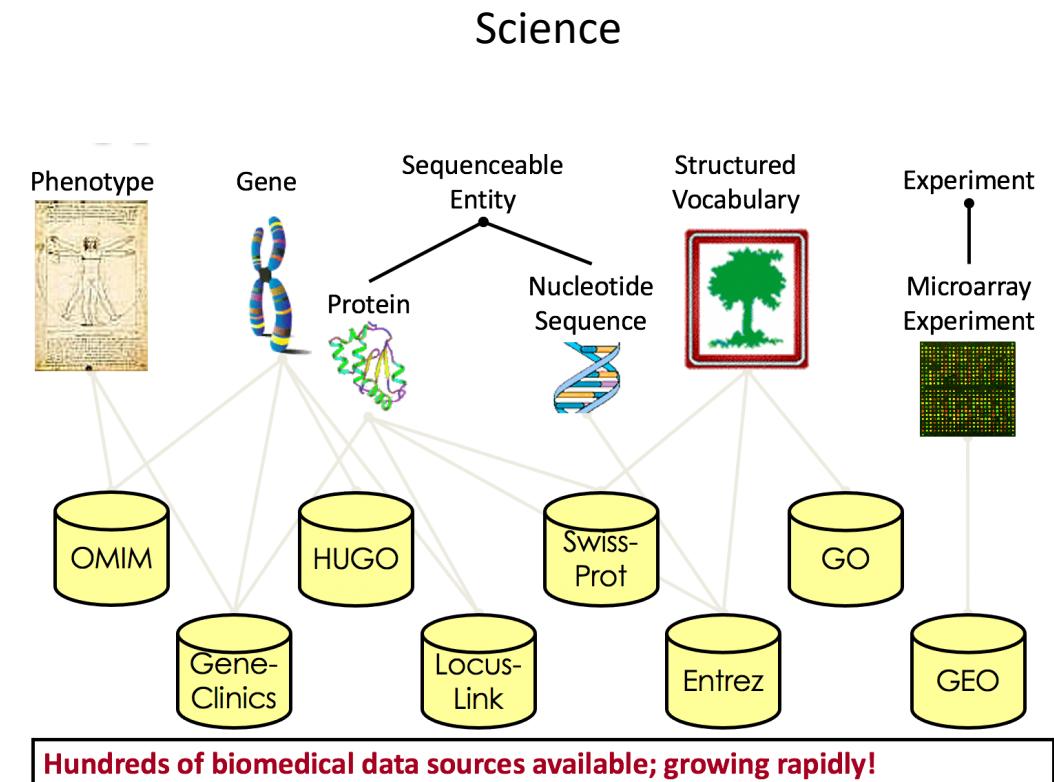
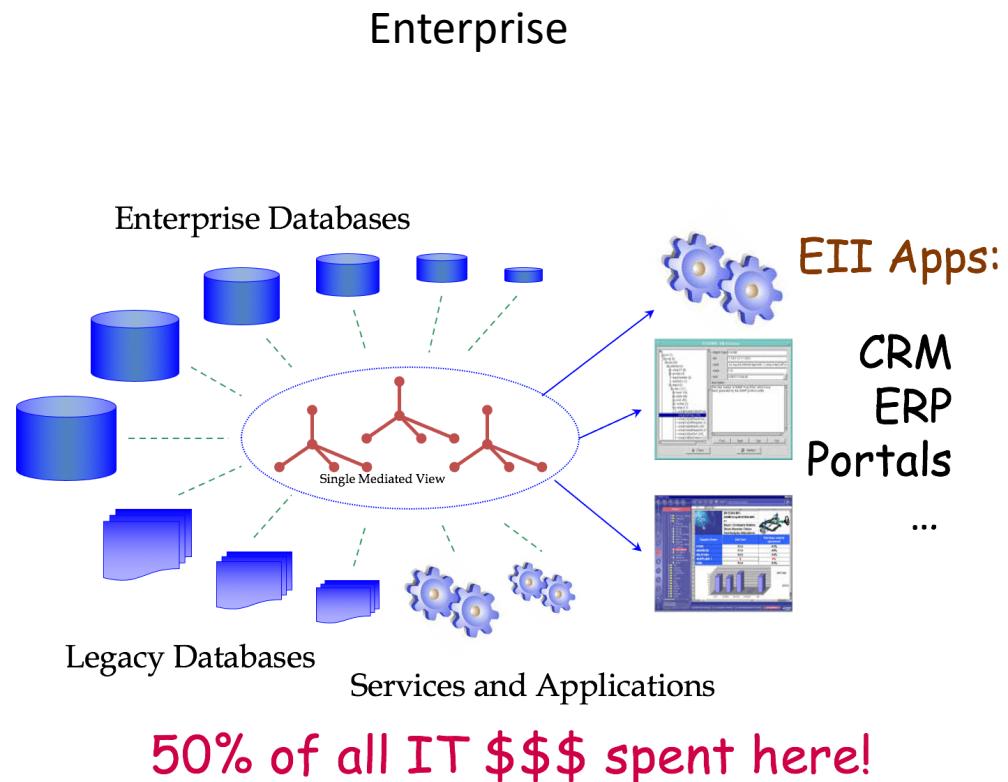


