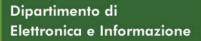


> POLITECNICO DI MILANO



Search Computing

November 7, 2011

Stefano Ceri ... and the SeCo Project Team

Adnan Abid, Mamoun Abu Helu, Davide Barbieri, <u>Daniele Braga</u>, Marco Brambilla, <u>Alessandro Bozzon</u>, Alessandro Campi, Davide Chicco, Emanuele Della Valle, Piero Fraternali, Nicola Gatti, Giorgio Ghisalberghi, Michael Grossniklaus, Davide Martinenghi, Marco Masseroli, Maristella Matera, Chiara Pasini, <u>Silvia Quarteroni</u>, Marco Tagliasacchi, Luca Tettamanti, Salvatore Vadacca, Serge Zagorac

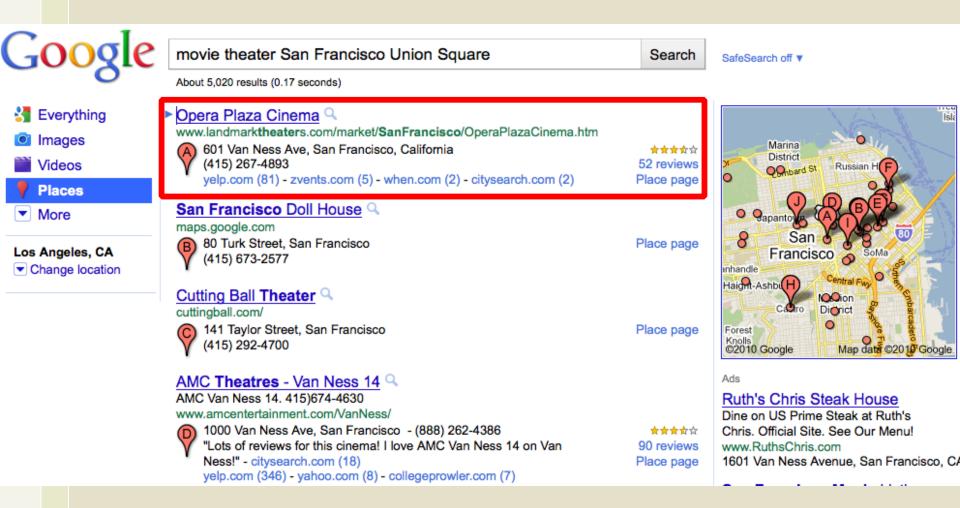
Motivating Examples – What Search Engines can't do 2

 "Where can I find a theater close to Union Square, San Francisco, showing a recent thriller movie, close to a good steak house?"

Search For a Solution Using All Keywords

Google	movie theater San Francisco Union Square steak house 🛛 🗙 Search	Instant is on ▼ SafeSearch off ▼
000310	About 2,620,000 results (0.19 seconds) Advanced search	
 Everything Images Videos More Los Angeles, CA Change location Show search tools 	 Century San Francisco Centre 9 - Union Square - San Francisco, CA ****** 383 reviews San Francisco, CA 94103. Neighborhoods: Union Square, SOMA When I first heard about this movie theatre I thought what an ideal location www.yelp.com//century-san-francisco-centre-9-san-francisco - Cached - Similar Tad's Steakhouse - Union Square - San Francisco, CA ****** 234 reviews - Price range: \$\$ My boyfriend and I had an hour to kill before our movie and we didn't want www.yelp.com/biz/tads-steakhouse-san-francisco-2 - Cached - Similar * Show more results from yelp.com San Francisco Restaurants Downtown & Union Square Restaurants is located in San Francisco's Theater District downtown near Union Square is located in San Francisco's Theater District downtown near Union Square www.sanfrancisco.com/restaurants/union-square. Shopping, Dining Morton's The Steakhouse 400 Post Street San Francisco, CA 94108. (415) 986-5830 Reserve Now! Web Site. Hours Mon-Thu 5:30pm-11pm. Sat 5pm-11pm www.unionsquareshop.com/pages/mortons.html - Cached - Similar Union Square :: Shopping, Dining & Travel Guide :: San Francisco Union Square Experience the unique character of San Francisco's Union Square www.unionsquareshop.com/ - Cached - Similar 	Ads Ruth's Chris Steak House Enjoy a complete three-course meal for only \$39.951 Reserve today. www.ruthschris.com 6100 Topanga Canyon Blvd Suite 1360 See your ad here »

Split the task, and search for theaters first



Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

But there's no thriller!

Opera Plaza Cinemas

601 Van Ness Avenue, San Francisco, CA, United States - (415) 267-4893

 Today's Special

 1hr 39min - Rated R - Comedy - IMDb - : ☆☆☆☆☆

 2:25 4:55 7:25pm

Client 9: The Rise and Fall of Eliot Spitzer 1hr 57min - Rated R - Documentary - <u>Trailer</u> - <u>IMDb</u> 2:30 5:10 8:00pm Ahead of Time 1hr 13min - Documentary - <u>IMDb</u> - : ☆☆☆☆ 2:40 4:40pm

<u>Genius Within: The Inner Life of Glenn Gould</u> 1hr 49min - Documentary - <u>IMDb</u> - : ☆☆☆☆ 2:20 4:50 7:20pm

Try another theater: Found! (The Next Three Days) close enough to Union square....

Showtimes for Union Square, San Francisco, CA

Change Location

Change

Saturday

Sunday

All times

Morning

Evening

Theaters Movies

Show list view

Show map view

Late

Afternoon

> Today Tomorrow Century San Francisco Centre 9 and XD

835 Market Street, San Francisco, CA - (800) 3263264 x991#

Burlesque

1hr 40min - Rated PG-13 - Drama - <u>Trailer</u> - <u>IMDb</u> - : ☆☆☆☆ <u>10:50am</u> <u>1:00</u> <u>1:50</u> <u>4:00</u> <u>7:00</u> <u>8:00</u> <u>10:00</u> <u>10:45pm</u>

Tangled in Disney Digital 3D 1hr 40min - Rated PG - Animation/Comedy 11:00am 12:00 1:40 2:40 4:25 5:25 7:10 9:50pm

The Chronicles of Narnia: The Voyage of the Dawn Treader in Digital 3D 1hr 52min - Rated PG - Action/Adventure/Scifi/Fantasy - IMDb -: ***** 12:01am

