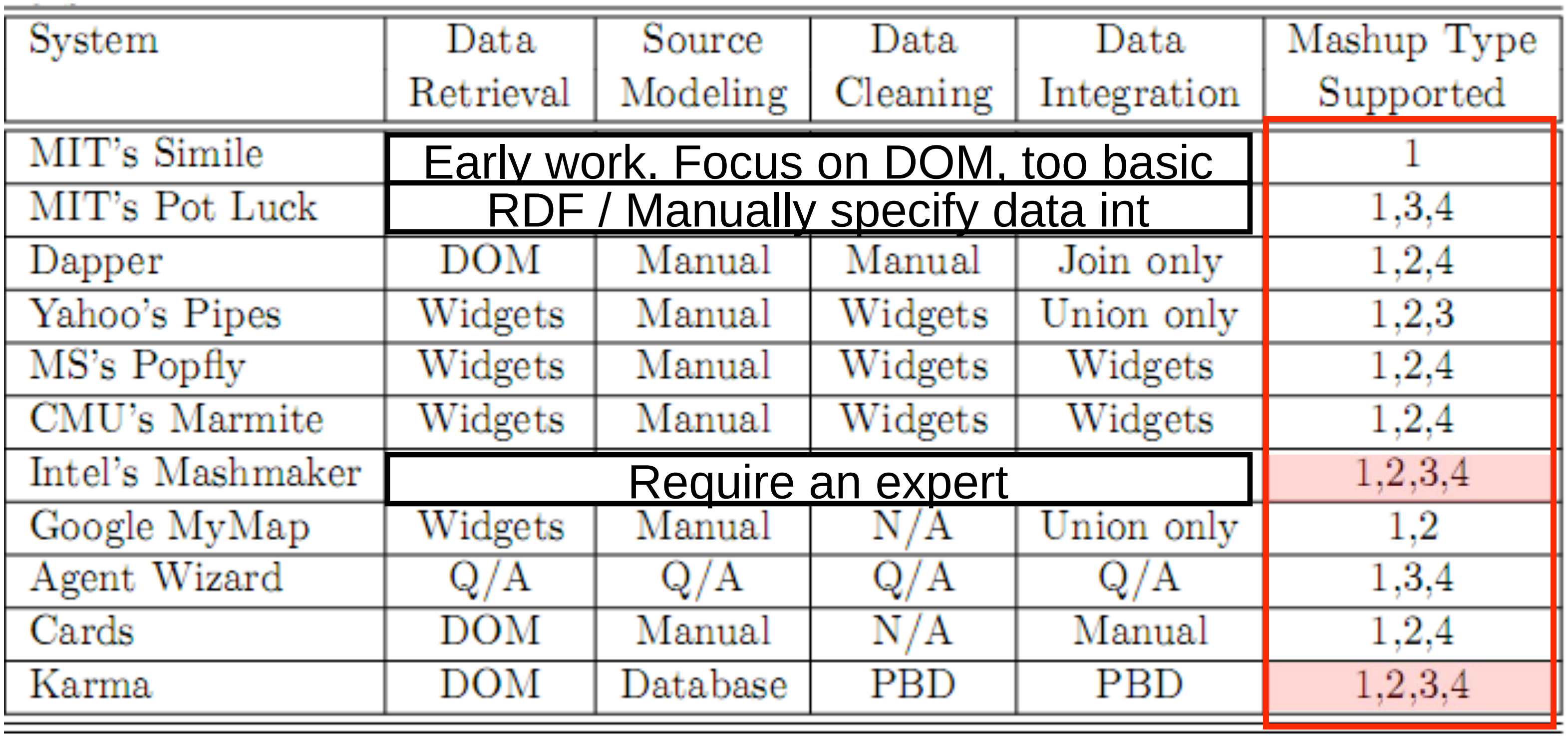
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	DOM	Manual	N/A	N/A	1
MIT's Pot Luck	RDF	Manual	PBD	Manual	1,3,4
Dapper	DOM	Manual	Manual	Join only	1,2,4
Yahoo's Pipes	Widgets	Manual	Widgets	Union only	1,2,3
MS's Popfly	Widgets	Manual	Widgets	Widgets	1,2,4
CMU's Marmite	Widgets	Manual	Widgets	Widgets	1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Widgets	Manual	N/A	Union only	1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