Information integration on the web

Mayank Kejriwal

Introduction and applications

Before going into definitions, let's look at some intuitive applications of 'information integration'



Web and deep web

Historically...



Today



...

Airfare prediction (very Web-specific example)

The screenshot shows the Kayak website interface for flight search. At the top, there's a navigation bar with tabs for HOTELS, FLIGHTS, CARS, and PACKAGES. The main search area displays a flight from LAX to IAD on Jan 21 (depart) and Jan 23 (return), in Economy cabin for 1 traveler. The search results are sorted by price (low to high), showing 554 of 612 flights.

Daily Low Fare History

Advice: Wait Confidence: 79%
Prices may fall within 7 days

Stops

- nonstop \$558
- 1 stop \$398
- 2+ stops \$529

Times

Take-off Los Angeles (LAX)
Wed 12:00a - Thu 12:00a

Take-off Washington (IAD)
Fri 5:00a - 10:00p

Tip: Buy - Fares rising or steady
Lowest fare: \$291 · Average fare: \$48

LAX ↔ IAD Jan 21 depart → Jan 23 return Economy cabin 1 traveler Change

Sort by: price (low to high) ▾ 554 of 612 flights show all Round-trip | Segment NEW

\$558 Fly with Virgin America

Fly with WiFi, on-demand food and drinks, live TV, movies, music, and more.

Select nonstop Virgin America \$558 Select
1 stop Virgin America \$566 Select

\$398 JetBlue Airways / Virgin America

Hacker Fare jetBlue 12:45p LAX → 11:00p IAD 7h 15m 1 stop (JFK) nonstop
9:00a IAD → 11:50a LAX 5h 50m

Show details ▾ Economy

Ad

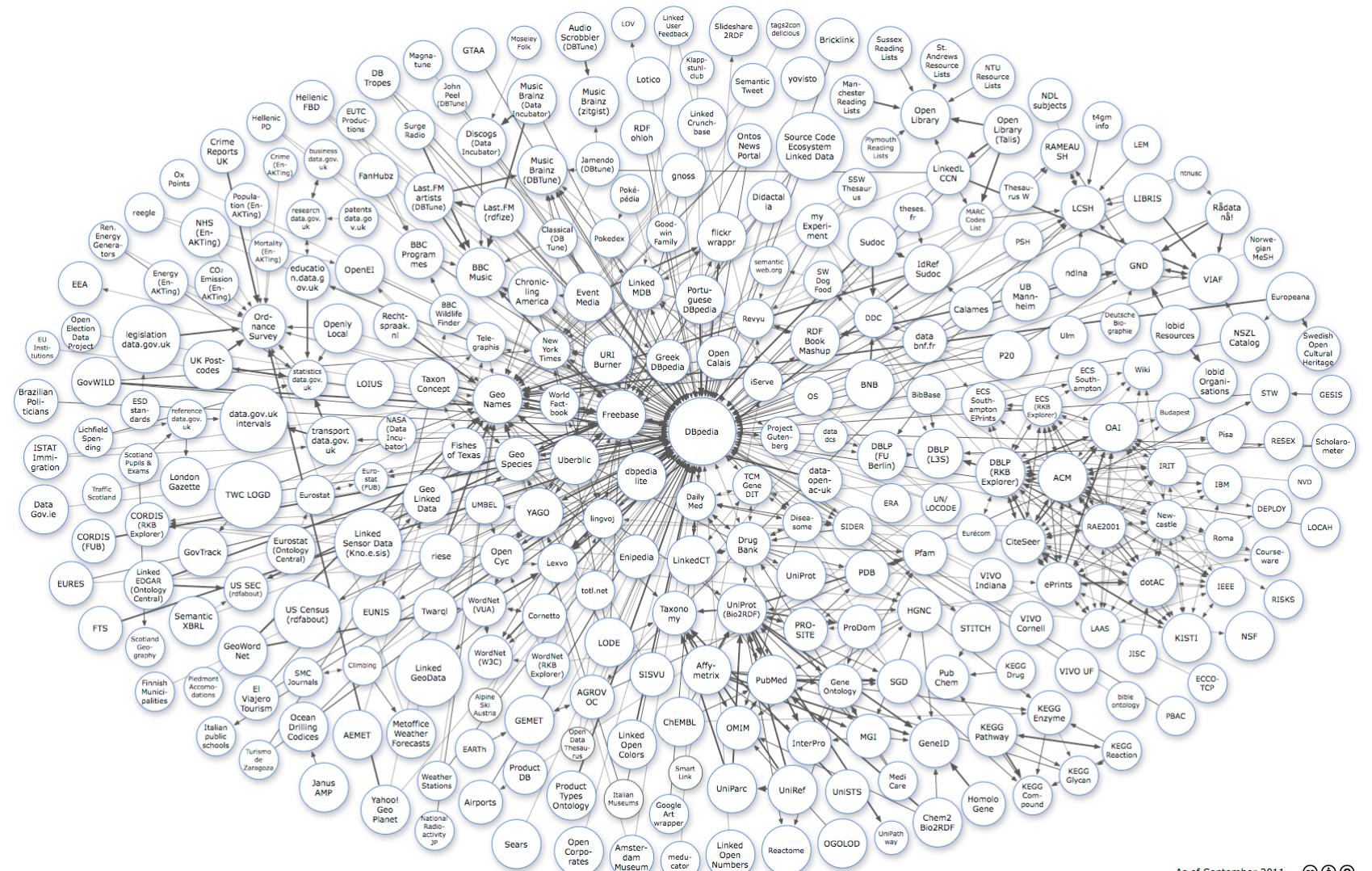
ibl. Book Today.

