

Building Data Integration Queries by Demonstration

Rattapoom Tuchinda, Pedro Szekely,
and Craig A. Knoblock

Information Science Institute
University of Southern California

Intelligent User Interface. Jan 30, 2007

Vision and Motivation

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 [GlobeExplorer](#) and Suppliers | © 2007 Proxim

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

Zillow.com

Los Angeles County Office of the Assessor

Property Information

Assessor's ID Number	4313-005-033
Site Address	3767 CLARINGTON AVE LOS ANGELES CA 90034
Property Type	Multi-Family Residential
Region/Cluster	09 09A31
Tax Rate Area (TRA)	00067

[Click Here to View Assessor's Map](#)
[Click Here to View Index Map](#)

Recent Sale Information

Latest Sale Date	
Indicated Sale Price	

[Search for Recent Sales](#)

2006 Roll Values

Recording Date	05/07/1982
Land	\$1,792,303
Improvements	\$8,704,658
Personal Property	\$61,950
Fixtures	\$0
Homeowners' Exemption	\$0
Real Estate Exemption	\$0
Personal Property Exemption	\$0
Fixture Exemption	\$0

Tax properties

38.2347 How many digits of precision do we need?

Address 3767 Clarington Ave
Los Angeles CA 90034
(34.023086, -118.401261)

Latitude 34.023086 °
N 34 ° 1' 23.1"
34 ° 1.3852' (degree
m.mmmm)

Longitude -118.401261 °
W 118 ° 24' 4.5"
-118 ° 24.0757' (degree
m.mmmm)

Search for another address:

3767 clarington ave, Los Angeles, CA 90034

Geocoder

Map interface showing a list of search results on the left and a map view on the right. The list includes items like 'yln: mtrhw bloc', 'yln: mtrhw bloc', etc. The map shows a street grid with several red markers indicating property locations.

REALTOR.com Official Site of the National Association of REALTORS®

Find a Home ▾ Rentals ▾ Home Finance ▾ Moving ▾ Home & Garden ▾

Welcome, Visitor

Sign Up to:
Save Searches
Save Listings
Sign Up Now!
Already a member?
Sign In

Homebuying Tools
Find a Lender
Find a Mover
Market Conditions
Neighborhood Tour

Find a Home
Over 3 million properties for sale!

- Enter a City - State/Province - OR - Zip/Postal Code -

Minimum Price to Maximum Price

Beds Baths [More Search Options](#)
[Map Search](#)

What's Your Home Worth?

Realtor

RAND OBJECTIVE ANALYSIS. EFFECTIVE SOLUTIONS.

A non-profit institution that addresses the challenges facing the public and private sectors around the world.