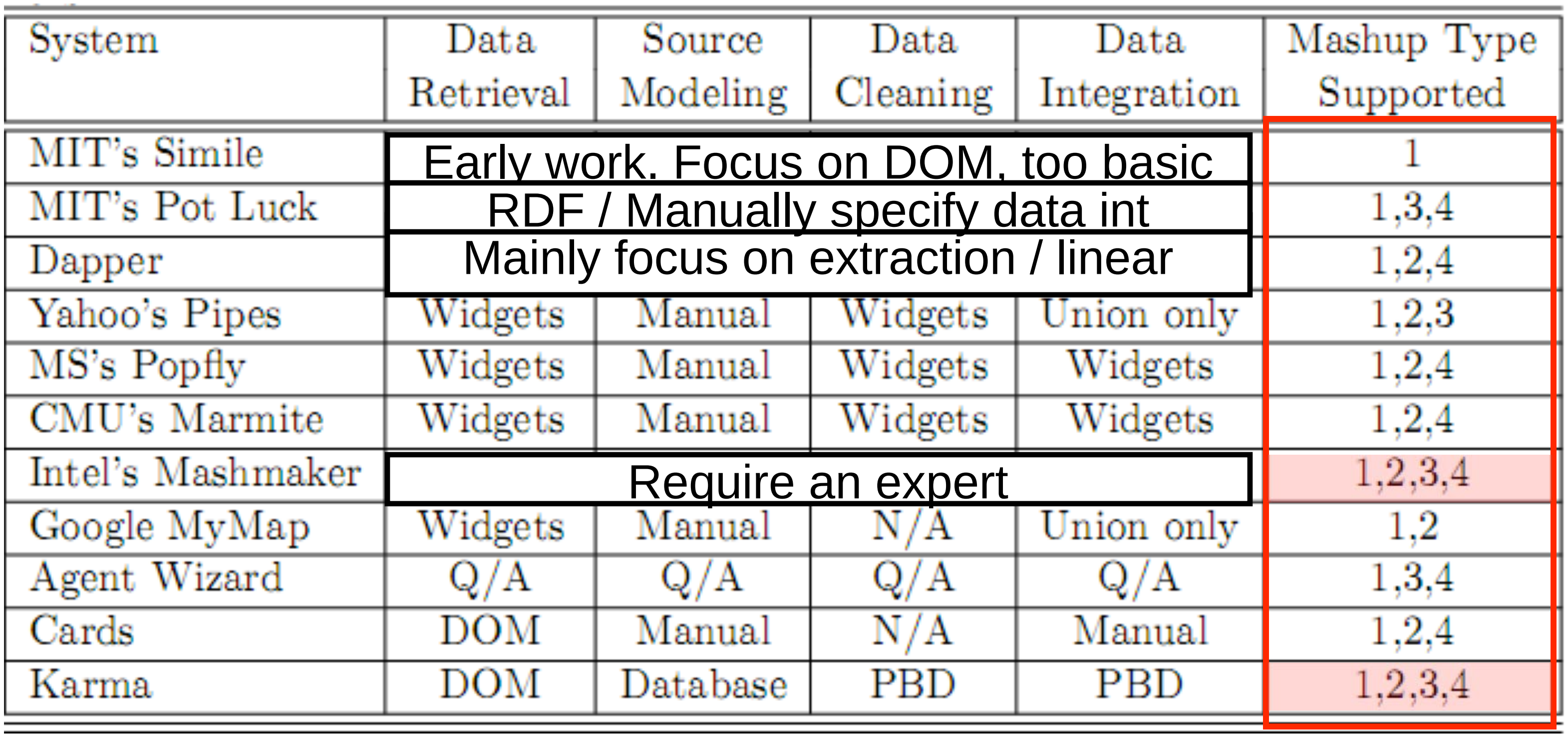
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF	Manual	PBD	Manual	1,3,4
Dapper	DOM	Manual	Manual	Join only	1,2,4
Yahoo's Pipes	Widgets	Manual	Widgets	Union only	1,2,3
MS's Popfly	Widgets	Manual	Widgets	Widgets	1,2,4
CMU's Marmite	Widgets	Manual	Widgets	Widgets	1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Widgets	Manual	N/A	Union only	1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



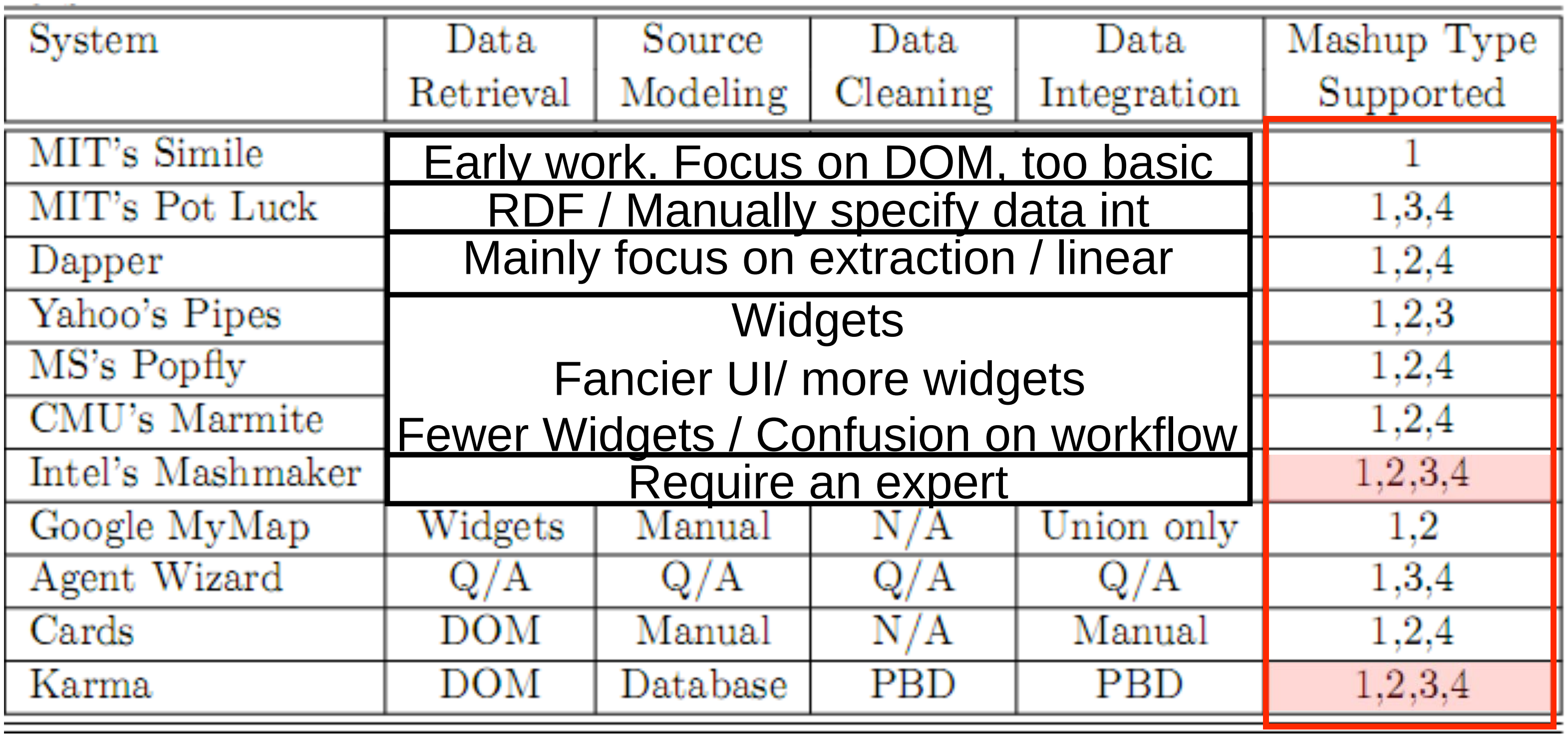
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF / Manually specify data int				1,3,4
Dapper	DOM	Manual	Manual	Join only	1,2,4
Yahoo's Pipes	Widgets	Manual	Widgets	Union only	1,2,3
MS's Popfly	Widgets	Manual	Widgets	Widgets	1,2,4
CMU's Marmite	Widgets	Manual	Widgets	Widgets	1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Widgets	Manual	N/A	Union only	1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