\$550
33
More filters ▾

is Price ▾

\$323
Book

Semantic Web



SATURDAY, MARCH 13, 2004

GOOGLE TO HARVEST RDF-A

Search giant goes semantic

Norris Center for the Performing Arts, 27579 Crossfield Drive in RHE, on Saturday, March 20 at 2 and 4 p.m. Tickets are \$38 for adults and \$18 for students for the evening performance and \$30 for adults and \$15 for students at the matinee performance. For reservations, call 544-0403.

* UPCOMING — The Palos Verdes Peninsula Unified School District and Friends of School Music host the 15th Palos Verdes Elementary Choral Festival on March 23, 24 and 25 at the Norris Center for the Performing Arts, 27579 Crossfield Drive in RHE. All shows begin at 7:30 p.m. For tickets, call the Norris box office at 544-0403.

* ONGOING — The Distinctive Edge, 29050 S. Western Ave., Suite 113 in RHE, continues "Third Time's a Charm," an exhibit of 3-D Collages by artist Steve Jacobsen, through March 30. For gallery hours, call 833-3613.

* ONGOING — "Natural Treasures" exhibition contin-

Two years ago, U.S. Navy personnel and their families assigned to the Atsugi Navy base, home of the U.S. Kittiwake, were treated to a rare experience when Terry Fleming and his local Irish/American band, Innisfree, travelled to the base to entertain them on St. Patrick's Day. Fleming and the other five members of Innisfree were delighted and honored to be able to go to Japan and lift the spirits, if only for a few hours, of the Navy personnel and their families.

For the third year in a row, Fleming — a local insurance broker in Rolling Hills by day and an entertainer by night — and the band travel to entertain the Navy men, women and families at various bases throughout Japan.

Fleming, the leader of the band on accordion and harmonica, actually is the only member of the band from Ireland. Other members include lead singer Julie Delaney, a civil engineer in Newport Beach; Terry Doyle, guitarist, a news director with CBS news; Dennis Doyle, Celtic harpist, a professor at Glendale College; Kevin Wied, keyboards; and bugger, music

teacher and amateur director of the Orange County Symphony; and Mike Tiffany, bass, a computer engineer. The band has been playing the length and breadth of California for the past 25 years. They have played at graduations, weddings, birthdays and on occasions where there was little excuse for throwing a party.

Fleming says it was by coincidence the band got the opportunity to travel to Japan. Another band was unable to travel at the last minute and so he and his band were offered the opportunity to go in their place.

With some trepidation they made their first trip and with the overwhelming response they had were quickly allayed. On a damp St. Patrick's Day, hundreds of families, clad in many shades of green, whooped it up,

shaking their hearts out and danced up a storm. As the evening wore on, many in the audience were emboldened to try their hand or foot at the Irish jig, with much encouragement from the band.