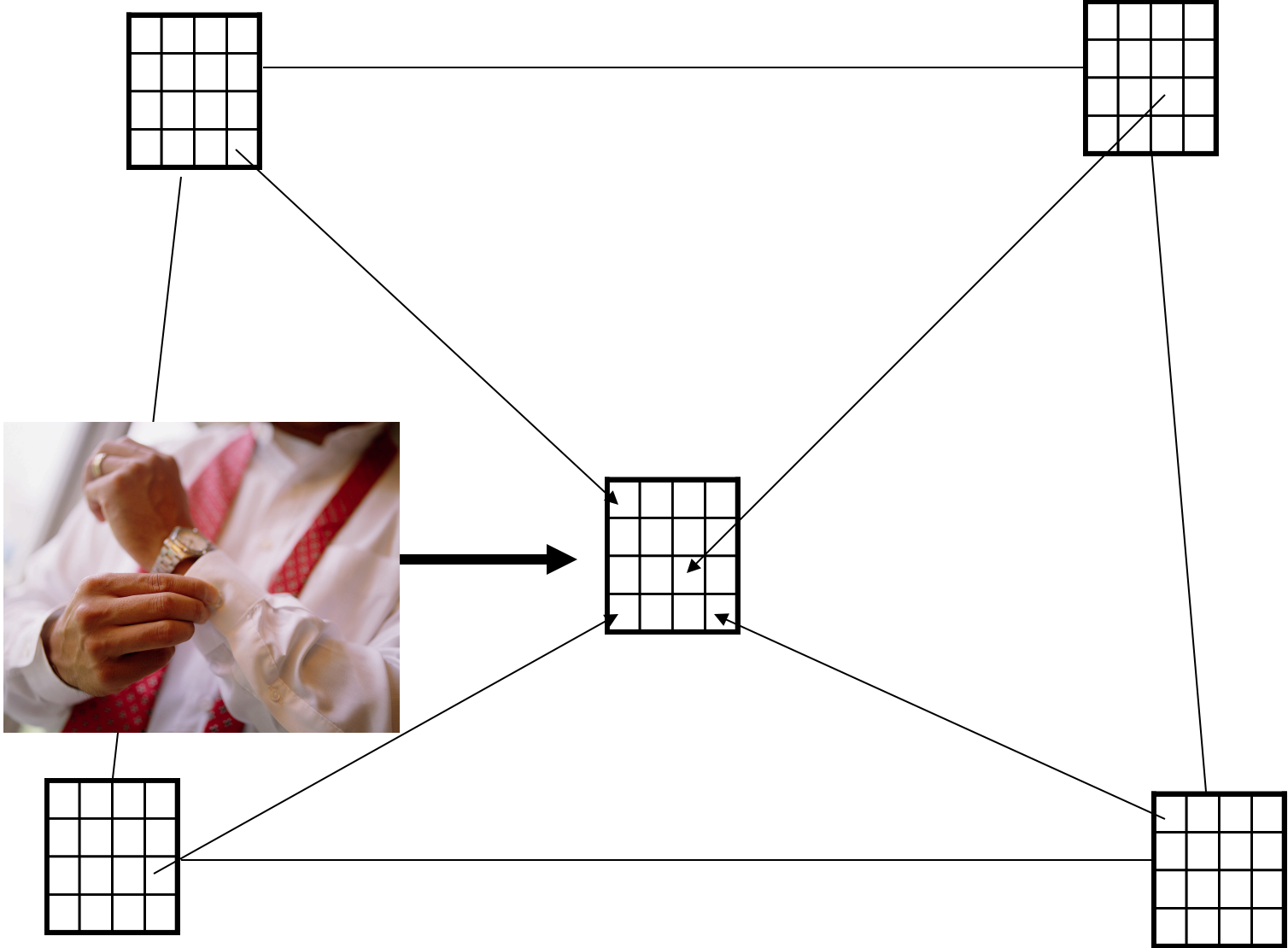
For DONORS
For CONGRESS
For NEWS MEDIA
For JOB SEEKERS
For EXISTING CLIENTS
For PROSPECTIVE CLIENTS

Delivering Quality and Objectivity

For nearly 60 years, the RAND name has been synonymous with high-quality, objective research and analysis.

RAND's quality standards emphasize the importance of RAND's core values of quality and objectivity in all aspects of our work.

And many more..



Why is it hard?

Web Source Characteristics:

1. The number of sources is huge
2. Overlapping data between sources

As a result, it's difficult to write sql queries.

User Characteristics:

1. Don't know how to program
2. Don't always know what sources are available
3. Do know partial "data" (data value) that they want but may not know the "semantic" (attribute). E.g. hyatt (hotelname), waikiki (city)

Our Approach



Our Approach

Web [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

when will i grow

when will i grow	41,200,000 results
when will i graduate	44,600,000 results
when will i grow facial hair	
when will i grow chest hair	
when will i grow up	

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2006 Google



Our Approach

Web [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

when will i grow

when will i grow	41,200,000 results
when will i graduate	44,500,000 results
when will i grow facial hair	
when will i grow chest hair	
when will i grow up	

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2006 Google

Intuition

Crime rate

City	Crime Rate
Detroit	88%
NYC	60%
Los Angeles	45%

white pages

Name	Address	City	State
John Smith	500 Bundy	Los Angeles	CA
Mary Smith	100 Main	Long Beach	CA
Joe Smith	744 Temple	Ventura	CA