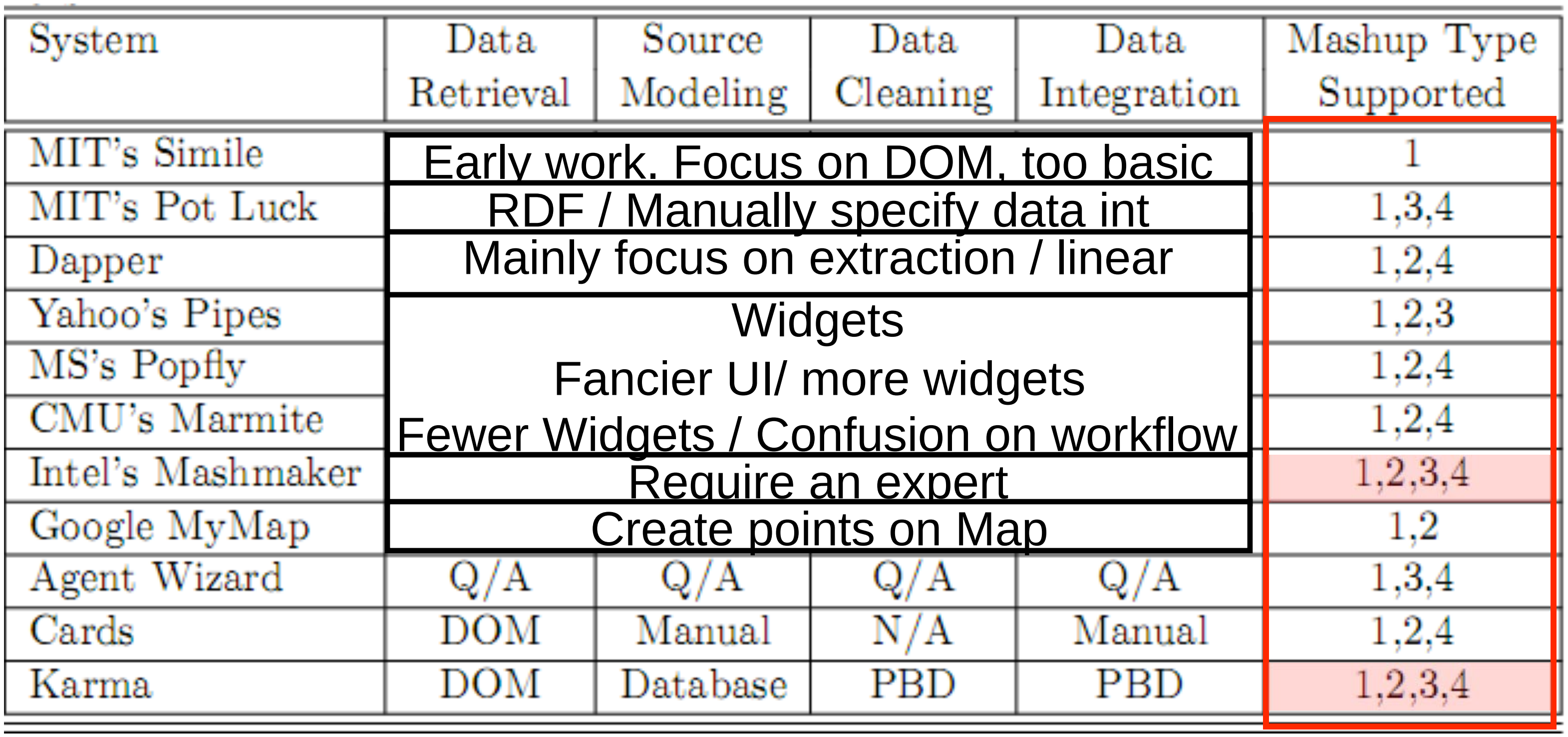
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF / Manually specify data int				1,3,4
Dapper	Mainly focus on extraction / linear				1,2,4
Yahoo's Pipes	Widgets	Manual	Widgets	Union only	1,2,3
MS's Popfly	Widgets	Manual	Widgets	Widgets	1,2,4
CMU's Marmite	Widgets	Manual	Widgets	Widgets	1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Widgets	Manual	N/A	Union only	1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



