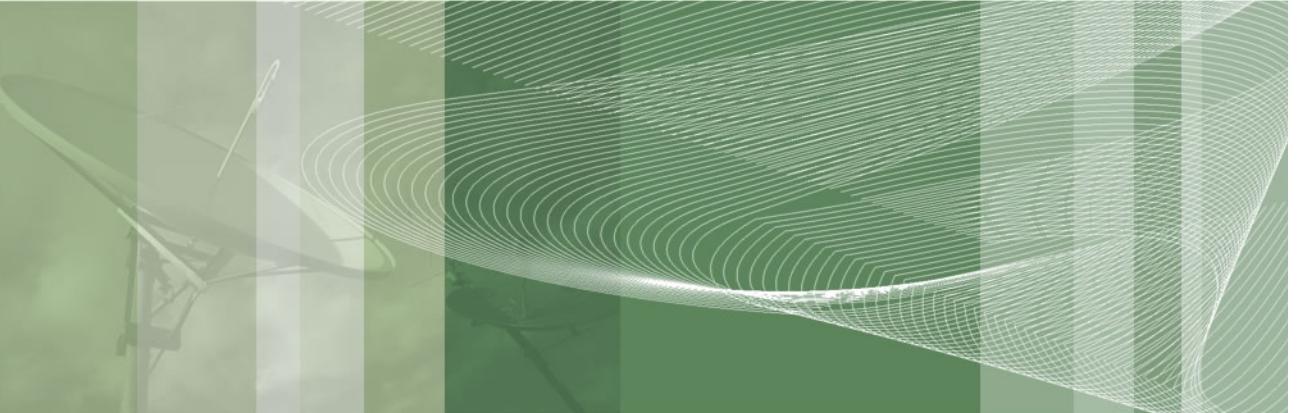


POLITECNICO DI MILANO

Dipartimento di
Elettronica e Informazione



Search Computing Introduction

October 10, 2011

Stefano Ceri ... and the SeCo Project Team

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

Motivation: multi-domain queries with ranking

2

A class of queries search engines are not good at

- “Where can I attend an interesting scientific **conference** in my field and at the same time relax on a beautiful beach nearby? ”
- “Retrieve **jobs** as Java developer in the Silicon Valley, nearby affordable fully-furnished **flats**, and close to good schools”
- “Find a **theater** close to Union Square, San Francisco, showing a recent thriller movie, close to a steak house? ”
- With a complex notion of “best”
 - with many factors contributing to optimality
- Involving several different data **sources**
 - possibly **hidden** in the deep Web
 - typically returning **ranked** results (search services)
- With possibly articulated “join” conditions
 - capturing **search sessions** rather than one-shot queries

Due to query complexity, not data heterogeneity or unavailability

Motivating Examples – What Search Engines can't do 3

- “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



movie theater San Francisco Union Square steak house

About 2,620,000 results (0.19 seconds)

[Advanced search](#)

Instant is on ▾
SafeSearch off ▾

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 ...](#)



Steakhouse. A sleek, modern **San Francisco steakhouse**, cityhouse in the Parc is located in **San Francisco's Theater District** downtown near **Union Square**. ...

[www.sanfrancisco.com/restaurants/union-square.html](#) - Cached - Similar

► [Morton's The Steakhouse :: 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

[+ Show more results from unionsquareshop.com](#)

Ads

[Ruth's Chris Steak House](#)

Enjoy a complete three-course meal for only \$39.95! Reserve today.

[www.ruthschris.com](#)

6100 Topanga Canyon Blvd Suite 1360

[See your ad here »](#)

Split the task, and search for theaters first

Google

movie theater San Francisco Union Square

Search

SafeSearch off ▾

About 5,020 results (0.17 seconds)

Everything

Images

Videos

Places

More

Los Angeles, CA

Change location

Opera Plaza Cinema

www.landmarktheaters.com/market/SanFrancisco/OperaPlazaCinema.htm

A 601 Van Ness Ave, San Francisco, California
(415) 267-4893

yelp.com (81) - zevents.com (5) - when.com (2) - citysearch.com (2)

★★★★★

52 reviews

Place page

San Francisco Doll House

maps.google.com

B 80 Turk Street, San Francisco
(415) 673-2577

Place page

Cutting Ball Theater

cuttingball.com/

C 141 Taylor Street, San Francisco
(415) 292-4700

Place page

AMC Theatres - Van Ness 14

AMC Van Ness 14. 415)674-4630

www.amcertainment.com/VanNess/

D 1000 Van Ness Ave, San Francisco - (888) 262-4386
"Lots of reviews for this cinema! I love AMC Van Ness 14 on Van Ness!" - citysearch.com (18)
yelp.com (346) - yahoo.com (8) - collegeprowler.com (7)

★★★★★

90 reviews

Place page

Map data ©2010 Google

Ads

Ruth's Chris Steak House

Dine on US Prime Steak at Ruth's Chris. Official Site. See Our Menu!

www.RuthsChris.com

1601 Van Ness Avenue, San Francisco, CA 94109

S F M N P G

Inspect Theatre Details: Looks good...

[Get Directions](#) [My Maps](#)

[Edit this place - Business owner?](#)

[Print](#) [Email](#) [Link](#)

Opera Plaza Cinema

601 Van Ness Ave, San Francisco, CA 94102

(415) 267-4893

[landmarktheaters.com](#)

[Directions](#) [Search nearby](#) [more▼](#)

Category: Movie Theater

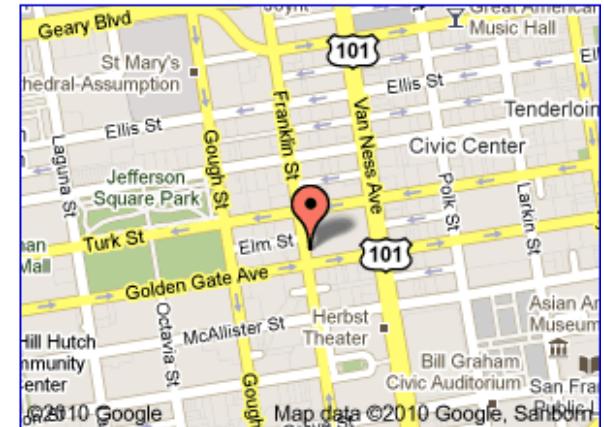
Transit: [Metro Civic Center Station/Downtn](#) (0.4 mi) F, J, K, L, M, N, T, ... 6, 9, 9L, ...

