Search Computing

November 7, 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
Motivating Examples – What Search Engines can’t do

- “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

Century San Francisco Centre 9 - Union Square - San Francisco, CA

San Francisco, CA 94103. Neighborhoods: Union Square, SOMA .... When I first heard about this movie theatre I thought what an ideal location. ...

Tad's Steakhouse - Union Square - San Francisco, CA

My boyfriend and I had an hour to kill before our movie and we didn't want ...

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. ...

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 ...

Union Square :: Shopping, Dining & Travel Guide :: San Francisco

Union Square Experience the unique character of San Francisco's Union Square ...
Split the task, and search for theaters first
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 46min - 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

<table>
<thead>
<tr>
<th>Theaters</th>
<th>Show Time</th>
<th>Duration</th>
<th>Rating</th>
<th>Genre</th>
<th>IMDb Rating</th>
</tr>
</thead>
</table>
| Century San Francisco Centre 9 and XD | 10:50am | 1hr 40min - Rated PG-13 - Drama - Trailer - IMDb: **★★★★★
  1:00 1:50 4:00 7:00 8:00 10:00 10:45pm** | 10:50am | 1hr 40min - Rated PG-13 - Drama - Trailer - IMDb: **★★★★★
  1:00 1:50 4:00 7:00 8:00 10:00 10:45pm** | 10:50am |
| Tangled in Disney Digital 3D | 11:00am | 1hr 40min - Rated PG-13 - Drama - Trailer - IMDb: **★★★★★
  12:00 1:40 2:40 4:25 5:25 7:10 9:50pm** | 11:00am | 1hr 40min - Rated PG-13 - Drama - Trailer - IMDb: **★★★★★
  12:00 1:40 2:40 4:25 5:25 7:10 9:50pm** | 11:00am |
| The Chronicles of Narnia: The Voyage of the Dawn Treader in Digital 3D | 12:01am | 1hr 52min - Rated PG - Action/Adventure/SciFi/Fantasy - IMDb: **★★★★★
  12:01am** | 12:01am | 1hr 52min - Rated PG - Action/Adventure/SciFi/Fantasy - IMDb: **★★★★★
  12:01am** | 12:01am |
| Love and Other Drugs | 11:18am | 1hr 53min - Rated R - Drama - Trailer - IMDb: **★★★★★
  12:30 2:00 3:30 4:50 6:30 7:40 9:20pm** | 11:18am | 1hr 53min - Rated R - Drama - Trailer - IMDb: **★★★★★
  12:30 2:00 3:30 4:50 6:30 7:40 9:20pm** | 11:18am |
| Tangled | 11:46am | 1hr 40min - Rated PG - Animation/Comedy - Trailer - IMDb: **★★★★★
  10:40pm** | 11:46am | 1hr 40min - Rated PG - Animation/Comedy - Trailer - IMDb: **★★★★★
  10:40pm** | 11:46am |

### 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

Google search results for "steak house san francisco union square"

1. **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 Martini, Real San Francisco.
   - 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
   - 202 reviews

2. **San Francisco Steakhouse Restaurant | Morton's The Steakhouse San Francisco**
   - Morton's The Steakhouse San Francisco is the top steak restaurant and fine dining...
   - 400 Post Street, San Francisco - (415) 986-5830
   - "All in all not really worth the price tag of about $295 for three." - opentable.com
   - 169 reviews

3. **Espetus Churrascaria Brazilian Steak House**
   - 1686 Market St., San Francisco - (415) 552-8792
   - "We enjoyed dinner very much. Pricey, but the quality and flavors were excellent." - opentable.com
   - 289 reviews
Done! Close enough!
(data integration and ranking in the user’s brain)
VISION
The Search Computing Project

- ERC-founded project
- 5 years – Started in 2009, now at month 36
- Build theories, methods and tools to support search-oriented multi-dimensional queries
  - 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

- 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

GOOD, 30 results, 10 calls
GOOD, 30 results, 5 seconds, 50 calls
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 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...
TECHNOLOGICAL FRAMEWORK
Search Computing architecture: overall view

High level query
“Where can I attend a DB scientific conference close to a beautiful beach reachable with cheap flights?”

Sub query 1
“Where can I attend a DB scientific conference?”

Sub query 2
“place close to a beautiful beach?”

Sub query 3
“place reachable with cheap flight?”

Presented results
ESWC-Crete-Olympic
CAISE- Hammamet – Alitalia
TOOLS-Malaga-EasyJet

Low level query 1
ConfSearch(“DB”, placeX, dateY)

Low level query 2
TourSearch(“Beach”, PlaceX)

Low level query 3
Flight(“cost<200”, PlaceX, DateY)

Query plan

Services invocations and operators execution

Main Query flow

<Uses> relation

Domain Framework
Domain Repository
Service Repository
WS World

Database Management
Prof. Stefano Ceri
POLITECNICO DI MILANO
Dipartimento di Elettronica e Informazione
Search Computing architecture: incremental prototyping

Prototype 1: Core behaviour of the system.
- Query engine
- Domain repository
- Service repository
- Result presentation