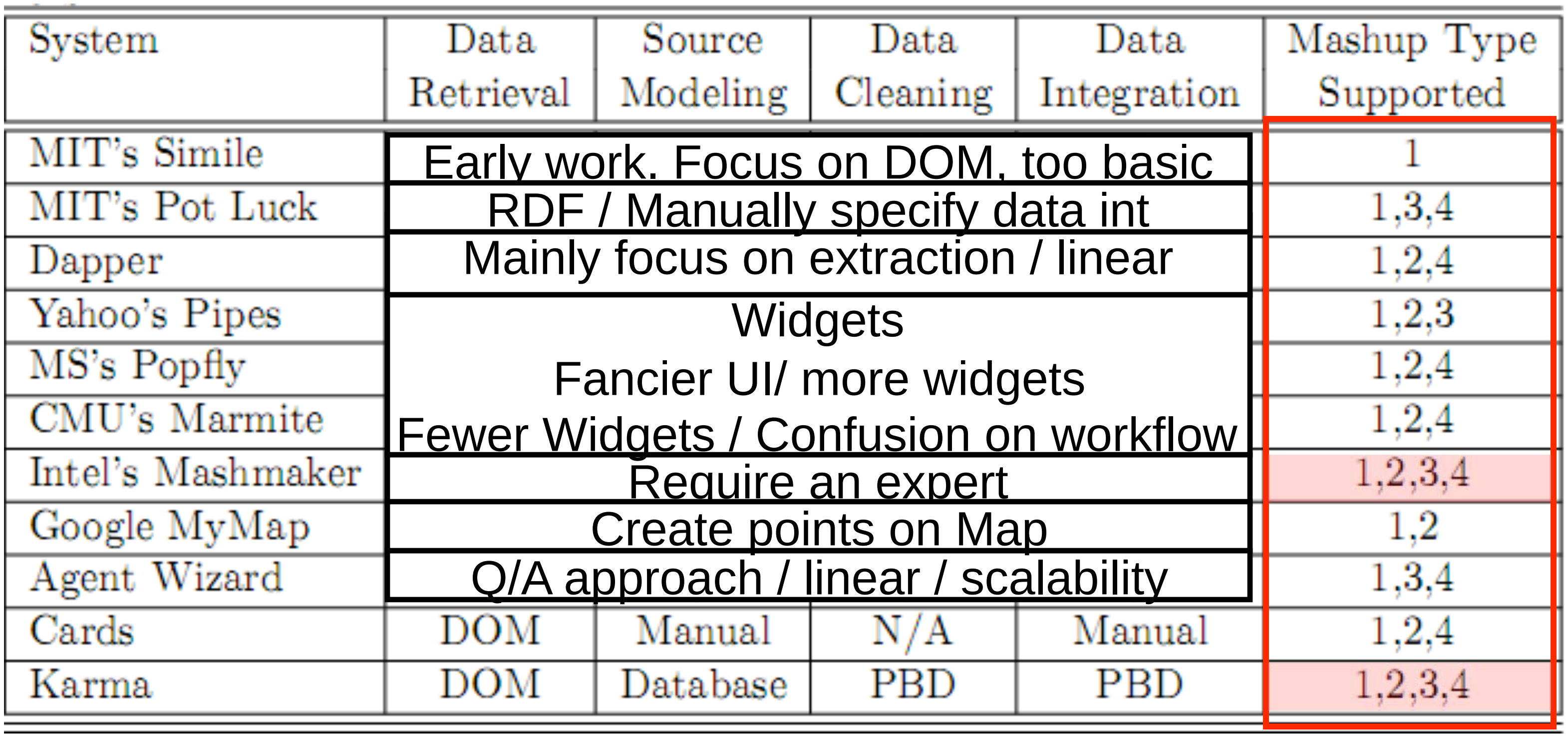
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF / Manually specify data int				1,3,4
Dapper	Mainly focus on extraction / linear				1,2,4
Yahoo's Pipes	Widgets				1,2,3
MS's Popfly	Fancier UI/ more widgets				1,2,4
CMU's Marmite	Fewer Widgets / Confusion on workflow				1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Widgets	Manual	N/A	Union only	1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