 Love and Other Drugs

 1hr 53min - Rated R - Drama - Trailer

 11:15am
 12:30

 2:00
 3:30
 4:50
 6:30

Tangled 1hr 40min - Rated PG - Animation/Comedy - <u>Trailer</u> - <u>IMDb</u> <u>10:40pm</u> Due Date

1hr 35min - Rated R - Comedy/Drama - <u>Trailer</u> - <u>IMDb</u> - : ☆☆☆☆ 12:20 2:50 5:20 7:50 10:10pm

The Next Three Days 2hr 2min - Rated PG-13 - Drama/Romance/Suspense/Thriller -Trailer - IMDb 1:30 4:30 7:30 10:35pm

 Skyline

 1hr 40min - Rated PG-13 - Suspense/Thriller/Scifi/Fantasy

 <u>Trailer</u> - <u>IMDb</u> - : ☆☆☆☆

 11:20am
 1:50
 4:40
 7:20
 9:40pm

 The Tempest (2010/II)

 1hr 50min - Rated PG-13 - Scifi/Fantasy/Drama/Romance - Trailer

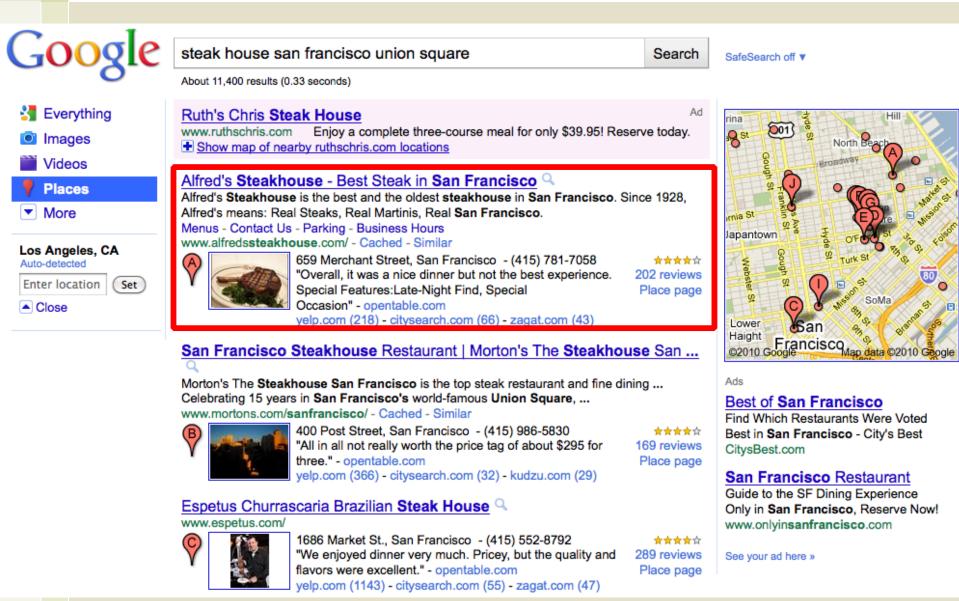
 - IMDb - : ★★★☆☆

 12:05am

Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Independent search for steak house



Database Management

Prof. Stefano Ceri

POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Done! Close enough! (data integration and ranking in the user's brain)

Google maps

steak house san francisco union square

Search Maps

Show search options

Get Directions My Maps

Alfred's Steak House

659 Merchant Street, San Francisco, CA 94111-2505 (415) 781-7058 <u>alfredssteakhouse.com</u> <u>Directions</u> <u>Search nearby</u> <u>more</u>▼

"Great food, great value, good atmosphere" - tastyr.com ... "Service slow and inattentive" - zagat.com ... "This is about steak cooked to perfection in a great traditional atmosphere" - tastyr.com ... "Overall very fine meat-emphasized food and traditional elegant ambience" - citysearch.com ... "Classy unprentious dining" - zagat.com

Details

Cuisine: Specialties, Steakhouse Dining Style: Casual Elegant Reservations Policy: Suggested (Highly) Year Opened: 1928 Meals Served: Dinner, Lunch Menu: Menu



Edit this place - Business owner?

citysearch.com



Print 🖂 Email 📾 Link



Ads

San Francisco Coupons

1 ridiculously huge coupon a day. Get 50-90% off the Bay's best! www.Groupon.com/San-Francisco

VISION

Database Management

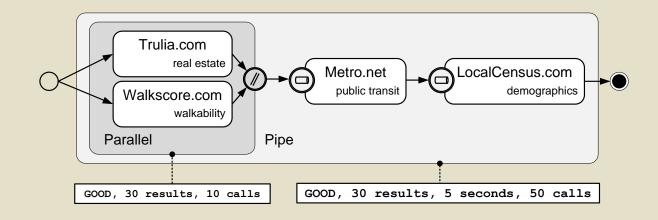
8

The Search Computing Project

- ERC-founded project
- 5 years Started in 2009, now at month 36
- Build theories, methods and tools to support <u>search-oriented multi-dimensional queries</u>
 - Given a **multi-domain** query
 - Build global solutions by integrating data produced by search services
 - Rank global solutions according to a global rank function and output results in rank order
 - Support user-friendly interfaces for query definition and result browsing, which allow adding search domains while the search process proceeds and possibly change the relative weight of each ranking

Search Computing = Search Service Composition 11

- Searching the Web of Data requires demand-driven service composition
- Composition abstractions should emphasize few elements: service invocations, fundamental operations, precedences, global constraints on execution
- Data composition should be search-driven producing few top results very fast



Modular software view of search applications

- New generation software for building focused search applications
- Covering the functionalities of vertical search systems (e.g. "expedia", "amazon") on more focused application domains (e.g. localized real estate or leasure planning, sector-specific job market offers, support of biomed research, ...)
- Should be easy-to-build, easy-to-query, easy-to-maintain, easy-to-scale...