Prototype 2: Vertical solutions
- ER Domain description
- Query planner
- Application design tools

Prototype 3: Ontology-driven search
- Ontological query interpretation
- Ontological description & annotation of services

Prototype 4: NL or keyword queries

Diagram:
- Query Engine
- Domain Framework
- Service Repository
- Domain Repository
- WS-Framework
- Admin Interface
- Result Transformation
- Low-level queries
- Merged Results
- Concrete Query Plan
- Sub-queries
- Concrete Query Plan
- Sub-queries
- Final User Results
- <Uses> relation
LIQUID QUERY INTERFACE
Liquid query definition

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

\[ = \text{inputs, outputs} \quad + \quad \text{GR} = \text{global ranking} \]
Liquid query definition

... And then characterizing the user interaction

Plus:

- Parametrization of global ranking
- Data visualization options
- .. and so on
Exploration of the Service Space

Entity Selection

- Real Estate: Sale and Rent of Real Estate
- Conference: list of conference
- 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
Exploration of the Service Space

Service Selection

Entity Selection
Exploration of the Service Space

Entity Selection

Service Selection

Query !!
Result Presentation

Tabular Representation

| Local Filter Projection
| Order
| Ranking Bar

---

<table>
<thead>
<tr>
<th>Tuple ID</th>
<th>Global Score</th>
<th>tupledId</th>
<th>tupleScore</th>
<th>state</th>
<th>zipCode</th>
<th>city</th>
<th>street</th>
<th>longitude</th>
<th>latitude</th>
<th>type</th>
<th>uscOl</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.41</td>
<td>gsp8CKMaYll9YlyEni48RHX 0.41</td>
<td>NY</td>
<td>11237</td>
<td>Brooklyn</td>
<td>Jefferson St</td>
<td>73.92305755615234</td>
<td>40.70718002319336</td>
<td>makeMeMove</td>
<td>Multi Family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>kSat91vvCVI0vQHdGvshHzhC 0.4</td>
<td>NY</td>
<td>11221</td>
<td>Brooklyn</td>
<td>Suydam St</td>
<td>73.9254081542969</td>
<td>40.6999643066406</td>
<td>makeMeMove</td>
<td>Multi Family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>0e18fizbqXGNalsakZVeibGc 0.35</td>
<td>NY</td>
<td>11211</td>
<td>Brooklyn</td>
<td>Devoe St</td>
<td>-73.93654632563836</td>
<td>40.7139892578125</td>
<td>makeMeMove</td>
<td>Single Family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>J1Ln1JucS9iGpEqkT2cgo8G 0.25</td>
<td>NY</td>
<td>11211</td>
<td>Brooklyn</td>
<td>Leonard St APT J3</td>
<td>73.94757080078125</td>
<td>40.71459579467734</td>
<td>makeMeMove</td>
<td>Condo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.205</td>
<td>AN9oZdNK5Hnq,JnGF5ZX4Y 0.205</td>
<td>NY</td>
<td>11211</td>
<td>Brooklyn</td>
<td>Humboldt St APT 1</td>
<td>73.94300842285156</td>
<td>40.717288970947266</td>
<td>makeMeMove</td>
<td>Condo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>eB1LYR9znwIE3SnWJTPQk 0.2</td>
<td>NY</td>
<td>11211</td>
<td>Brooklyn</td>
<td>242 S 1st St APT 1C</td>
<td>73.95755686035156</td>
<td>40.712318420410156</td>
<td>makeMeMove</td>
<td>Condo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Result Presentation (Map)
Exploration options from a given state

Related Entities

Object list

Expansion from Zillow By Coord (1)
- School: School institutes and academics
- Walkscore: Index of walkability for a location
- Hotel: Hotel, B&B, inn etc
- Doctor: Doctors and Medical Specialists
- Theatre: Theatre and Cinema
- Event: A social event
- Rented Car: Car Rental Services
- Statistics: Statistics and Census information
- News: News articles
- Job: Jobs and open positions
- Restaurant: Restaurants and Diners

Service list for: Doctor By City

- doctor.com by city: This service returns available doctors, including geographical coordinates, nearby a given location.
Result Presentation (Atom View)

Association

Real Estate Service

Doctor Service
Result Visualization – Combinations on Maps
SERVICE REGISTRATION
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...

User Inputs

Concert
- "San Francisco"
- 1 mile
- 10/03/2010
- 12/03/2010
- "Jazz"
- location
- radius
- minDate
- maxDate
- genre
- name
- date
- lat
- long
- distance
- price
- address

Restaurant
- 1 mile
- "Vegetarian"
- 3 stars
- radius
- category
- minRating
- location
- name
- address
- rating
- lat
- long
- distance
- url

Hotel
- 1 mile
- "Hotel"
- 3 stars
- location
- radius
- category
- minRating
- name
- phone
- rating
- lat
- long
- distance
- url
Service Framework in SeCo

- A Service Description Framework coupled with a Semantic Annotation Framework

- Conceptual representation: group services by core entities
- Logical representation: i/o fields and transitions as domain entities/relationships
- Physical representation: as shipped by data provider