52 reviews **Your rating:**

"As for rudeness, I had no problems nor seen any for the 2 visits I was there" - citysearch.com ... "The Best Movies in Town: Yes the best!" - yahoo.com ... "Its edgy, uplifting and full of hope" - zevents.com ... "Incidentally, I've also gotten good movie advice here" - citysearch.com ... "Como No te voy a Querer?" - when.com



sfgate.com



Details

Showtimes Schedule: [Showtimes Schedule](#)

Reputation Trend: [Reputation Trend](#)

High Resolution Image: [High Resolution Image](#)

[judysbook.com](#), [wcities.com](#), [google.com](#)

[More details »](#)

Ads

[Universal Studios in L.A.](#)

Buy a Day, Get a Week Free.

Official site. Buy now.

[www.universalstudioshollywood.com](#)

Photos

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 - [Trailer](#) - [IMDb](#)
2:30 5:10 8:00pm

Ahead of Time

1hr 13min - Documentary - [IMDb](#) - : ★★★★★
2:40 4:40pm

Genius Within: The Inner Life of Glenn Gould

1hr 49min - Documentary - [IMDb](#) - : ★★★★★
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](#)

› [Today](#)

[Tomorrow](#)

[Saturday](#)

[Sunday](#)

› [All times](#)

[Morning](#)

[Afternoon](#)

[Evening](#)

[Late](#)

› [Theaters](#)

[Movies](#)

› [Show list view](#)

[Show map view](#)

Century San Francisco Centre 9 and XD

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

Burlesque

1hr 40min - Rated PG-13 - Drama - [Trailer](#) - [IMDb](#) - : ★★★★★
[10:50am](#) [1:00](#) [1:50](#) [4:00](#) [7:00](#) [8:00](#) [10:00](#) [10:45pm](#)

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](#) - [IMDb](#)
[11:15am](#) [12:30](#) [2:00](#) [3:30](#) [4:50](#) [6:30](#) [7:40](#) [9:20pm](#)

Tangled

1hr 40min - Rated PG - Animation/Comedy - [Trailer](#) - [IMDb](#)
[10:40pm](#)

Due Date

1hr 35min - Rated R - Comedy/Drama - [Trailer](#) - [IMDb](#) - : ★★★★★
[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 - [Trailer](#) - [IMDb](#) - : ★★★★★
[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](#)

Independent search for steak house



steak house san francisco union square

Search

SafeSearch off ▾

About 11,400 results (0.33 seconds)

Everything

Images

Videos

Places

More

Los Angeles, CA

Auto-detected

Enter location

Set

Close

Ruth's Chris Steak House

www.ruthschris.com Enjoy a complete three-course meal for only \$39.95! Reserve today.

[Show map of nearby ruthschris.com locations](#)

Ad

Alfred's Steakhouse - Best Steak in San Francisco

Alfred's Steakhouse is the best and the oldest steakhouse in San Francisco. Since 1928, Alfred's means: Real Steaks, Real Martinis, Real San Francisco.

Menus - Contact Us - Parking - Business Hours

www.alfredssteakhouse.com/ - Cached - Similar



659 Merchant Street, San Francisco - (415) 781-7058

"Overall, it was a nice dinner but not the best experience.

Special Features: Late-Night Find, Special Occasion" - [opentable.com](#)

yelp.com (218) - [citysearch.com](#) (66) - [zagat.com](#) (43)

★★★★★

202 reviews

[Place page](#)

San Francisco Steakhouse Restaurant | Morton's The Steakhouse San ...

Morton's The Steakhouse San Francisco is the top steak restaurant and fine dining ...

Celebrating 15 years in San Francisco's world-famous Union Square, ...

www.mortons.com/sanfrancisco/ - Cached - Similar



400 Post Street, San Francisco - (415) 986-5830

"All in all not really worth the price tag of about \$295 for three." - [opentable.com](#)

yelp.com (366) - [citysearch.com](#) (32) - [kudzu.com](#) (29)

★★★★★

169 reviews

[Place page](#)

Espetus Churrascaria Brazilian Steak House

www.espetus.com/



1686 Market St., San Francisco - (415) 552-8792

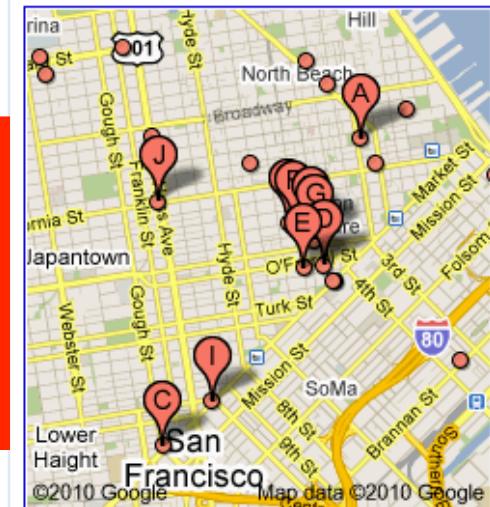
"We enjoyed dinner very much. Pricey, but the quality and flavors were excellent." - [opentable.com](#)

yelp.com (1143) - [citysearch.com](#) (55) - [zagat.com](#) (47)

★★★★★

289 reviews

[Place page](#)



Ads

Best of San Francisco

Find Which Restaurants Were Voted Best in San Francisco - City's Best CitysBest.com

San Francisco Restaurant

Guide to the SF Dining Experience Only in San Francisco, Reserve Now! www.onlyinsanfrancisco.com

[See your ad here »](#)

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](#)

[Edit this place](#) - [Business owner?](#)

[Print](#) [Email](#) [Link](#)

Alfred's Steak House

659 Merchant Street, San Francisco, CA 94111-2505

(415) 781-7058

[alfredssteakhouse.com](#)

[Directions](#) [Search nearby](#) [more▼](#)

Category: Steak House

Price: \$\$\$ **Reservations:** [opentable.com](#)

Hours: Today 11:30 AM – 2:30 PM 5:30–9:30 PM

Transit: [Metro Montgomery Station/Downtown](#) (0.4 mi) F, J, K, L, M, N,

T, ... 5, 6, 9, ...

