Research @ Google

Dr. Alfred Z. Spector
VP, Research and Special Initiatives
Google, Inc.
Asia, September 2009

Abstract

Research @ Google

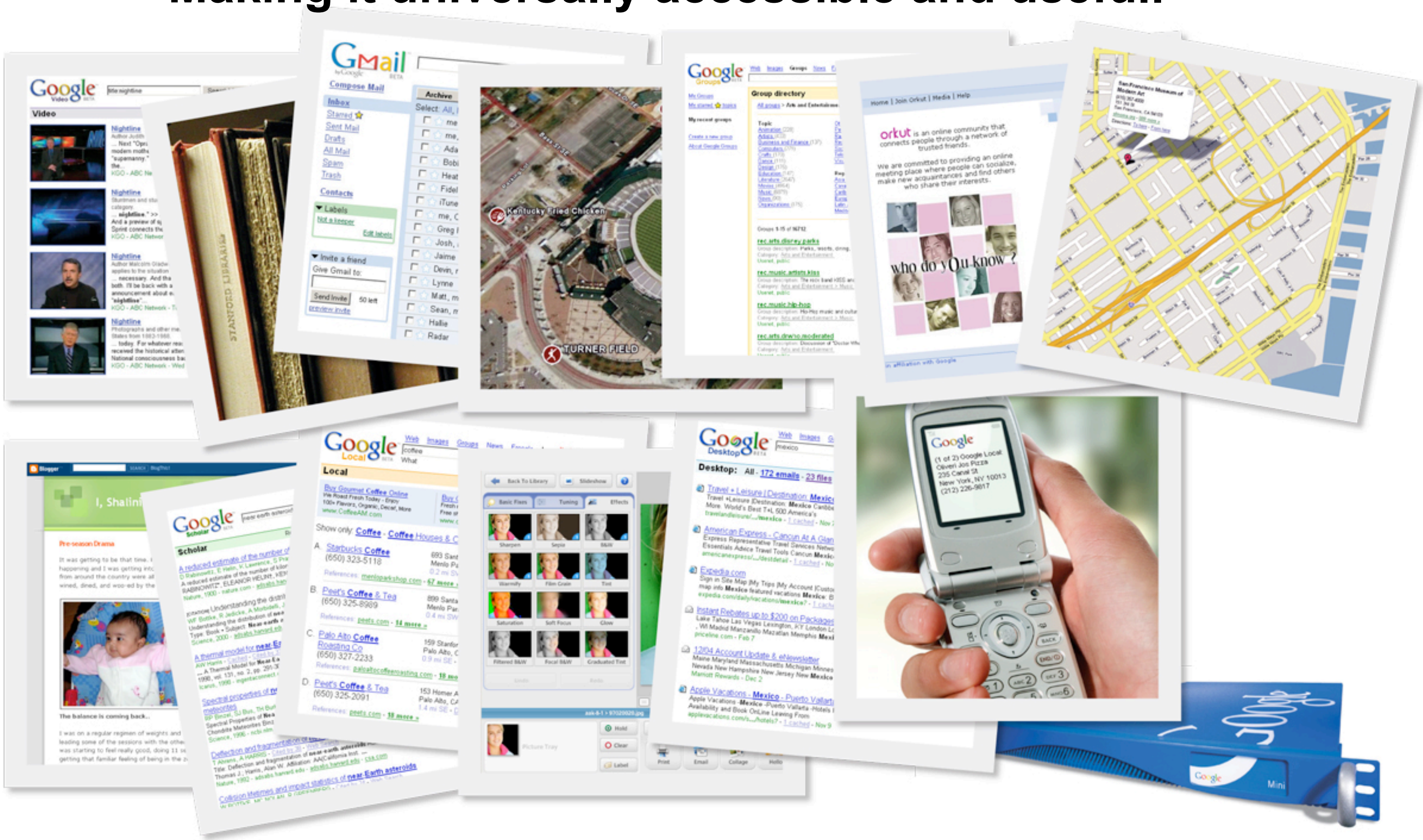
At its core, Google's mission is to organize the world's information and make it universally accessible and useful. The breadth of this mission, coupled with our services based delivery model, provides Google great opportunities to perform research and to innovate in many areas of technology. In this presentation, I'll summarize our approach to innovation and the results we have achieved in many domains; for example, translation, speech, and vision. I'll also discuss some of our focus areas moving forward, and our general approach to research organization. I'll conclude by discussing our interactions with the research world around us, a world with which we desire strong, mutually beneficial connections.