- Capturing of service semantics via Knowledge Base lookup
- Entities/relationships “mentioned” in SDF
Semantic Framework: Domain Diagram and Access Patterns

TheatrebyMovie

ActorByTitle

MovieByTitle

PrizeByDirector

Actor

Movie

Theatre

Current Date

Current Position

Film_Director

Prize

Domain concept

Access Pattern
SECO ENGINE
The Query Processor in the Big Picture
The Query Processor in the Big Picture

Browser
- Liquid queries UI
  - Queries
    - API
      - Logical queries
    - Query Analyzer
      - Low level queries
    - Query Planner
      - Execution plans
    - Execution Engine
      - Service calls
    - Service Invocator
      - Invocations
      - Wrapper Environment
        - Adapted calls
    - Existing Services

Query Engine
- Interaction Designer Tool
- SeCo Developer
- SeCo Expert
- Service Registration Tool
- Service Publisher
- Service Developer
- Compiled Queries
- Query Management Tool
- Workbench testing tool
- Wrapper Deployment Tool
- Wrapper definitions
- Services
- UI definitions
- Final User
- Main Data Flow
### SeCoQL

```
DEFINE QUERY Nightclub ((Address : String, City : String, Country : String)) AS
  SELECT *
  FROM theatre, address, city, country, manager
  WHERE manager.manager = theatre.manager
  AND manager.address = theatre.address
  AND manager.city = theatre.city
  AND manager.country = theatre.country
  AND theatre.country = country.country
  AND theatre.city = city.city
  AND theatre.address = address.address
  AND theatre.name = 'Nightclub'
  AND manager.name = 'Manager'
  AND manager.email = 'ManagerEmail'
  AND manager.phone = 'ManagerPhone'
  AND theatre.year = 2020
  RANK BY (T + 0.4, N + 0.5, W = 0.1)
  LIMIT 10
```

### Logical Level

- **Movie**
- **Restaurant**
- **Theatre**

### Physical Level

- **SeCoQL**
- **SeCo Developer**
- **SeCo Expert**
- **Service Registration Tool**
- **Service Invocator**
- **Query Management Tool**

### Query Processor

- **Query Engine**
  - **Query Analyzer**
    - Low level queries
  - **Query Planner**
    - Execution plans
  - **Execution Engine**
    - Service calls
  - **Service Invocator**
    - Invocations
      - Wrapper Environment
        - Adapted calls
      - Existing Services
        - Wrapper definitions

### Workbench testing tool

- **Logical Level**
  - **Interaction Designer Tool**
  - **Query Definitions**
  - **Compiled Queries**
  - **IN**
  - **OUT**
  - **IN**
  - **OUT**

### Panta Rhei

- **Aeroflot**
- **Airport Taxi/Travel**
- **Ticketmaster**
- **HRS**
An old drama movie showing tonight in a theatre close to a good restaurant

NightOut(Piccadilly, London, UK)

DEFINE QUERY NightOut ( $Address : String , $City : String , $Country : String ) AS
SELECT *
FROM theatre_by_address ( iAddress : $Address , iCity : $City , iCountry : $Country ) AS T USING Google_Theatre_by.Addr
JOIN Movies_by_Genre ( iGenre : Drama ) AS M USING IMDB_Movies_by_Genre
  ON I.title = M.title
JOIN Restaurant_by_Address ( iAddress : T.address , iCity : $City , iCountry : $Country ) AS R USING GMaps_Rest_by.Addr
WHERE M.year < 2008
RANK BY ( T = 0.4 , R = 0.5 , M = 0.1 )
LIMIT 5 TUPLES AND 4 CALLS
First step: from conjunctive queries to logical plans

```
DEFINE QUERY NightOut ( $Address : String, $City : String, $Country : String ) AS
    SELECT *
    FROM theatre_by_address ( iAddress : $Address, iCity : $City, iCountry : $Country ) AS T USING Google_Theatre_by_Addr
    JOIN Movies_by_Genre ( iGenre : Drama ) AS M USING IMDB_Movies_by_Genre
    ON T.title = M.title
    JOIN Restaurant_by_Address ( iAddress : T.address, iCity : $City, iCountry : $Country ) AS R USING GMaps_Rest_by_Addr
    WHERE M.year < 2008
    RANK BY ( T = 0.4, R = 0.5, M = 0.1 )
    LIMIT 5 TUPLES AND 4 CALLS
```
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

Query Lifecycle

Query Execution Report
THEORY
Proximity Rank Join (Martinenghi-Tagliasacchi, VLDB 2010)

Order by
- individual scores
- proximity from query
- inter-object proximity
and return the top-K results

Too many results!
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
Results after 36 months

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

- **Research results**
  - LNCS: Search Computing Challenges and Directions (2010)
  - Publications (incl. VLDB, WWW, SIGMOD, TODS)
  - US Patent filed (top-k method, random & sequential services)

- **Site:** [www.search-computing.eu](http://www.search-computing.eu)

- **Blog:** [http://blog.search-computing.com/](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
Accesses to Web Site & Blog (2011)

Visits: 27% USA, 10% India, 8% Italy, 4% Germany – UK - France
Events in 2011

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

问题?