	Los Angeles	CA	

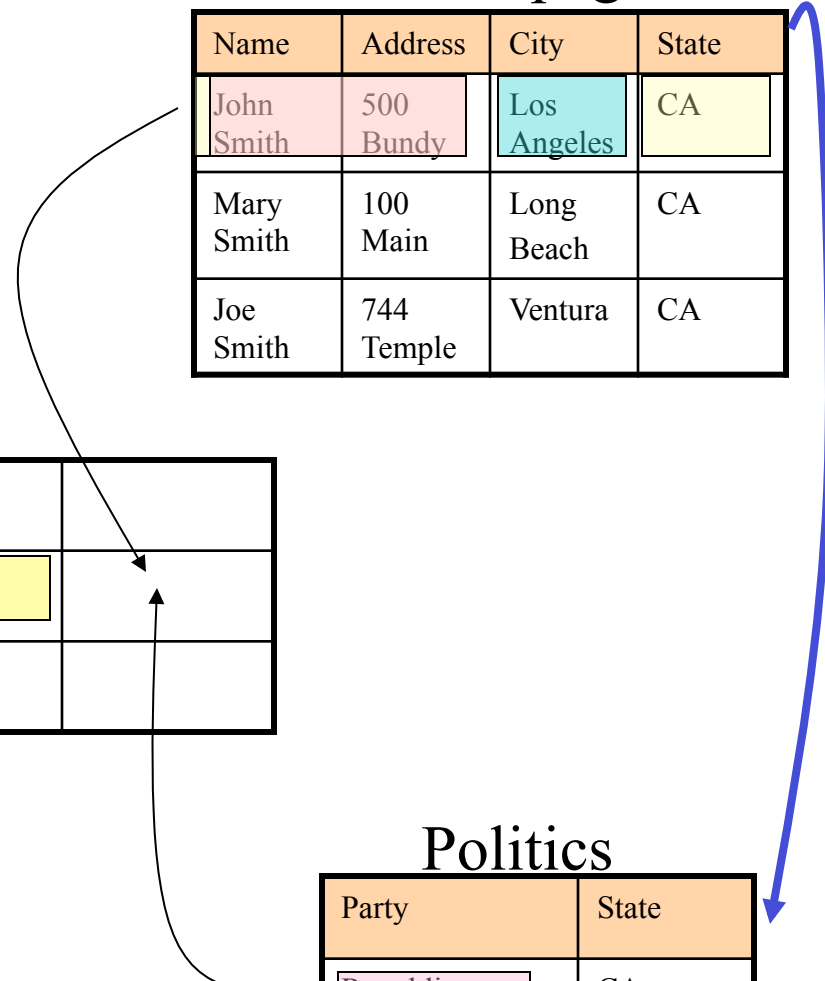
Hotels

Hotel name	Address	City	Phone
Hyatt	15 fiji	Salt Lake	801-534-4900
Radisson	10 Slauson	Los Angeles	310-666-6666
Marriott	102 Ames St.	Cambridge	617-353-1432

Politics

Party	State
Republican	CA
Democrat	MA
Republican	TX

Los



Single Column Example

City
Los Angeles
Honolulu

$a \in$ all attribute

$v \in$ all value

Los Angeles : $\{(\mathbf{city}, \text{tax_properties}),$
 ~~$(\text{song_name}, \text{pop_music})\}$~~

$v \in$ all values \wedge attributeOf(v) \in $\{\mathbf{city}, \text{song name}\}$

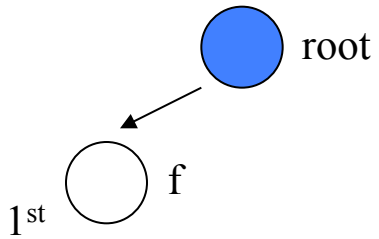
Honolulu : $\{(\mathbf{city}, \text{tax_properties}),$
 $(\mathbf{city}, \text{favorite vacation spot})\}$

Can we determine the attribute now? Yes

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

$\{v\} = \text{val}(a,s)$ where $a \in \{x\} \wedge s$ is any source where $\text{att}(s) \cap \{x\}$