202 reviews 1 Your rating:

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

Details

Cuisine: Specialties, Steakhouse

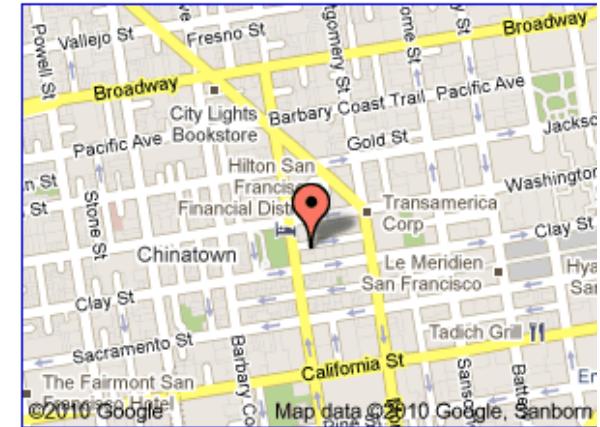
Dining Style: Casual Elegant

Reservations Policy: Suggested (Highly)

Year Opened: 1928

Meals Served: Dinner, Lunch

Menu: [Menu](#)



Ads

[San Francisco Coupons](#)

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

Motivating Examples – Why Search Engines can't do it? 10

- Query is about distinct domains that should be linked
- Query deals with multiple rankings, although hard to compute (“close” theatre, “recent” thriller, “good” steak house)

Note that

- enough data is on the Web but
- not on a single web page.

VISION

Observed trends

- More and more data sources become accessible through Web APIs (as services)
 - Surface & deep Web
- Data sources are often coupled with **search APIs**
- Publishing of structured and interconnected data is becoming popular (*Linked Open Data*)



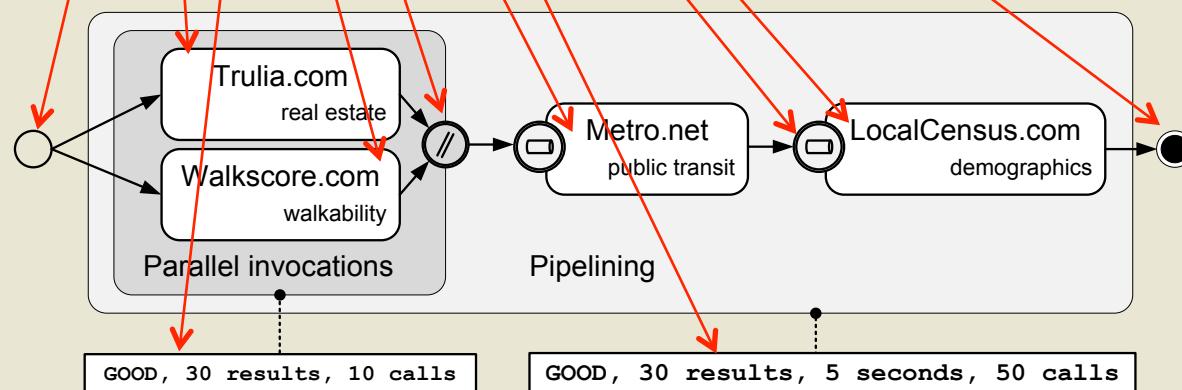
Opportunity for building **focused search systems**
composing results of several data source

- easy-to-build, easy-to-query, easy-to-maintain, easy-to-scale...
- covering the functionalities of vertical search systems (e.g. “expedia”, “amazon”) on **more focused application domains**
 - (e.g. **localized** real estate or leisure planning, **sector-specific** job market offers, support of biomedical **research**, ...)

Search Computing = service composition “on demand” 13

- Composition **abstractions** should emphasize few aspects:
 - service invocations
 - fundamental operations (parallel invocations, joins, pipelining, ...)
 - global constraints on execution
- Data composition should be search-driven
 - aimed at producing **few top results** very fast

A house in a walkable area, close to public transportation and located in a pleasant neighborhood



Modular software view of search applications

14

- 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 leisure planning, sector-specific job market offers, support of biomed research, ...)
- Should be easy-to-build, easy-to-query, easy-to-maintain, easy-to-scale...
- Key ingredient: **Search Service** – a data service extracting top results for queries on data sources

The Search Computing “Manifesto”

- Build theories, methods, and tools to support search-oriented multi-domain queries

Given a **multi-domain query** over a set of **search services**

- Build global answers by **combining** data from each service
- Rank global answers according to a **global ranking** and output results in ranking order
- Support user-friendly query **formulation** and browsing of results
- Include new domains while the **search process** proceeds
- Possibly change the relative weight of each partial ranking

“Searching via interactive/dynamic mashups of ranked data sources”