Even though far from home, the Atsugi base — situated a few hours

south of Tokyo — felt like home away from home, with its lush green rolling landscape and its multitude of cherry blossom trees. "Yes," Fleming says, "one was struck by the commitment and dedication of our men and women in uniform as they played their part in protecting and serving in the world."

A couple of the families even took time out from their busy schedule to host Terry and the band members. They treated them to a guided tour of the base and accompanied them on a few exciting trips off the base, visiting beautiful ancient temples, monuments and several giant Buddhas.

On a visit to downtown Tokyo, the band came across what they assumed was a very rare sight, a place called "Scruffy Murphy," an Irish pub located in the heart of a bustling downtown. Upon checking the establishment out, they discovered a real authentic Irish Pub with excellent Guinness and good pub grub. It also happened to be open mic night, so the band members took over the stage

and entertained the locals for a few fun-filled hours. It turned out that it was just one of many establishments in the city.

A special bond developed between the band members and these families and already exchange visits have occurred when the same families were on leave in the United States.

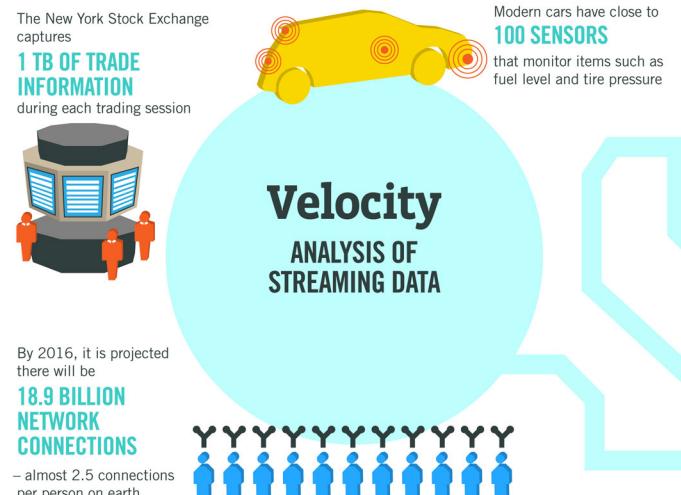
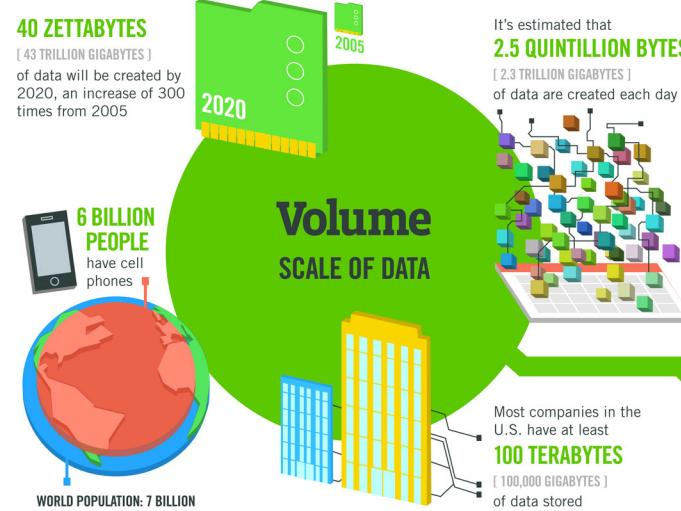
For more information about the band, log on to www.bellisfreemusic.com.



<http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html>

slide from Frank van Harmelen

Big Data



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States.

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES [161 BILLION GIGABYTES]

30 BILLION PIECES OF CONTENT

are shared on Facebook every month



Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month



400 MILLION TWEETS are sent per day by about 200 million monthly active users

1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate



Veracity UNCERTAINTY OF DATA

Challenges

Why is information integration challenging...

- Accessing the data
- Understanding the data
- Resolving differences
 - Schema-level
 - Data-level
- Efficiency
 - Query evaluation
 - Data transformation

...and what makes it even more challenging
on the Web?

- The Web is an incredible source of data
- New challenges arise:
 - Need to turn web pages into structured data
 - Don't have control over the data
 - Sources have input/output constraints
 - Distributed nature of the web can make integration slow
 - How to relate information across sources

Going beyond simple keywords is still a difficult problem in Web Search

- Try Googling answers to these questions
 - Which US cities have more than 500,000 people?
 - Which congress members voted “No” on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts?
 - How many Twitter followers does Obama have in each state?
 - Find all stores that sell a Canon 7D for less than \$1,700

Architectural overview

Typical Web information integration architecture