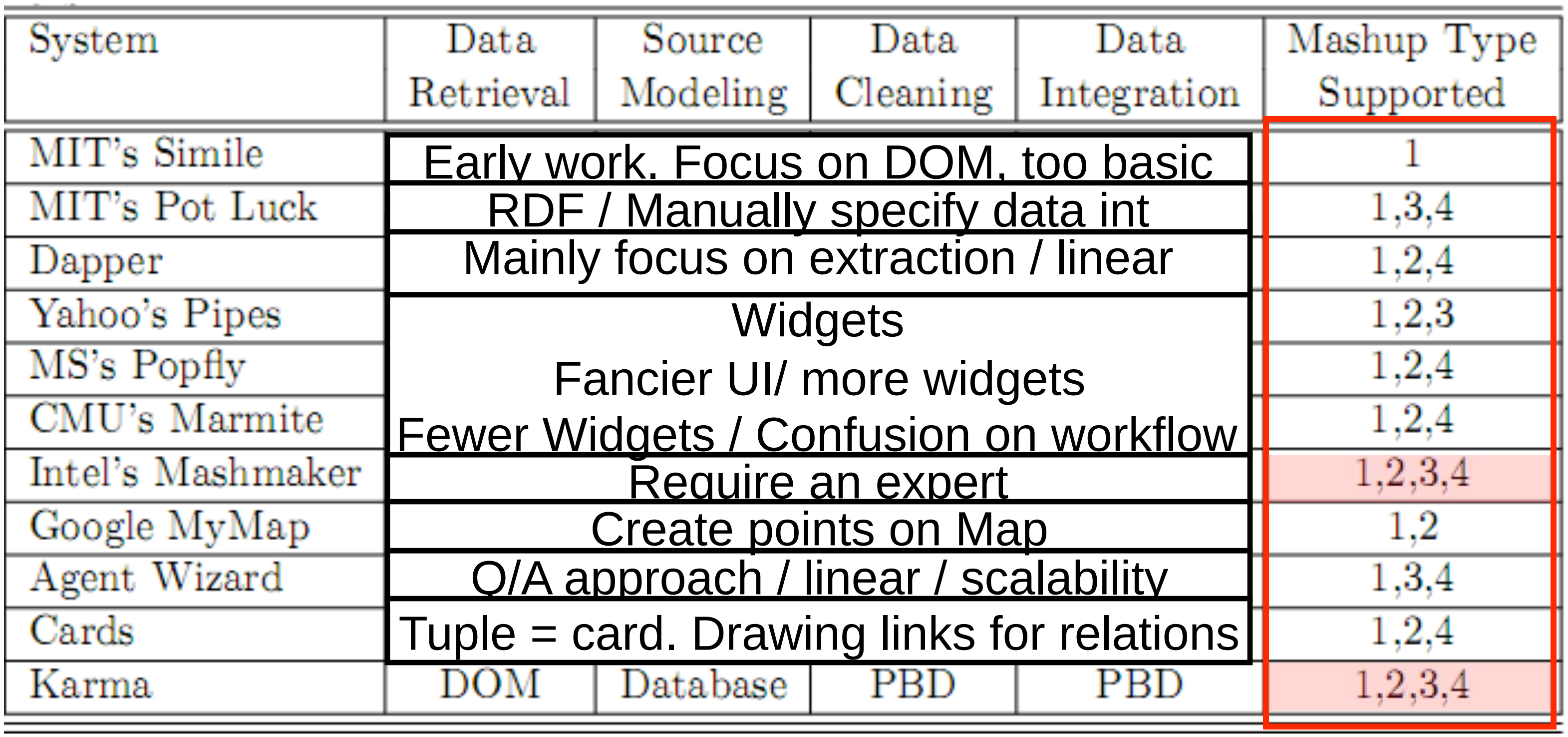
System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF / Manually specify data int				1,3,4
Dapper	Mainly focus on extraction / linear				1,2,4
Yahoo's Pipes	Widgets				1,2,3
MS's Popfly	Fancier UI/ more widgets				1,2,4
CMU's Marmite	Fewer Widgets / Confusion on workflow				1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Create points on Map				1,2
Agent Wizard	Q/A	Q/A	Q/A	Q/A	1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF / Manually specify data int				1,3,4
Dapper	Mainly focus on extraction / linear				1,2,4
Yahoo's Pipes	Widgets				1,2,3
MS's Popfly	Fancier UI/ more widgets				1,2,4
CMU's Marmite	Fewer Widgets / Confusion on workflow				1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Create points on Map				1,2
Agent Wizard	Q/A approach / linear / scalability				1,3,4
Cards	DOM	Manual	N/A	Manual	1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Mashup Systems



System	Data Retrieval	Source Modeling	Data Cleaning	Data Integration	Mashup Type Supported
MIT's Simile	Early work. Focus on DOM, too basic				1
MIT's Pot Luck	RDF / Manually specify data int				1,3,4
Dapper	Mainly focus on extraction / linear				1,2,4
Yahoo's Pipes	Widgets				1,2,3
MS's Popfly	Fancier UI/ more widgets				1,2,4
CMU's Marmite	Fewer Widgets / Confusion on workflow				1,2,4
Intel's Mashmaker	Require an expert				1,2,3,4
Google MyMap	Create points on Map				1,2
Agent Wizard	Q/A approach / linear / scalability				1,3,4
Cards	Tuple = card. Drawing links for relations				1,2,4
Karma	DOM	Database	PBD	PBD	1,2,3,4