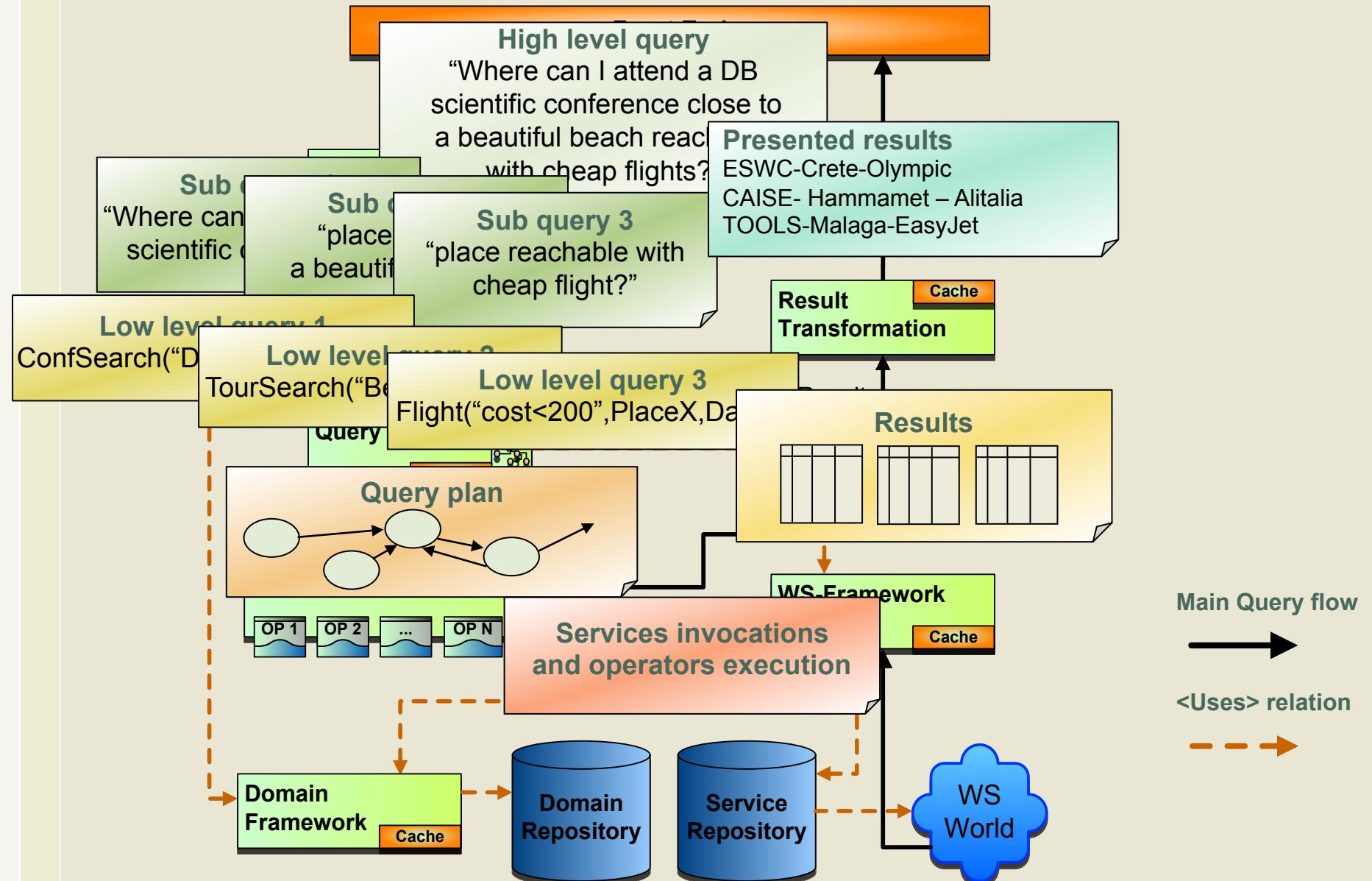
TECHNOLOGICAL FRAMEWORK

Motivating Examples – What Search Engines can't do 17

- “Where can I attend an interesting scientific conference in my field and at the same time relax on a beautiful beach nearby?”

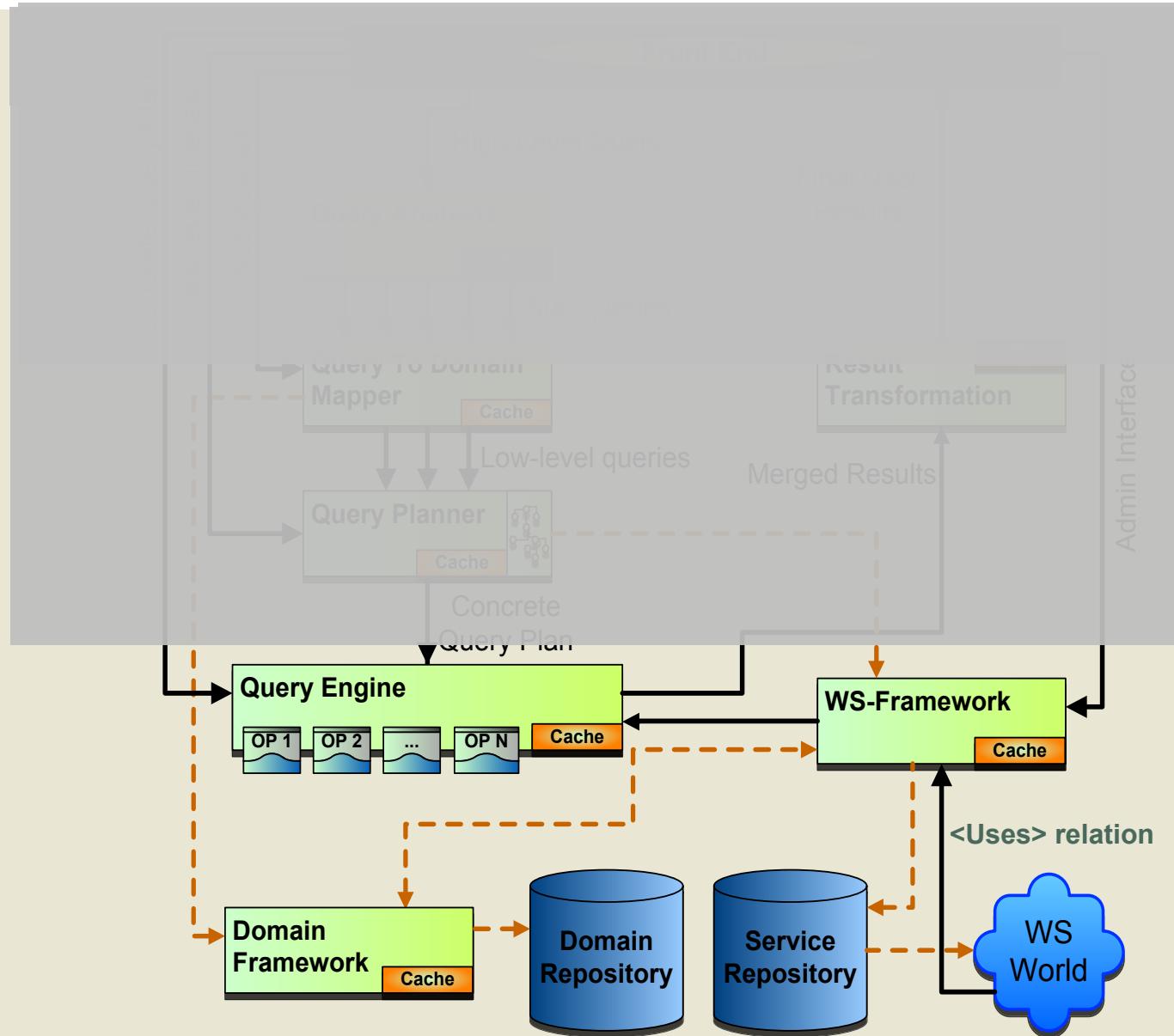
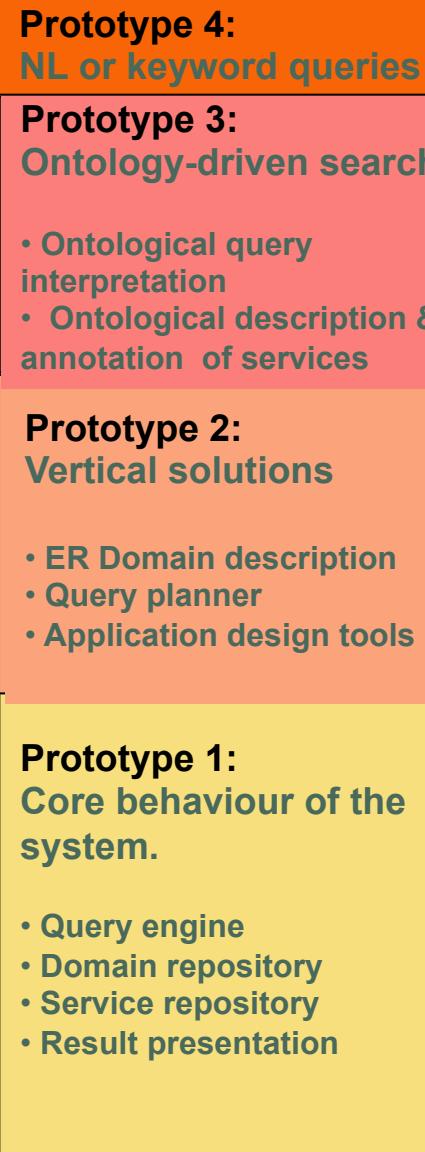
Search Computing architecture: overall view