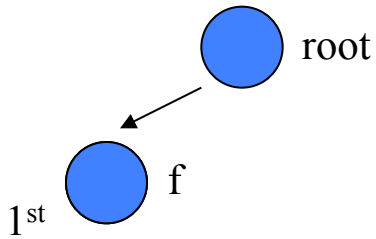
Query Transformation



	Los Angeles	

$f(a,s,v) = (?, ?, \text{Los Angeles})$

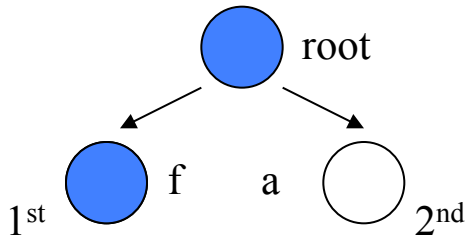
Query Transformation



City		
Los Angeles		
Honolulu		

$f(a,s,v) = (\text{city}, \text{tax_properties}, \text{Los Angeles})$

Query Transformation



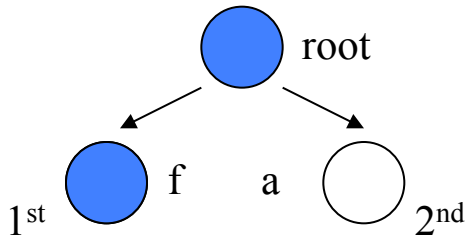
City	Zipcode	
Los Angeles		
Honolulu		

$f(a,s,v) = (?, ?, \text{Los Angeles})$

$f(a,s,v) = (\text{city}, \text{tax_properties}, \text{Los Angeles})$

$a(a,s,v) = (\text{zipcode}, \text{tax_properties}, \underline{\text{PLACE HOLDER}})$

Query Transformation



City	Zipcode	
Los Angeles		
Honolulu		

$f(a,s,v) = (?, ?, \text{Los Angeles})$

$f(a,s,v) = (\text{city}, \text{tax_properties}, \text{Los Angeles})$

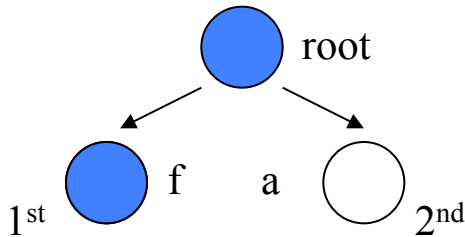
$a(a,s,v) = (\text{zipcode}, \text{tax_properties}, \underline{\text{PLACE HOLDER}})$

Select Zipcode

From tax_properties

Where City="Los Angeles"

Query Transformation



City	Zipcode	
Los Angeles		
Honolulu		

$f(a,s,v) = (?, ?, \text{Los Angeles})$

$f(a,s,v) = (\text{city}, \text{tax_properties}, \text{Los Angeles})$

$a(a,s,v) = (\text{zipcode}, \text{tax_properties}, \underline{\text{PLACE HOLDER}})$

Select Zipcode

From tax_properties

Where City="Los Angeles"

Karma

Back Forward Refresh Stop Compile/Browse

Google Web Toolkit BETA

http://localhost:8888/com.isi.Tabular/Tabular.html

Go

Karma
[About](#)

Rows: 1 - 10 of 10

Table Creation

Fill Table

-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

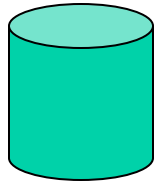
Cell Selected: 0,0

update

Filtering

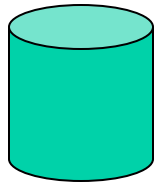
Data Import

Example Data Sources



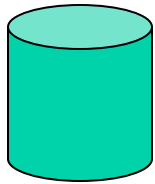
songs 2004

artist, album, song name, length



songs 2005

artist, album, song name, length



Review

artist, song name, reviewer, favorite

Karma

A screenshot of a web application interface. At the top, the word "Karma" is displayed in a large, black, serif font. Below it is a table with a light gray header and a white body. The table has six columns and several rows, each containing a single hyphen "-" character. The cell in the second row and first column is highlighted in light blue and circled with a red oval. Below the table is a gray bar with the text "Cell Selected: 2,0". At the bottom of the interface is a search bar containing the text "alph|", a dropdown menu with three items ("alpha", "herb alpert - lemon tree", and "weekends (alpha mix)"), and an "update" button. A red rectangle highlights the search bar, dropdown menu, and "update" button.

-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

Cell Selected: 2,0

alph|

- alpha
- herb alpert - lemon tree
- weekends (alpha mix)

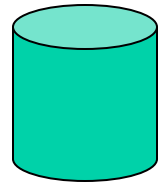
update

Karma

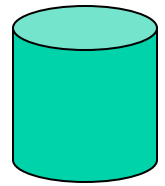
artist	-	-	-	-	-
alpha	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

Cell Selected: 2,0

alpha



songs 2004



songs 2005

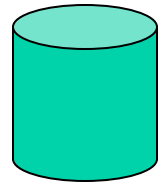
Karma

artist					
-	-	-	-	-	-
alpha	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

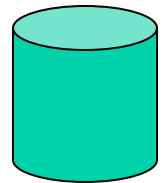
Cell Selected: 0,1

album
bit rate
favorite
genre
reviewer

update



songs 2004



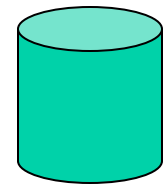
songs 2005

Karma

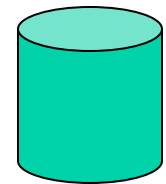
artist	album	song name	-	-	-
-	-	-	-	-	-
alpha	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