Related Work: Data Extraction

- Automatic extraction: table and lists only
  - RoadRunner (exploit HTML structure) [Crescenzi et al., 2001]
  - Adel (grammer induction to detect rows) [Lerman+ 2001]
  - VisualWeb (OCR technique to detect tables) [Gatterbauer+ 2007]
- Semi-Automatic: require more label examples
  - WIEN (inductive – less expressive than stalker) [Kushmerick 1997]
  - Stalker (Cotesting) [Muslea+ 1999]
  - SoftMealy (finite state transducer) [Hsu 1998]
  - WHISK (rigid format, exact delimiter) [Soderland 1998]
- DOM: rely on well-formed HTML and less labeling
  - Simile [Huynh+ 2005]
  - Dapper
  - Interactive Wrapper Generation (ML + prediction on DOM) [Irmak+ 2006]
  - PLOW (add natural language) [Allen+ 2007]
  - Cards [Dontcheva+ 2007]
  - Karma [Tuchinda+ 2008]

Related Work: Source Modeling

- 1:1 mapping, N:M mapping
  - Schema-level match
    - TranScm [Milo+ 98]
    - DIKE [Palopoli+ 99]
    - Artemis [Castano+ 01]
    - Delta [Clifton+ 97]
  - +Instance-based matcher
    - SemInt [Li 00]
    - LSD [Doan 01]
    - ILA [Etzioni 95]
    - iMapp [Dhamanka 04]
    - Clio (interactive) [Ling 01]
    - Inducing Source Description [Carman 07]
- **Karma leverages existing techniques to narrow candidate matches**
  - String Similarities [Cohen+ 2003]

Related Work: Data Cleaning

- Commercial Tools: Focus on writing transformation
  - ACR/Data, Migration Architect [Chaudhuri+ 1997]
- Discrepancy Detection: Use as a stepping stone for record linkage and cleaning system
  - Levenshtein distance [Needleman+ 70]
  - Vector based [Baeza-Yates+ 99]
  - EM [Ristad+ 98]
  - SVM [Bilenko+ 03]
- Record linkage & cleaning systems: Focus on ranking [Winkler 06]
  - Fuzzy Match [Chaudhuri+ 03]
  - Apollo [Michalowski+ 05]
  - Phoebus [Michelson+ 07]
  - Potter's wheel [Raman+ 01]
- Karma
  - Gained reference sources through source modeling process
  - Provided predefined transformations

Related Work: Data Integration

- Universal Relation: Make it easier to formulate the query but users still need to formulate the query [Ullman 1980, 1988]
- Query by example: Need to know which data sources to use and the query may not return results
  - QBE [Zloof 1975]
- Retrieval by formulation: Need to understand domain model to formulate partial description
  - Helgon [Fischer 1989]
  - RABBIT [Williams 1982]
- Graphical Query Language: Users still need to navigate through sources (graphs)
  - Gql [Benzi 1998, Haw 1994, Papantonakis 1988]
- Question-Answering Technique: Understanding about database operations required.
  - Agent Wizard [Tuchinda+ 2004]
- Interactive Schema/data integration: Understanding about source schema required
  - Clio [Ling 01]
- Karma is based on Programming by Demonstration [Cyper 2001; Lau2001]

Conclusion

- Mashup is a fast growing area
  - Need an efficient way to for casual web users to build it.
- Contributions
  - A PBD approach that uses a single table for building a Mashup
  - An integrated approach that links four Mashup buildling issues.
  - A query formulation technique that allows users to specify examples to build complicated queries.
- Evaluated the validity of the Karma approach
  - Subjects were able to complete Mashup building tasks in Karma
  - The overall improvement is at least 3.5

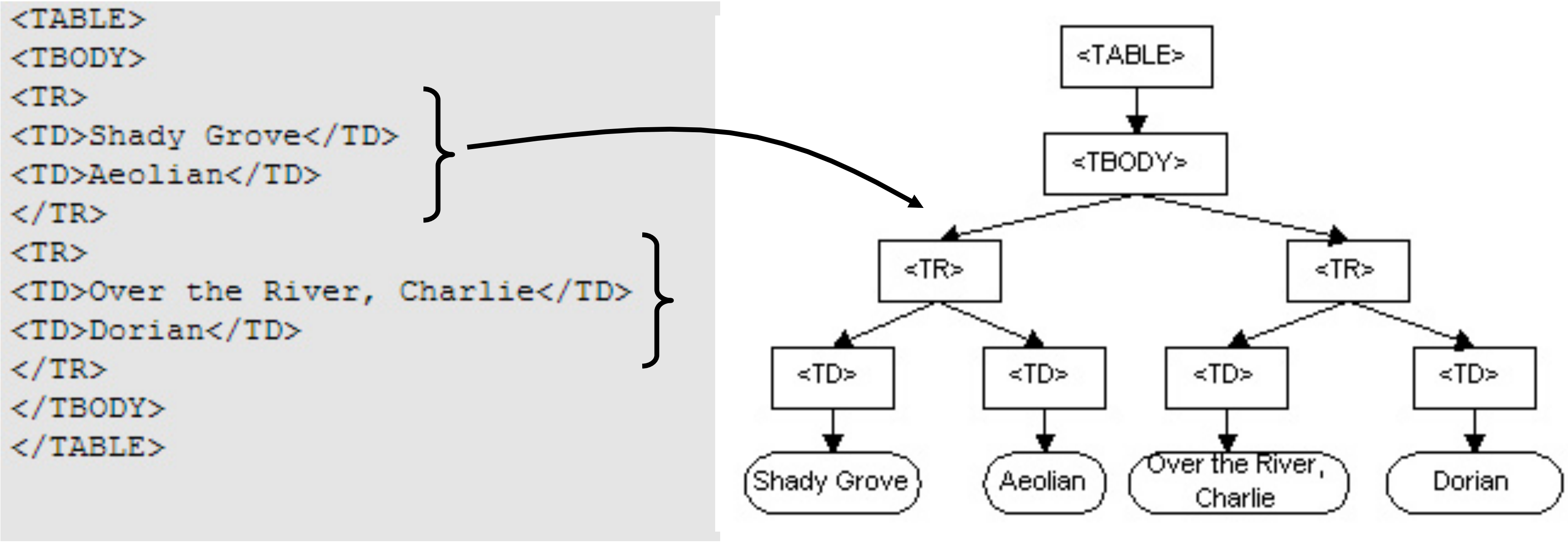
Future Work

- Customizing display by examples
- Implementing feedback and quality
- Adding planning components to handle dynamic data.