18



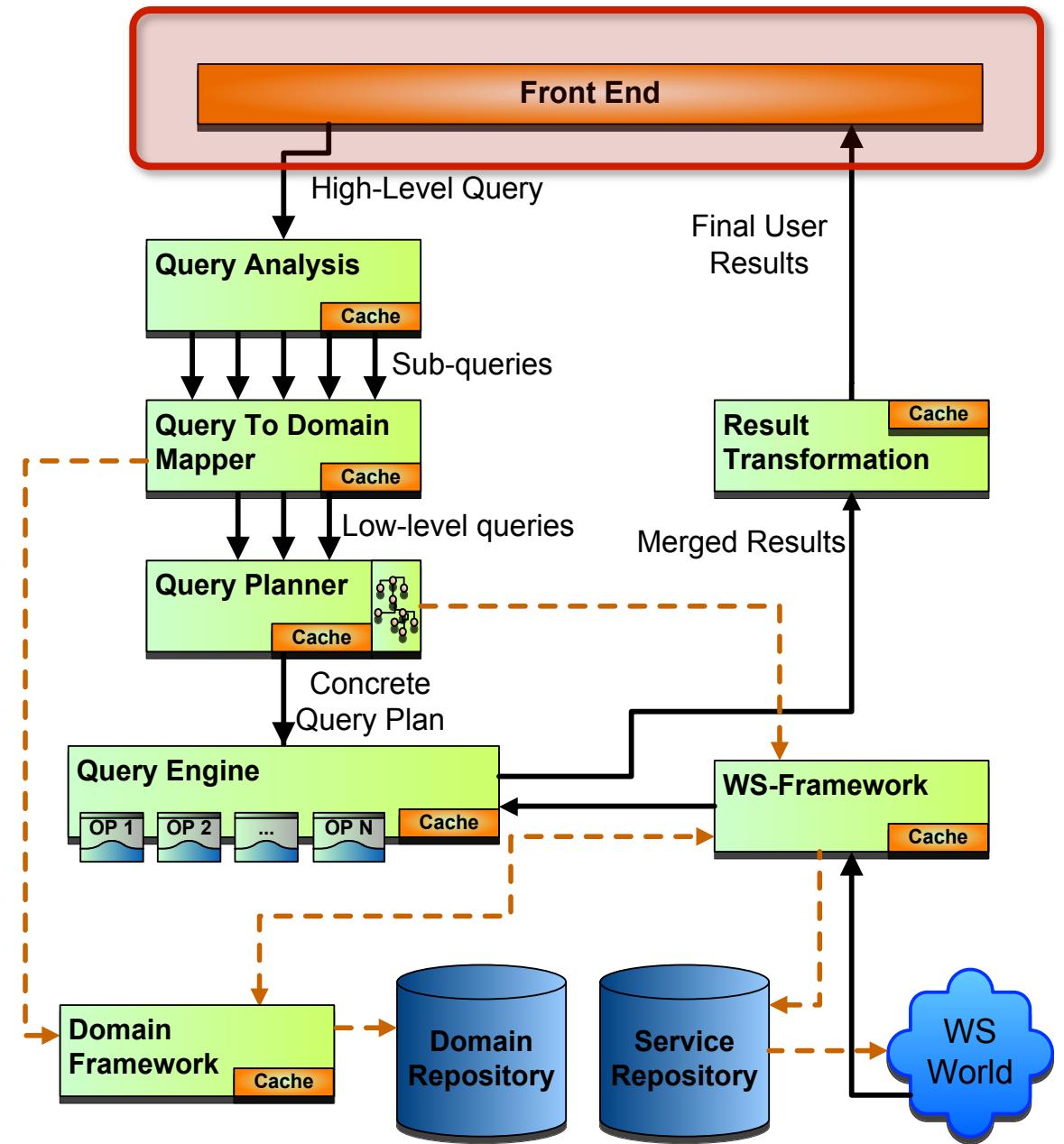
Search Computing architecture: incremental prototyping