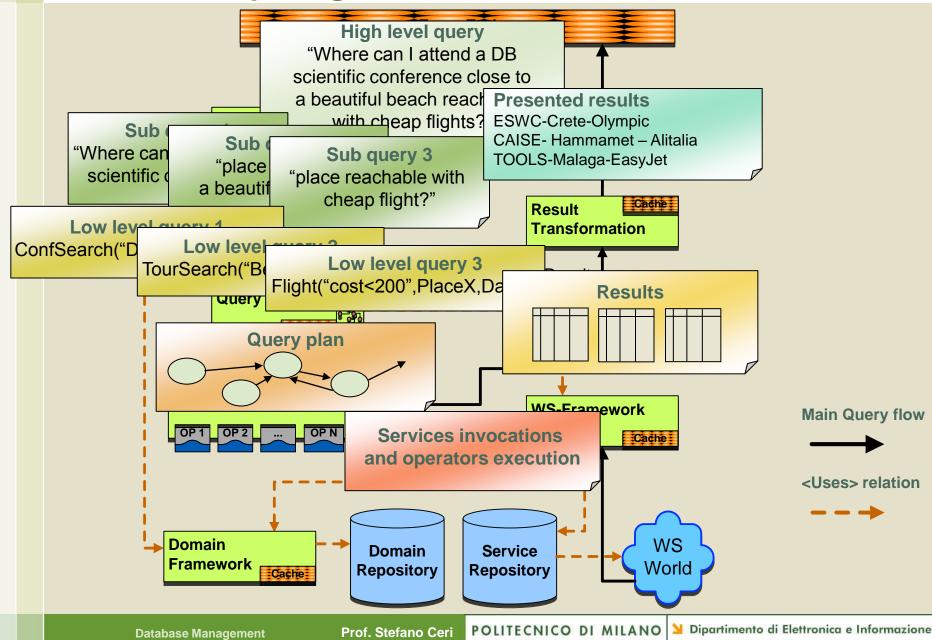
TECHNOLOGICAL FRAMEWORK

Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

13

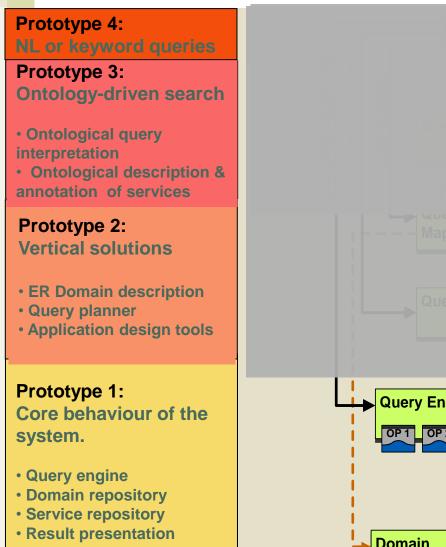
Search Computing architecture: overall view

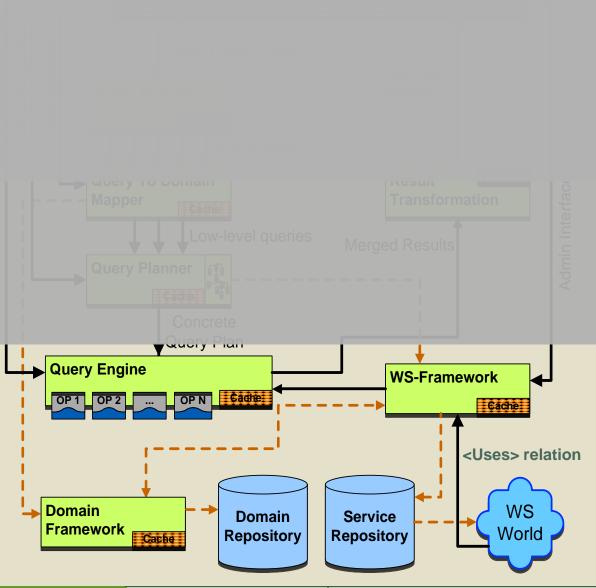


Search Computing architecture: incremental prototyping

Prof. Stefano Ceri

15





POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Database Management

LIQUID QUERY INTERFACE

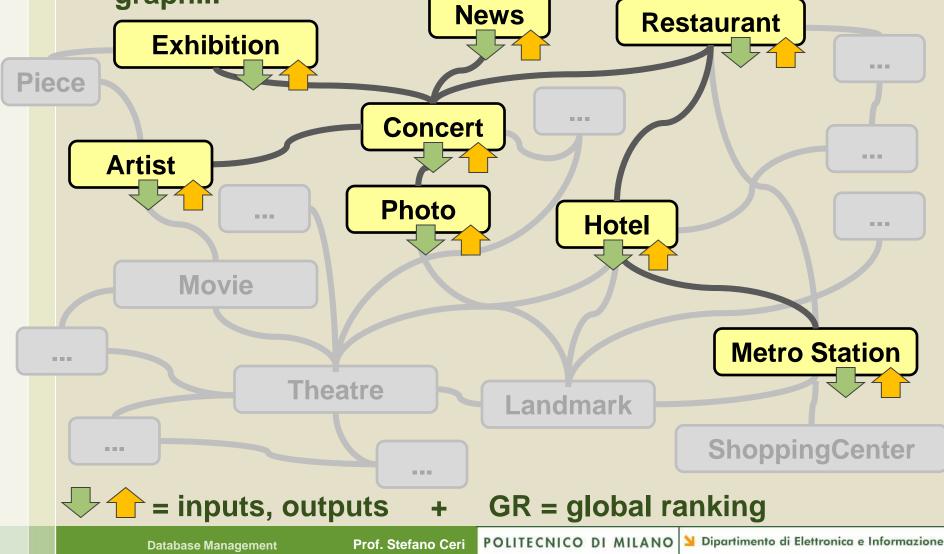
Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

16

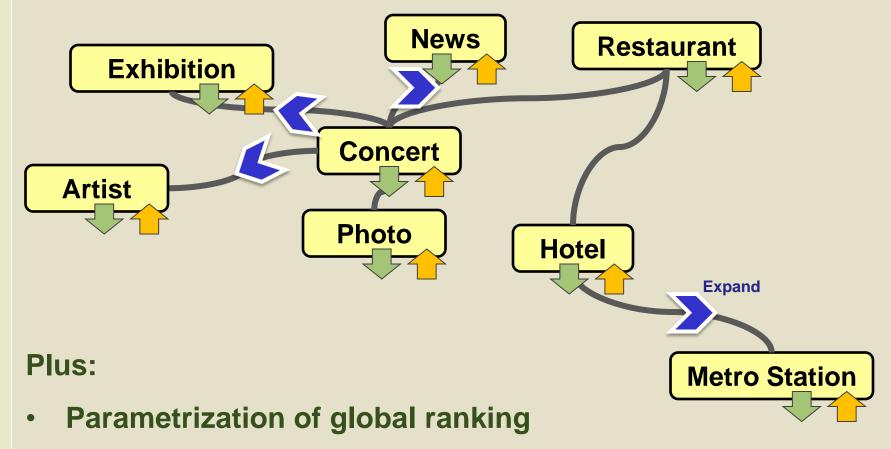
Liquid query definition

It consists of subsetting and parametrizing the resource graph...