Cell Selected: 0,2

song name



songs 2004



songs 2005

Karma

artist	album	song name	---
Alpha	Come From Heaven	Hazeldub	---
Alpha	The Impossible Thrill	AI Station	---
Alpha	Stargazing [Special Edition]	Once Round Town	---
Alpha	Stargazing [Special Edition]	Blue Autumn	---
Alpha	Come		
Alpha	Come		
Alpha	Come		
Alpha	Come		
Alpha	Come		
Alpha	Come		

```
(select artist, album, song name from songs_2004 where  
artist = "alpha")  
union  
(select artist, album, song name from songs_2005 where  
artist = "alpha")
```

Cell Selected: 0,2

song name

update

sort

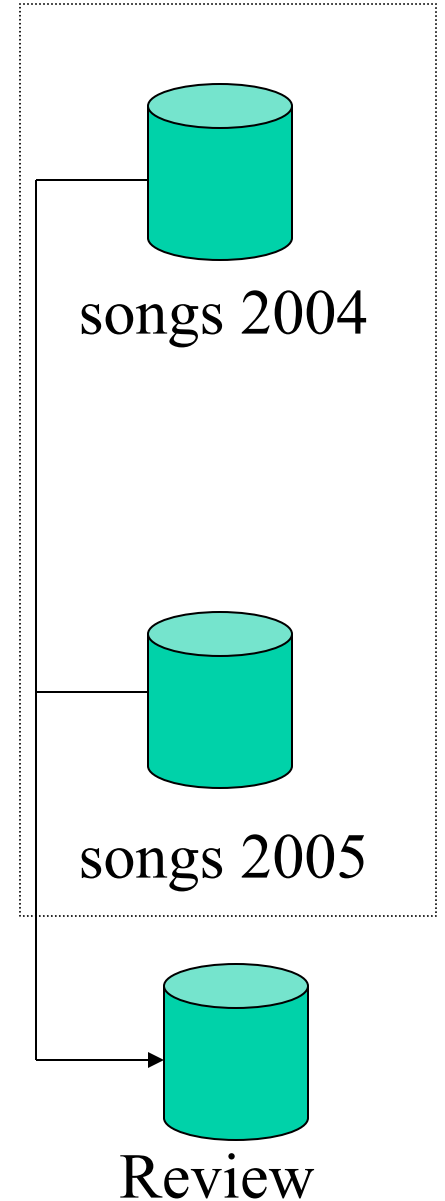
Karma

artist	album	song name	
Alpha	Come From Heaven	Hazeldub	--
Alpha	The Impossible Thrill	AI Station	--
Alpha	Stargazing [Special Edition]	Once Round Town	--
Alpha	Stargazing [Special Edition]	Blue Autumn	--
Alpha	Come From Heaven	Somewhere Not Here	--
Alpha	Come From Heaven	Firefly	--
Alpha	Come From Heaven	With	--
Alpha	Come From Heaven	Apple Orange	--
Alpha	Come From Heaven	Back	--
Alpha	Come From Heaven	Delaney	--

Cell Selected: 0,3

favorite
genre
reviewer
sample rate
size

update



Karma

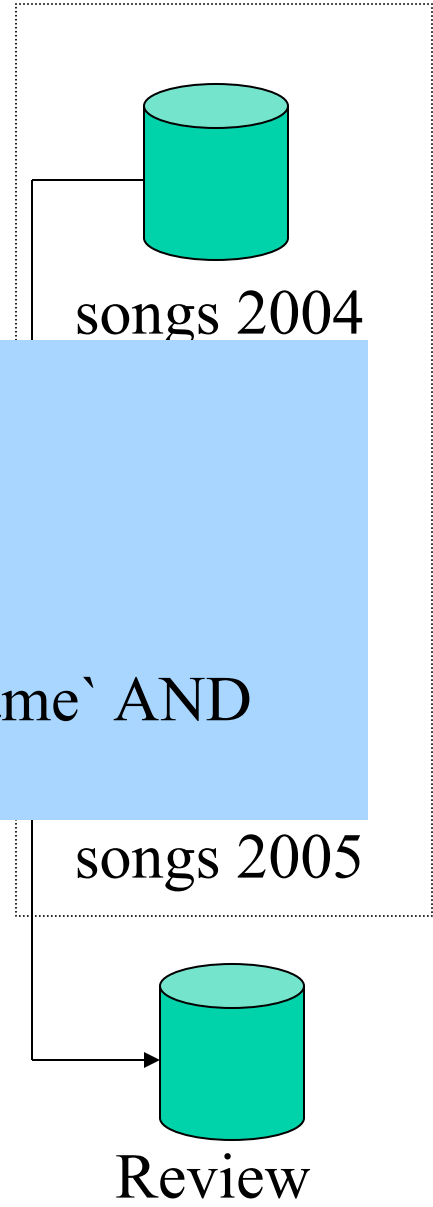
artist	album	song name
Alpha	Come From Heaven	Hazeldub
Alpha		
Alpha		
Alpha		
Alpha		
Alpha		
Alpha		
Alpha		
Alpha	Come From Heaven	Delaney