19

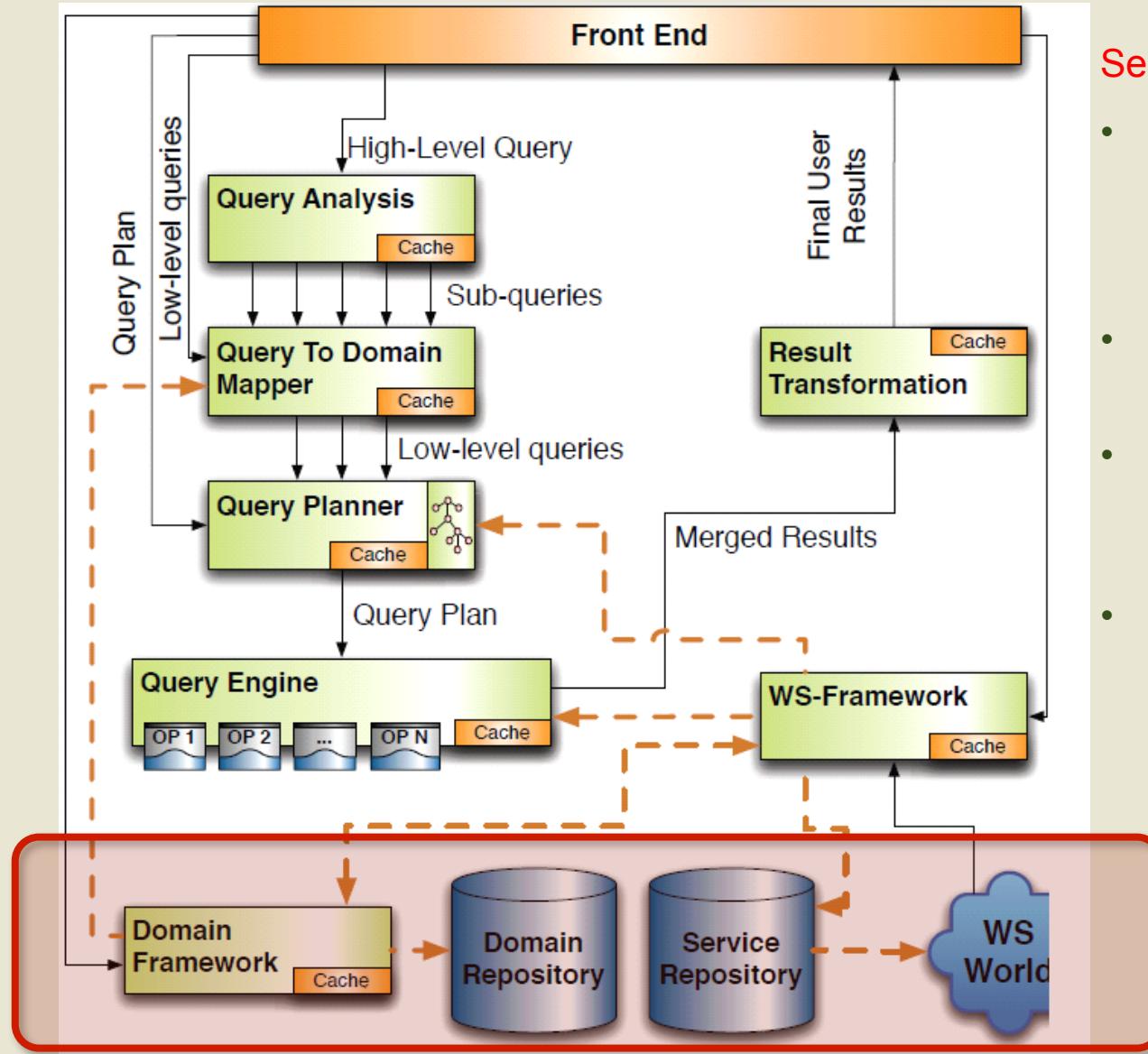


FOCUS ON: UI

- Liquid Query: showing multi-dimensional results
- Exploratory Interface: search as a process