Liquid query definition

... And then characterizing the user interaction



- Data visualization options
- .. and so on

Database Management

Exploration of the Service Space

Object list

Real Estate: Sale and Rent of Real Estate

Conterence: not or conterence

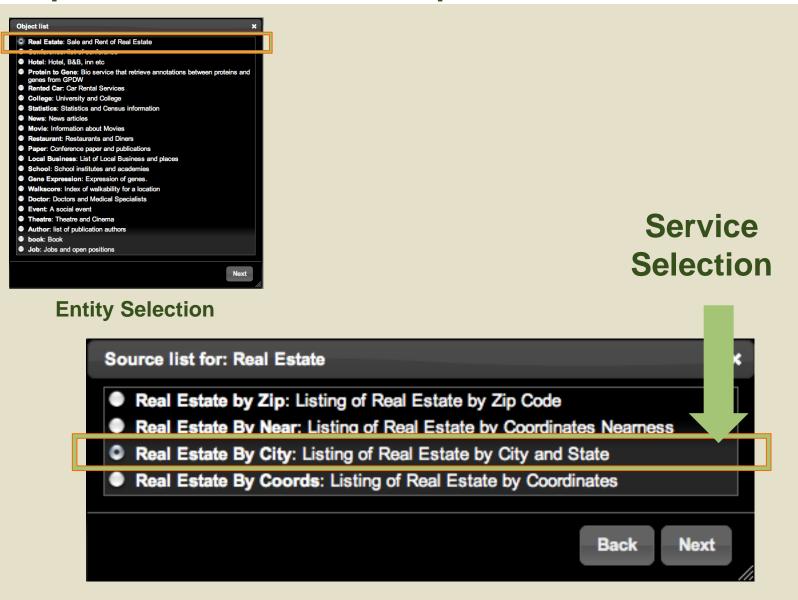
- Hotel: Hotel, B&B, inn etc
- Protein to Gene: Bio service that retrieve annotations between proteins and genes from GPDW
- Rented Car: Car Rental Services
- College: University and College
- Statistics: Statistics and Census information
- News: News articles
- Movie: Information about Movies
- Restaurant: Restaurants and Diners
- Paper: Conference paper and publications
- Local Business: List of Local Business and places
- School: School institutes and academies
- Gene Expression: Expression of genes.
- Walkscore: Index of walkability for a location
- Doctor: Doctors and Medical Specialists
- Event: A social event
- Theatre: Theatre and Cinema
- Author: list of publication authors
- book: Book
- Job: Jobs and open positions