Cell Selected: 0,3

favorite
genre
reviewer
sample rate
size

update

```
select favorite
from songs_2005, Review
where songs_2005.artist = Review.artist AND
songs_2005.artist = "Alpha" AND
songs_2005.`song name` = Review.`song name` AND
songs_2005.`song name` = "Firefly";
```



Advantages

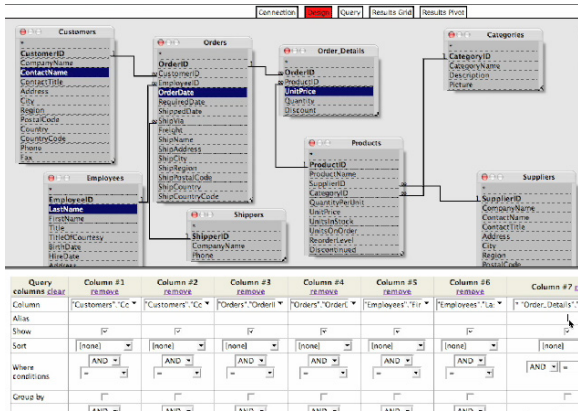
- No query writing
- Hides data sources from users
- Never produces empty result

Open Issues

- Foreign Key Requirement
- Scalability to hundreds of sources
 - Tested with 5 sources (3000 rows)
- Filtering

Related Work

Programming by
Demonstration
[Cypher 1993, Lau 2001]



QBE [Zloof 1975]

Choose database: qbb_ex1

Query
SELECT * FROM qbb_ex1 WHERE Overdraft > 0

Make a Query

- don't know / haven't decided
- yes I want it (click box)
- no I don't (click twice)

Data

	Name	Title	Wage	Overdraft
<input checked="" type="checkbox"/>	Fred	Mr	12000	500
<input checked="" type="checkbox"/>	John	Dr	20000	10000
<input type="checkbox"/>	Sue	Ms	10000	0
<input type="checkbox"/>	Diane	Mrs	2000	0
<input checked="" type="checkbox"/>	Tom	Mr	15000	100
<input checked="" type="checkbox"/>	Jane	Ms	20000	-5000
<input checked="" type="checkbox"/>	Dick	Mr	10000	50

Query By Browsing [Dix 1998]