FOCUS on: Service Registration

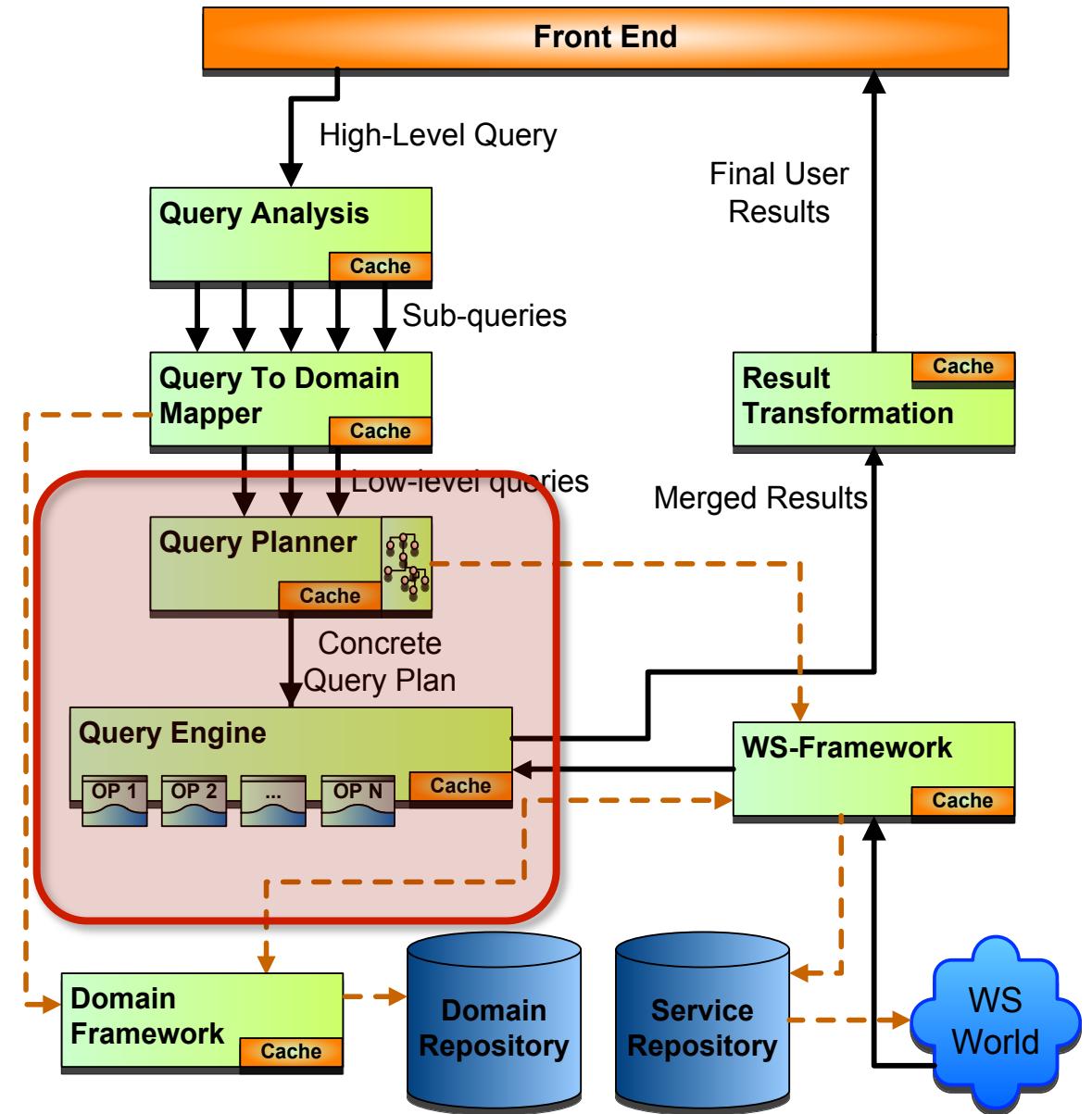


Service Registration

- Conceptual representation of resources as entities and connections
- Logical representation of signatures
- Physical representation as service implementations
- Capturing of service semantics through KB integration

FOCUS ON: Query Engine

- Query planner:
translates
queries into
execution plans
- Query engine:
executes query
plans



SUMMARY OF SECO RESULTS

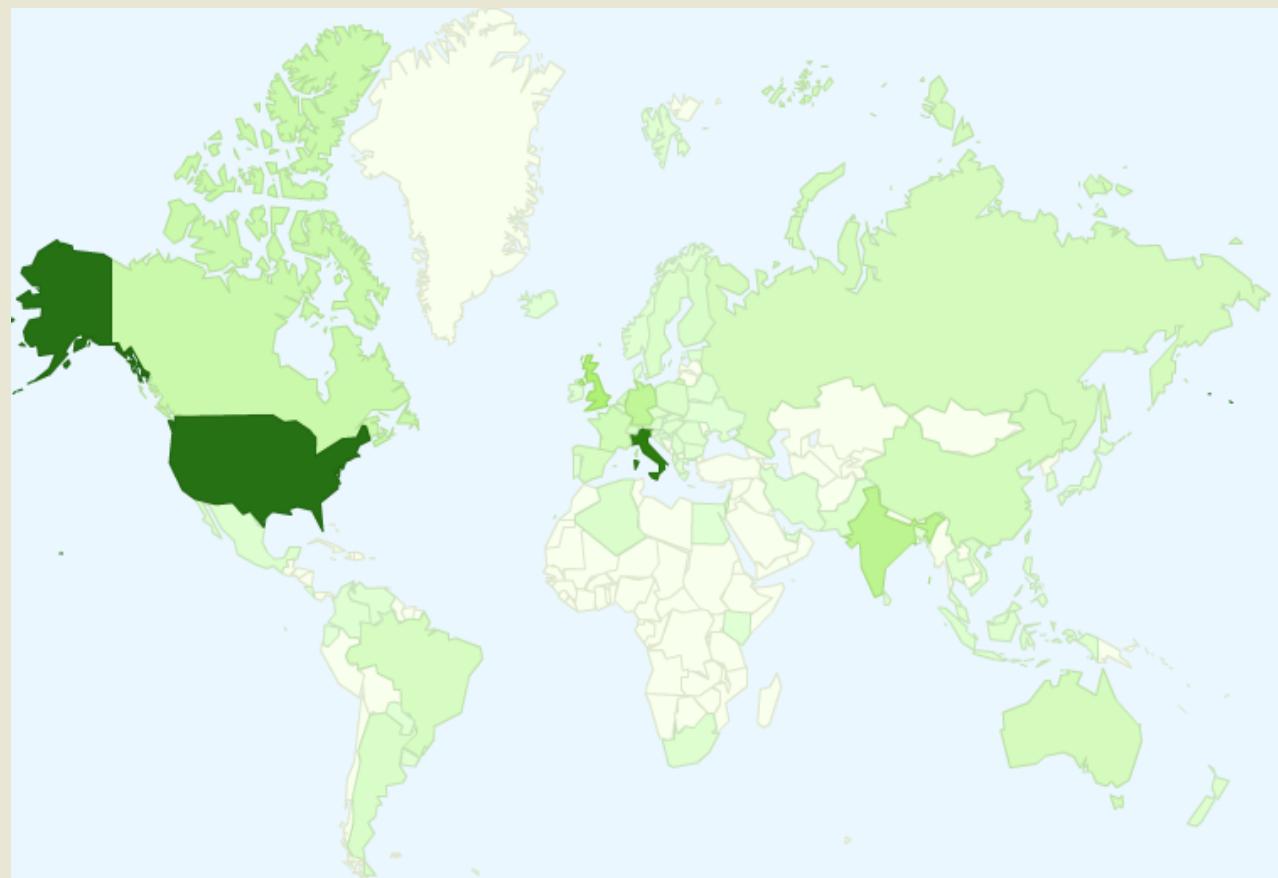
Results after 30 months

- 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)