Next

×

Entity Selection

Exploration of the Service Space



Exploration of the Service Space



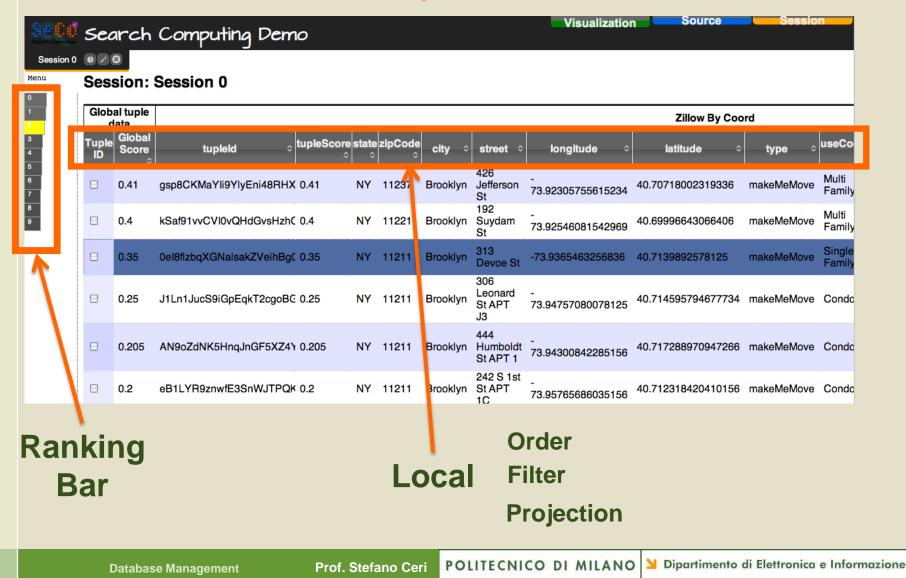
Entity Selection



Database Management

Result Presentation

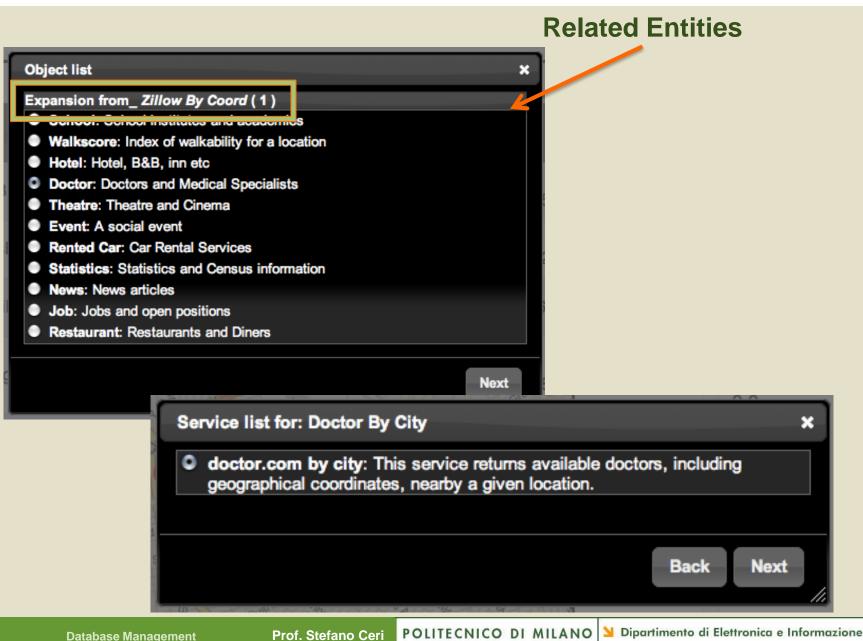
Tabular Representation



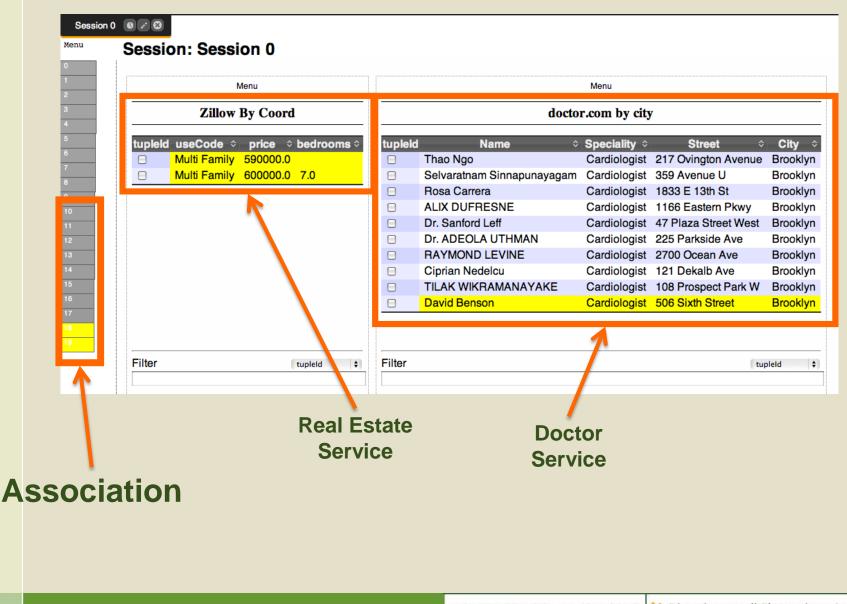
Result Presentation (Map)

Session 0 0 20 Menu ito S Huron 54th Rd Java St un69 66th St Brd Si \$ 55th Ave enAve 56th Rd 55th Dr G 00 3 Pust St 56th Dr FXI ordgs Ave 57th Rd 68th ã 5 Maspeth Ave Eliot Ave 5 FAIT 60th St 55 Grand St. 6 Madison St Woodwa Grand St. Dr S FDR Dr N Stagg St (eap 5 8 S 9th St holes St 王 Johnson Ave 9 à John St Water St Brook Flushing Ave ō ens Expy Park Ave (alworth ranklin Cooper Vemon Ave 80 đ Hart St Pulaski St Myrtle Ave Me 20 K allo De afavette Ave Greene Ave G 3 Quincy St Monroe St Clifton Coope 말 Evergreen Madison St Cemetery 30 Hatsey FA 52010 0.41 0.4 0.35 0.25 0.20 0.2 0.16 0.0 0.0 0.0 -Actions More One De-Filter More All Filter

Exploration options from a given state



Result Presentation (Atom View)



Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Result Visualization – Combinations on Maps



Database Management

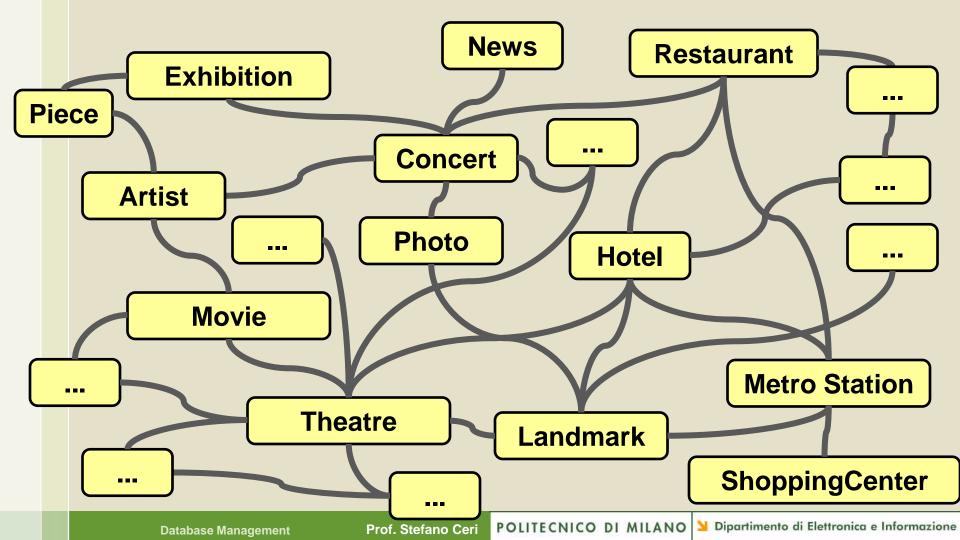
SERVICE REGISTRATION

Database Management

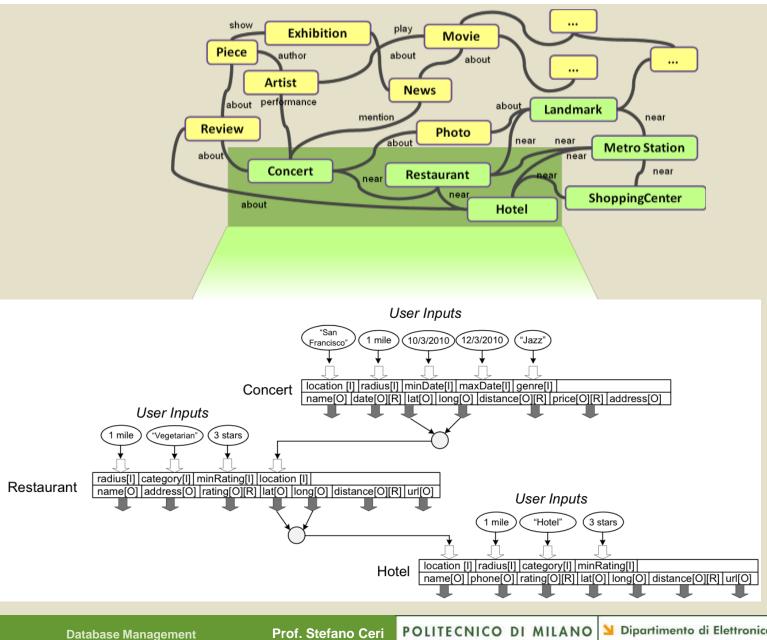
Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Rationale of Service Registration

- Providing a "Semantic Resource Framework" (SRF) where concepts of the real world are mapped to entities and interconnected by relationships
- Along the idea of the "web of objects" instead of the "web of pages"



Under the scene...