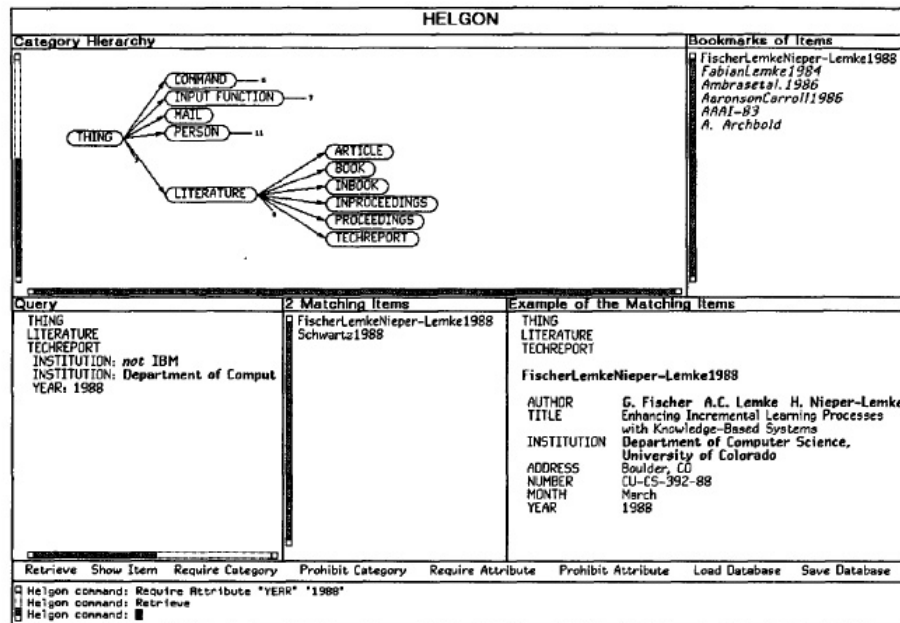
Integrate output from each service

How do you want to integrate 1999highschoolCA and housingSqftCA?

- merge the result together (union). [\(help\)](#)
- compare over the attribute. (join) [\(help\)](#)
- compare over multiple attributes. [\(help\)](#)

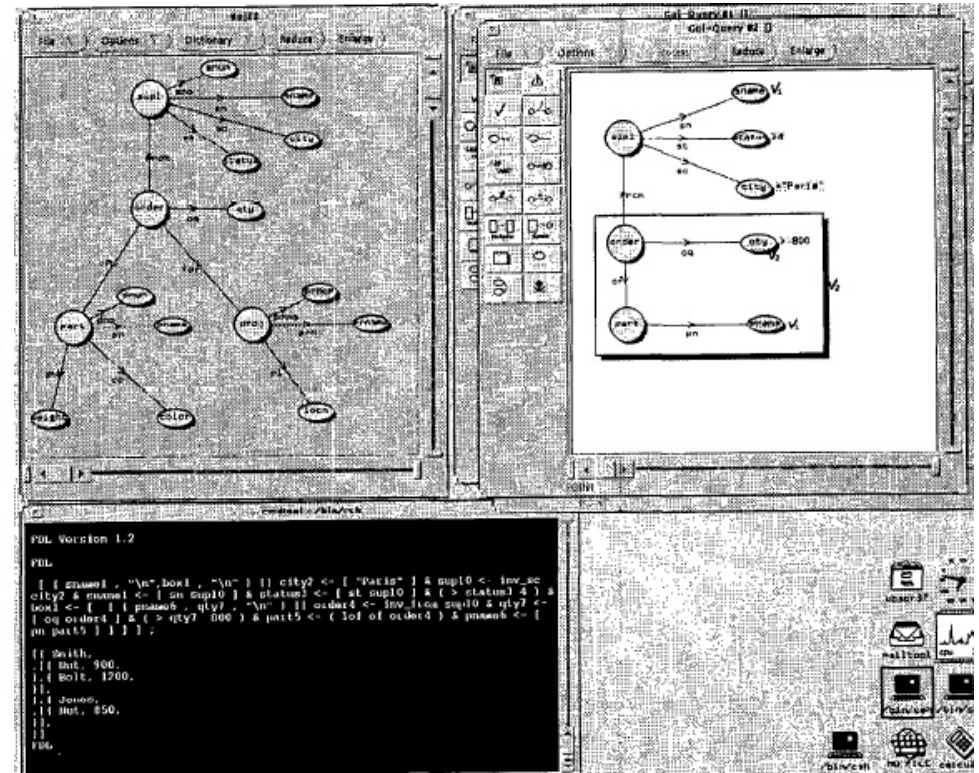
Agent Wizard [Tuchinda 2004]

Related Work



HELGON

Retrieval by formulation
HELGON[Fischer 1989]
RABBIT[Williams 1982]



Gql

Graphical Query Language
[Benzi 1998,
Haw 1994,
Papantonakis 1988]

Evaluation

	Clicks (c) and Key Strokes (k)	Cost
QBE A	$28c+16k$	$4a+2t+d$
Karma A	$17c+4k$	$3a+2t$
QBE B	$39c+28k$	$5a+3t+2d$
Karma B	$25c+7k$	$3a+3t$
QBE C	$78c+54k$	$2*(5a+6t+2d)$
Karma C	$37c+14k$	$3a+6t$

Typing in a value or Selecting a value = 1t unit

Selecting a data source to use = 1d unit

Selecting an attribute = 1a unit

Conclusion and Future Work

- Our contribution: An approach to data integration that
 - Does not require the user to know details about query writing, data sources, or existing values
 - Suggest valid possible values to the user
- In Progress: Filtering
- User studies