25

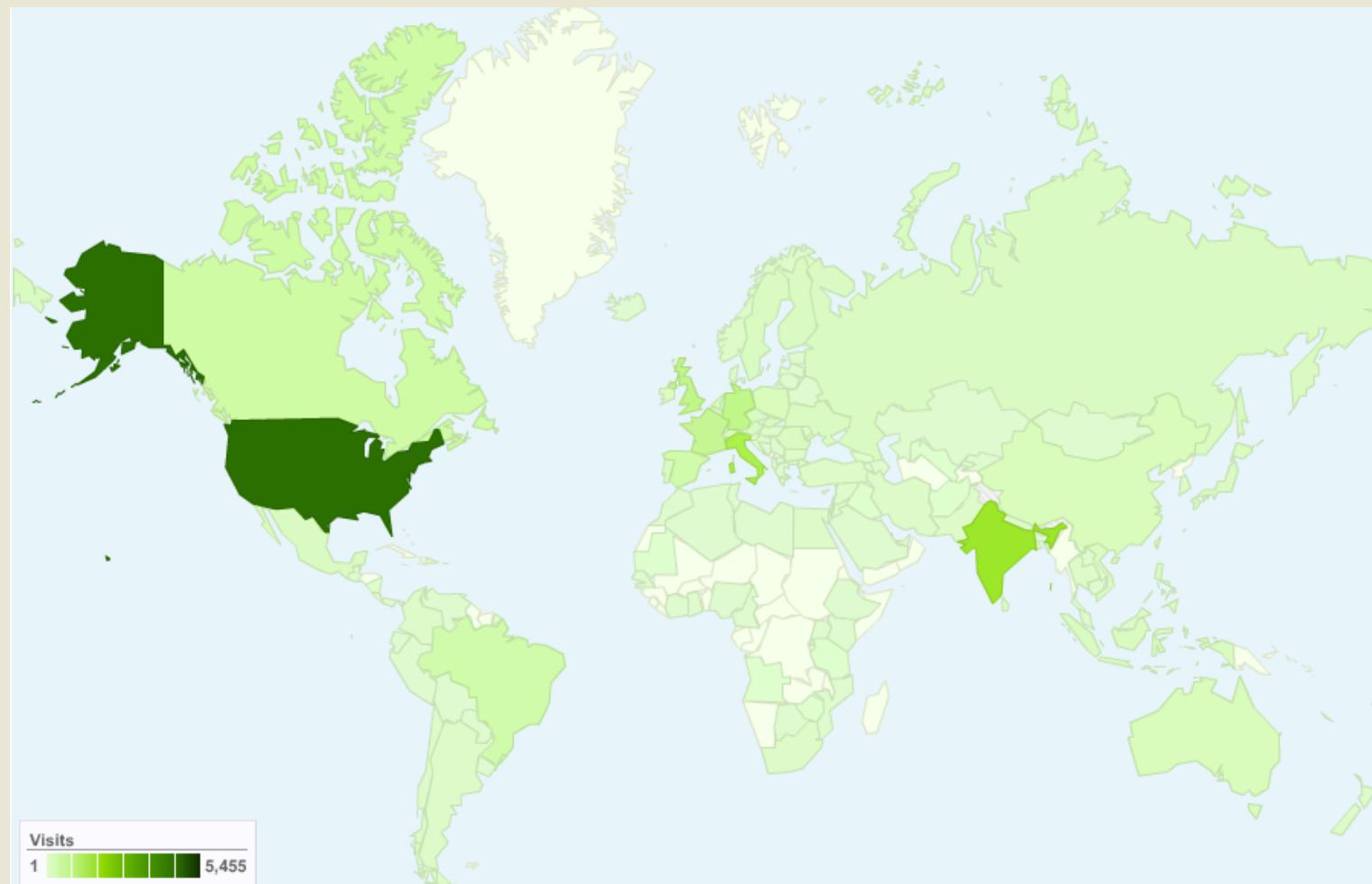
Visits: 20% USA, 18% Italy, 6% UK, 4% India, 4% Canada



Accesses to Web Site & Blog (2011)

26

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



Search Computing First Workshop

June 17-19, 2009

27



Search Computing Challenges and Directions (LNCS, vol. 5950, Ceri-Brambilla eds.)

28

■ Part 1: Vision

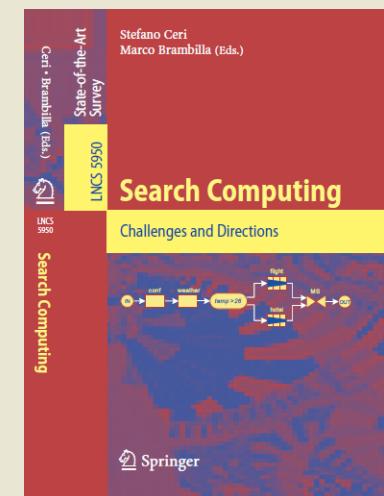
- Ceri: Search computing
- Baeza-Yates: Next generation search
- Weikum: Search for knowledge

■ Part 2: Technology Watch

- Della Valle-Buganza-Gatti: The search engine industry
- Casati-Daniel-Soi: Mashup technologies
- Baumgartner-Campi-Gottlob-Herzog: Web data extraction
- Hedeler-Belhajjame-Campi-Embry-Fernandez-Paton: Dataspaces
- Bozzon-Frernali: Multimedia and multimodal information retrieval

■ Part 3: Issues in Search Computing

- Campi-Ceri-Gottlob-Ronchi: Service marts
- Braga-Campi-Grossniklaus: Join methods and query optimization
- Ilyas-Martinenghi-Tagliasacchi: Rank aggregation
- Braga-Grossinklaus-Ceri: Panta Rhei, a query execution environment
- Brambilla-Ceri-Frernali-Manolescu: Liquid queries and liquid results
- Brambilla-Ceri: Software engineering of search computing applications
- Masseroli-Paton-Sasic: Search computing and the life sciences



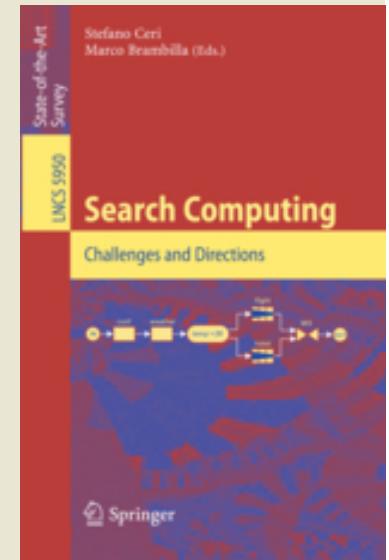
Second SeCo Workshop, May 2010

29



Second Workshop: Sessions

- Pre-Workshop (Milano, May 25)
 - Search as a Process
 - Business Models
- Workshop (Como, May 26-28)
 - Semantic Resource Framework
 - Wrapping Technology and Ontological Annotation
 - Design Tools and Mashup Languages
 - Search Computing and Research Evaluation
 - Query Processing
 - Rank Join
 - Search Computing for Bio-Medical Data Analysis
 - User-Centered Approach to Search Computing Applications
- Post-Workshop (Milano, May 31)
 - Visual Interfaces for Complex Search



Events in 2011

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



Events in 2011

- 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 ?

Future Research Directions

- NLP or keyword-based queries
 - focus on subquery partitioning & mapping to domains
- Panta Rhei evolution
 - adaptation, parallelism
- Application lifecycle
 - tools, best-practices, APIs
- Social dimension
 - crowd-searching to expert communities