Thank You!

Backup Slides

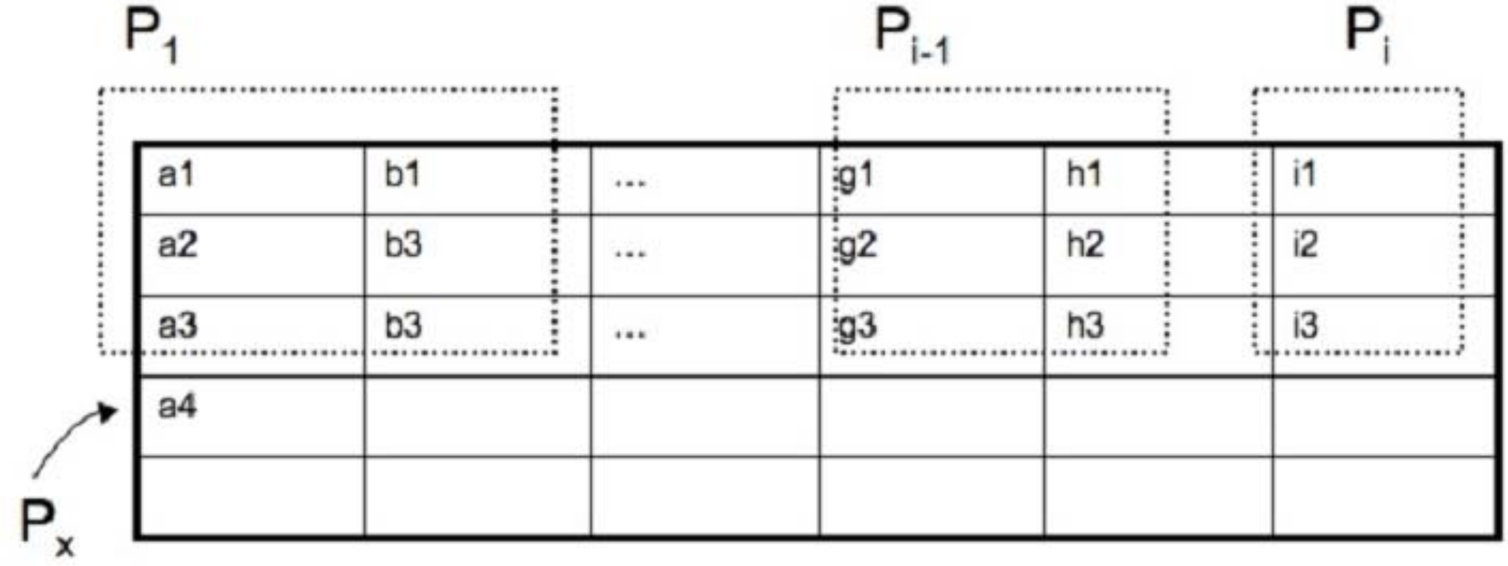
Document Object Model (DOM)



```
<TABLE>
<TBODY>
<TR>
<TD>Shady Grove</TD>
<TD>Aeolian</TD>
</TR>
<TR>
<TD>Over the River, Charlie</TD>
<TD>Dorian</TD>
</TR>
</TBODY>
</TABLE>
```

Source: http://www.w3.org

Vertical Expansion



a1	b1	...	g1	h1	i1
a2	b3	...	g2	h2	i2
a3	b3	...	g3	h3	i3
a4					

Table 8.4: Normalized data for task 1. E stands for data extraction, M stands for source modeling, and C stands for data cleaning. The asterisk indicates time substitution when failures happen. The data is reported in minutes.

Task1	Dapper/Pipes				Karma			
Subject	E	M	C	Total	E	M	C	Total
No.1	*5:00	0:20	*5:00	10:20	2:19	1:08	1:00	4:27
No.2	1:43	0:30	*5:00	7:13	1:00	0:40	0:29	2:09
No.3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
No.4	0:52	0:48	*5:00	6:40	1:12	1:00	0:50	3:02
No.5	5:00	0:35	*5:00	10:35	1:15	1:18	1:20	3:53
No.6	2:30	0:15	*5:00	7:45	1:00	0:54	0:28	2:22
No.7	1:20	0:22	*5:00	6:42	0:51	0:51	0:46	2:28
No.8	1:40	0:14	*5:00	6:54	1:04	0:41	0:33	2:19
No.9	1:26	0:16	*5:00	6:42	1:14	1:00	1:10	3:24
No.10	1:39	0:10	*5:00	6:49	0:53	0:42	0:50	2:26
No.11	2:00	0:19	*5:00	7:19	1:04	1:00	0:53	2:57
No.12	2:00	0:49	2:00	4:49	1:07	1:00	0:40	2:47
No.13	2:00	0:05	*5:00	7:05	0:58	0:50	0:56	1:44
No.14	2:46	0:15	*5:00	8:01	1:12	0:45	0:48	2:45
No.15	2:27	0:14	3:11	5:52	1:10	0:49	1:20	3:19
No.16	1:16	0:12	*5:00	6:28	0:58	0:42	0:25	2:05
No.17	n/a	n/a	n/a	n/a	2:00	1:00	0:50	3:50
No.18	2:30	0:14	*5:00	7:44	1:06	1:10	1:46	4:02
No.19	1:38	0:47	1:20	3:45	1:20	0:49	0:35	2:44
No.20	1:30	0:16	*5:00	6:46	1:04	0:44	0:35	2:23