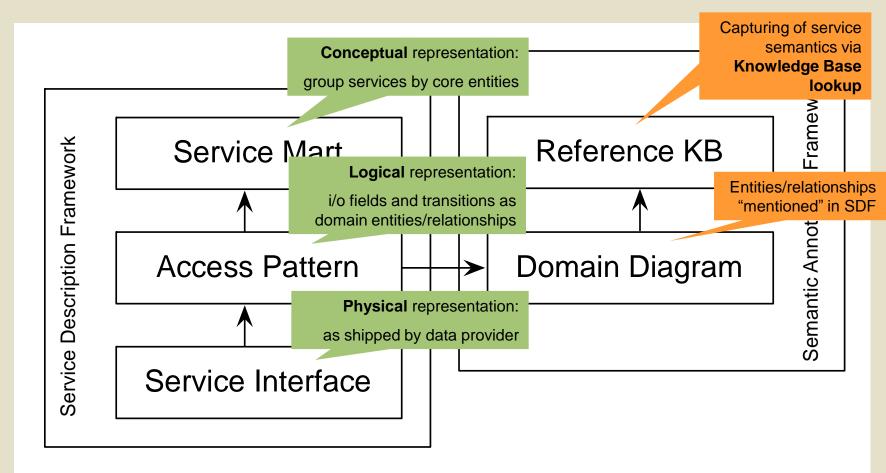


Prof. Stefano Ceri

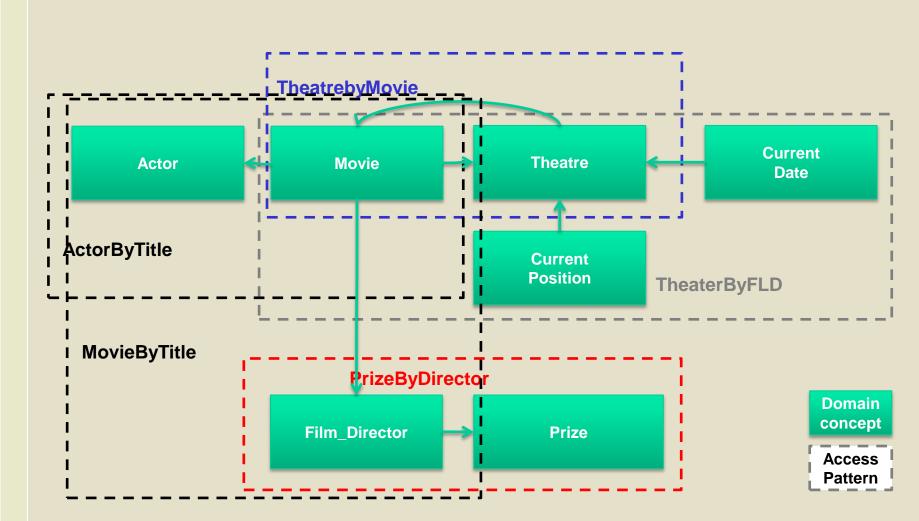
POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Service Framework in SeCo

• A Service Description Framework coupled with a Semantic Annotation Framework



Semantic Framework: Domain Diagram and Access Patterns



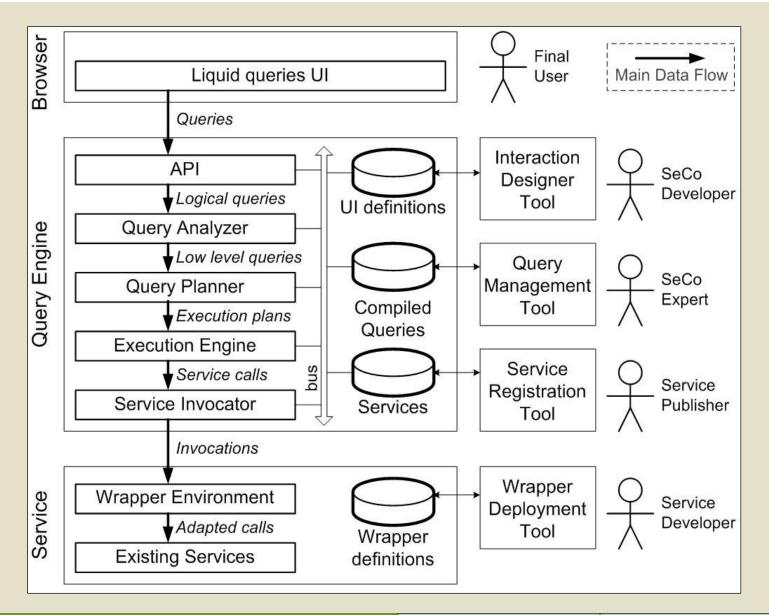
Database Management

SECO ENGINE

Database Management

33

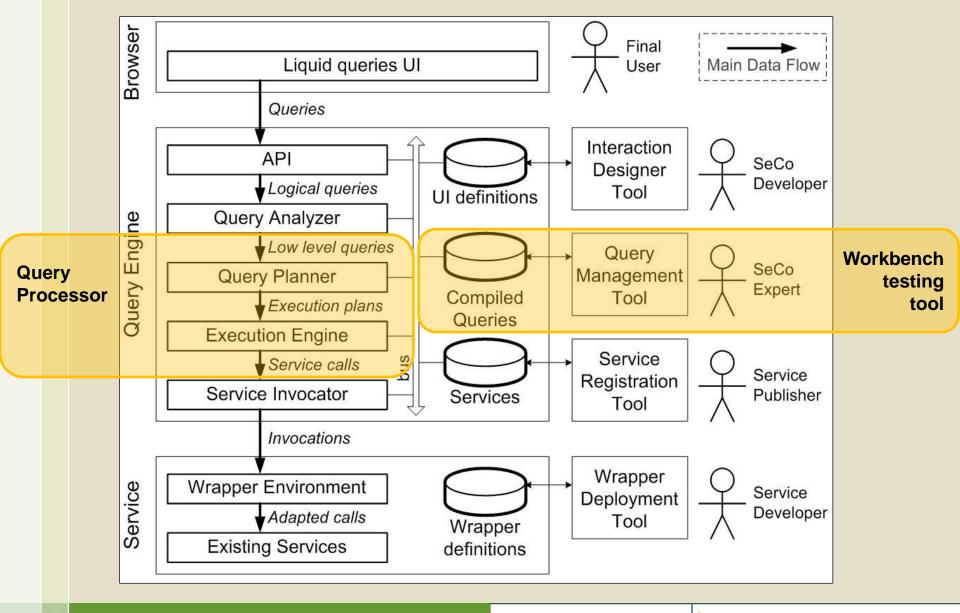
The Query Processor in the Big Picture



Database Management

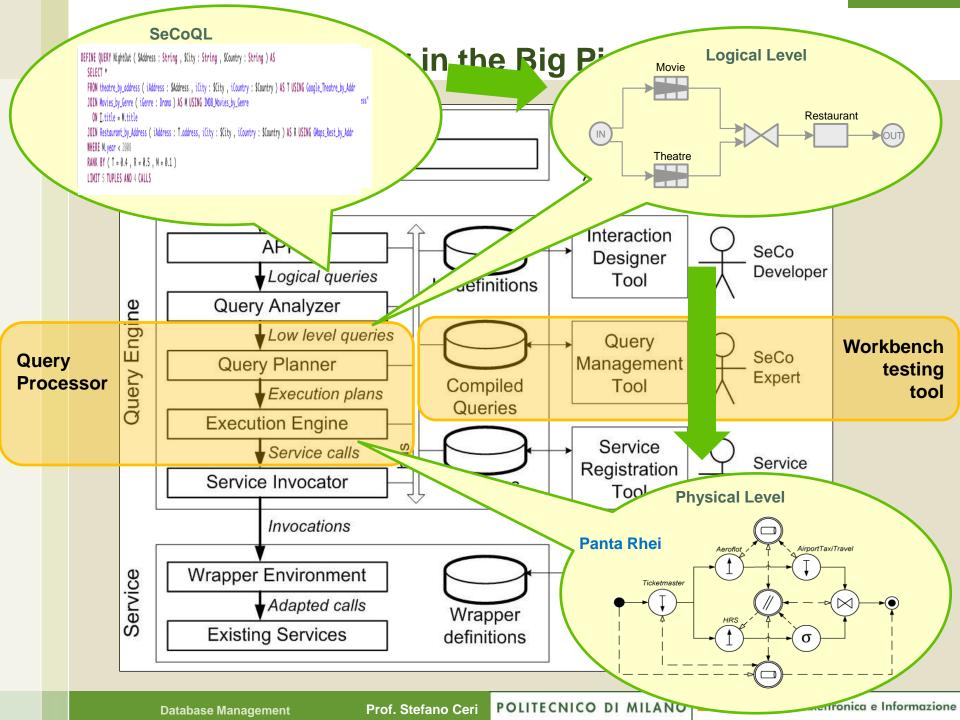
Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

The Query Processor in the Big Picture



Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione



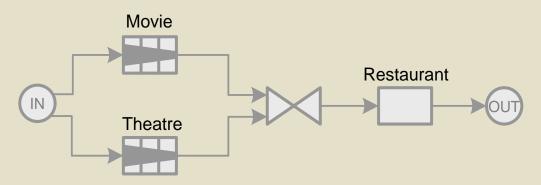
SeCoQL

An old drama movie showing tonight in a theatre close to a good restaurant

NightOut(Piccadilly, London, UK)

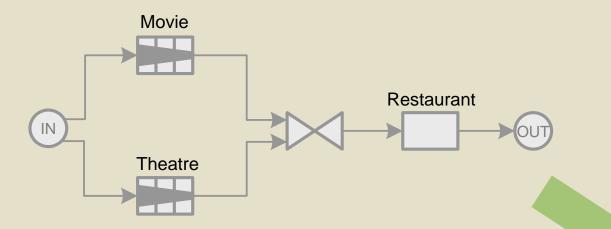
First step: from conjunctive queries to logical plans

Generation of a logical plan

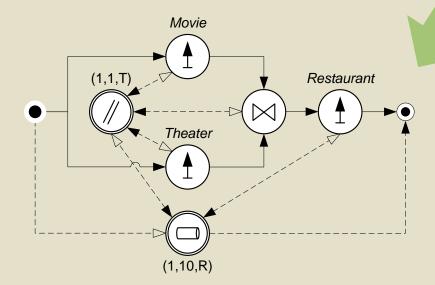




Second step: from logical to physical query plans

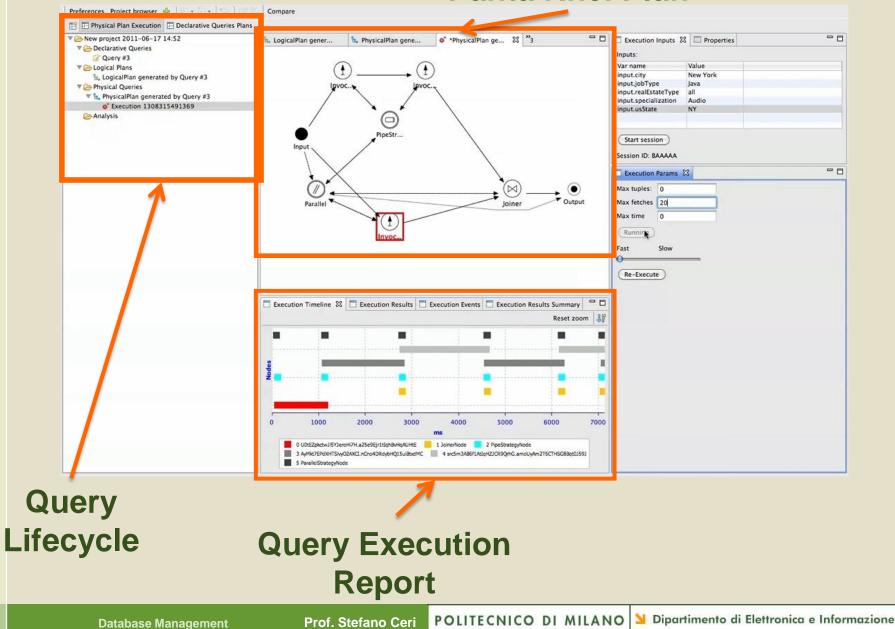


Then the planner generates a physical, executable query plan, expressed in *Panta Rhei*