Table 8.5: Normalized data for task 2. E stands for data extraction, M stands for source modeling, C stands for data cleaning, and I stands for data integration. The asterisk indicates time substitution when failures happen. The data is reported in minutes.

Task2	Dapper/Pipes					Karma				
Subject	E	M	C	I	Total	E	M	C	I	Total
No.1	4:38	0:22	2:45	1:15	9:00	1:26	0:43	0:43	0:00	2:52
No.2	1:35	0:12	3:30	0:12	5:29	0:50	0:57	0:57	0:00	2:44
No.3	*5:00	0:25	*5:00	*5:00	15:25	2:52	1:00	3:00	0:00	5:52
No.4	4:49	0:17	3:29	0:38	9:14	1:26	0:48	1:03	0:00	3:18
No.5	*5:00	0:29	1:44	1:16	8:29	1:43	0:45	1:20	0:00	3:48
No.6	*5:00	0:20	*5:00	*5:00	15:20	2:07	0:30	0:50	0:00	3:27
No.7	2:17	0:15	4:46	0:18	7:36	1:13	0:25	0:52	0:00	2:31
No.8	3:23	0:21	*5:00	*5:00	13:44	1:10	0:21	0:24	0:00	1:55
No.9	4:11	0:21	*5:00	*5:00	14:32	1:22	0:47	2:11	0:00	4:20
No.10	2:16	0:07	3:14	0:20	5:50	1:04	0:20	1:06	0:00	2:30
No.11	3:04	0:17	*5:00	*5:00	13:21	1:06	0:34	0:53	0:00	2:33
No.12	2:00	0:27	*5:00	0:20	7:47	1:23	0:30	0:37	0:00	2:30
No.13	*5:00	0:07	1:43	0:10	7:00	1:42	0:32	0:41	0:00	2:55
No.14	3:03	0:23	4:42	0:10	8:21	1:40	0:31	0:56	0:00	3:07
No.15	2:06	0:12	3:13	0:22	5:53	1:30	0:24	2:05	0:00	3:59
No.16	3:58	0:11	3:29	0:27	8:05	0:51	0:17	1:00	0:00	2:08
No.17	4:15	0:28	3:39	0:30	8:52	1:04	0:28	1:18	0:00	2:50
No.18	*5:00	0:23	*5:00	*5:00	15:23	1:17	0:30	1:10	0:00	2:57
No.19	4:01	0:14	2:42	0:21	7:16	1:39	0:21	0:50	0:00	2:50
No.20	1:36	0:43	0:36	0:22	3:17	1:07	0:28	0:40	0:00	2:15

Table 8.6: Normalized data for task 3. E stands for data extraction, M stands for source modeling, and I stands for data integration. The asterisk indicates time substitution when failures happen. The data is reported in minutes.

Task3	Dapper/Pipes				Karma			
Subject	E	M	I	Total	E	M	I	Total
No.1	1:30	0:26	*5:00	6:56	0:14	0:00	2:16	2:30
No.2	0:30	0:10	*5:00	5:40	0:25	0:00	0:26	0:54
No.3	1:00	0:15	*5:00	6:15	0:15	0:00	0:44	0:59
No.4	0:40	0:16	*5:00	5:56	0:20	0:00	1:06	1:26
No.5	0:40	0:14	*5:00	5:54	0:20	0:00	0:37	0:57
No.6	0:30	0:10	*5:00	5:40	0:20	0:00	0:31	0:51
No.7	0:27	0:10	*5:00	5:37	0:14	0:00	0:50	1:04
No.8	0:29	0:20	*5:00	5:49	0:30	0:00	0:51	1:21
No.9	0:40	0:23	*5:00	6:03	0:13	0:00	0:44	0:57
No.10	0:30	0:10	*5:00	5:40	0:20	0:00	0:35	0:55
No.11	0:51	0:20	*5:00	6:11	0:16	0:00	1:05	1:21
No.12	1:05	0:18	*5:00	6:23	0:30	0:00	0:46	1:16
No.13	0:31	0:14	*5:00	5:45	0:16	0:00	0:57	1:13
No.14	0:36	0:14	*5:00	5:50	0:14	0:00	2:00	2:14
No.15	0:26	0:21	*5:00	5:47	0:30	0:00	0:45	1:15
No.16	0:27	0:13	*5:00	5:40	0:15	0:00	0:56	1:11
No.17	0:33	0:38	1:56	3:07	0:30	0:00	0:46	1:16
No.18	1:03	0:07	*5:00	6:10	0:20	0:00	1:10	1:30
No.19	0:33	0:17	*5:00	5:50	0:25	0:00	1:20	1:45
No.20	0:18	0:13	*5:00	5:31	0:12	0:00	0:44	0:56