Workbench

Panta Rhei Plan



THEORY

Database Management

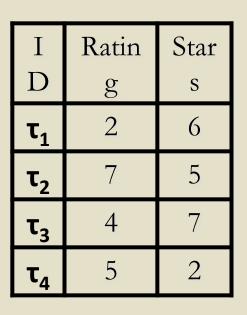
Proximity Rank Join (Martinenghi-Tagliasacchi, VLDB 2010)

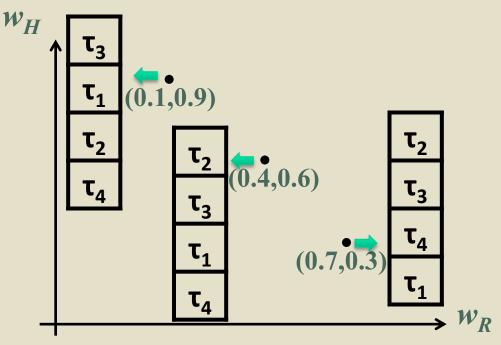


Ranking with uncertain scoring functions (Soliman, Ilyas, Martinenghi, Tagliasacchi, Sigmod 2011)

How to select the right weights?

```
SELECT *
FROM Restaurants R, Hotels H
WHERE R.Location = H.Location
ORDER BY w_R \cdot R.Rating + w_H \cdot H.Stars
LIMIT K
```





CURRENT STANDING

Database Management

Prof. Stefano Ceri POLITECNICO DI MILANO 🎽 Dipartimento di Elettronica e Informazione

Results after 36 months

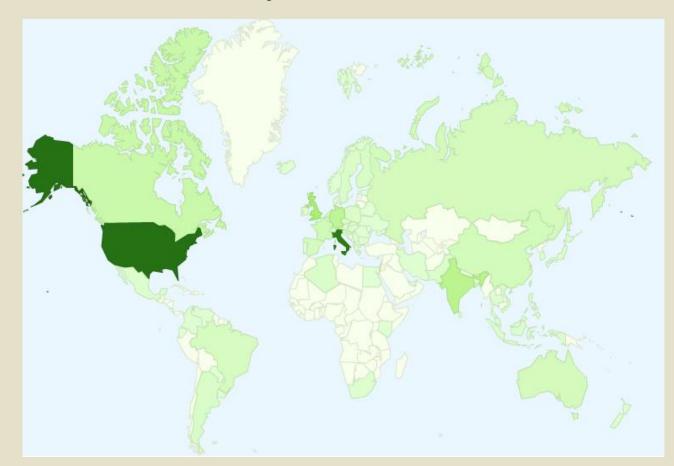
- Concepts
 - Service marts, join methods, panta rhei, liquid query

Research results

- LNCS: Search Computing Challenges and Directions (2010)
- LNCS: New Trends in Search Computing (2011)
- Publications (incl. VLDB, WWW, SIGMOD, TODS)
- US Patent filed (top-k method, random & sequential services)
- Site: www.search-computing.eu
- Blog: http://blog.search-computing.com/
- Temporary research positions (3 phd, 5 post-ms, 3 post-doc)

Accesses to Web Site & Blog (2010)

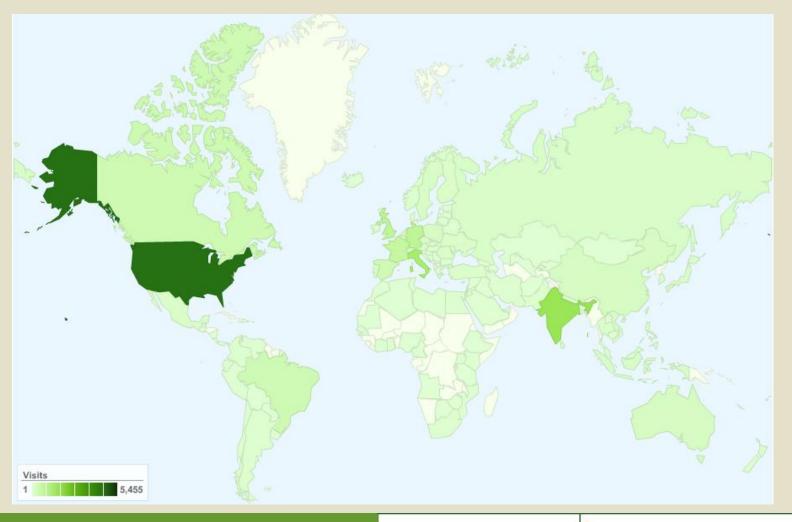
Visits: 20% USA, 18% Italy, 6% UK, 4% India, 4% Canada



Database Management

Accesses to Web Site & Blog (2011)

Visits: 27% USA, 10% India, 8% Italy, 4% Germany – UK - France



Database Management

Events in 2011

- Functionality Demo at WWW 2011 (Bangalore)
- Engine Demo at ACM-Sigmod (Athens)



Database Management

Events in 2011-12

- Workshops:
 - ExploreWeb (Brambilla, Fraternali, Schwabe) http://exploreweb.search-computing.org/
 - **DBRank** Workshop (Chakrabarti, Martinenghi)
 - Very Large Data Search (VLDS) (Brambilla, Casati, Ceri) http://vlds2011.search-computing.net/
 - Ordering and Reasoning (OrdRing) (Bozzon, Della Valle, Horrocks) http://ordring2011.search-computing.org/
 - DataView 2011 (Bozzon, Comai, Norrie) http://dataview.como.polimi.it/2011/
- Third LNCS Book ?
- Planned VLDB Journal special issue on "Web search over structured information and crowds" (tent. Title, Brambilla/Ceri/Halevy, Sept. 2012)

Future Research Directions

- NLP or keyword-based queries
 - focus on subquery partitioning & mapping to domains
- Social dimension
 - crowd-searching using social platforms
- Verticals
 - joint work with Diadem (Gottlob) on London real-estate
 - regional platforms for "quality of life"

And finally....