Outline

1. Google's Mission
2. Our Technical Approach
3. Innovation
4. Research @ Google
5. Key Themes
   A. Totally Transparent Processing
   B. The Rule of Distributed Computing
   C. Hybrid, not Artificial, Intelligence
6. Some Other Research Projects
7. Relationship with Academe and other Research Organizations
8. Summary and Perspective on Computer Science

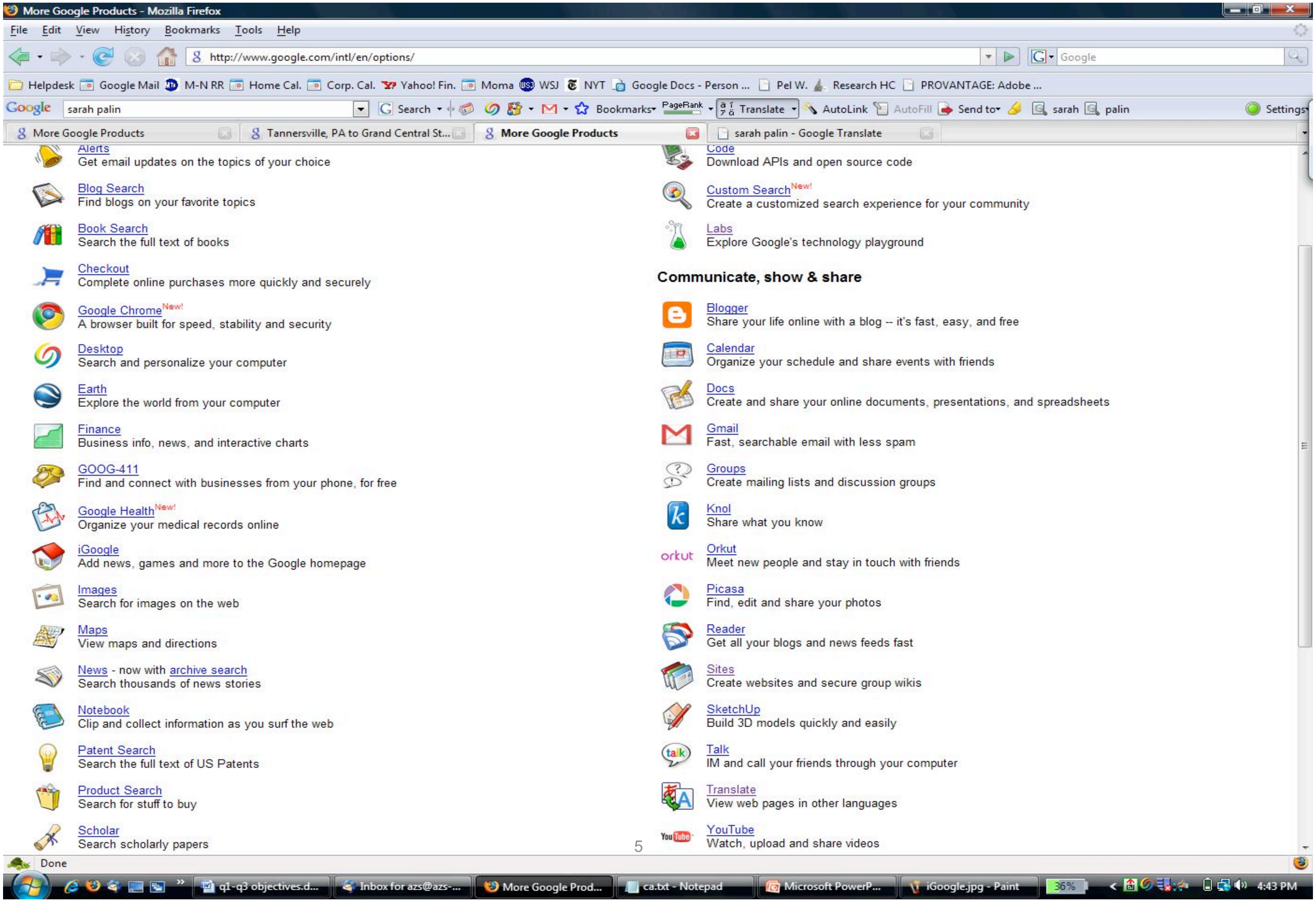
Google's Mission and Google Research

Organizing the world's information and
Making it universally accessible and useful.



A research organization optimized for in situ work

Search and more…



How (1): A Focus on Services

- **Google primarily delivers services to achieve it's mission**
- **Implications**
  - Lower cost of Development
  - Economies of scale
    - Lower cost of Installation
    - Lower cost of Operation
  - Resilience
  - Location transparency
  - Service Integration
  - Aggregated user feedback
  - …

How (2): The Google Common Distributed System

- **Vast**:
  - Data in the cloud
  - Processing in the cloud
  - Global Usage

  (Feasible due to large clusters using decades of distributed computing research)

- **Implications:**
  - Economies of scale from *shared* infrastructure
  - Low barriers to product launch
  - Decentralized development more feasible

Google™

How (3): <u>**Empiricism**</u> - Let Measurement & Feedback Rule

-0.54

2.7M

0.55060

108 milliseconds

$4.78 RPM

-0.0000339

2,800,000,000 views

425,440.01

56.76%

17.35

108 seconds/search

9995.55

1.3 searches per user

480,000,000 total pageviews

1607.44

10,400

6.55

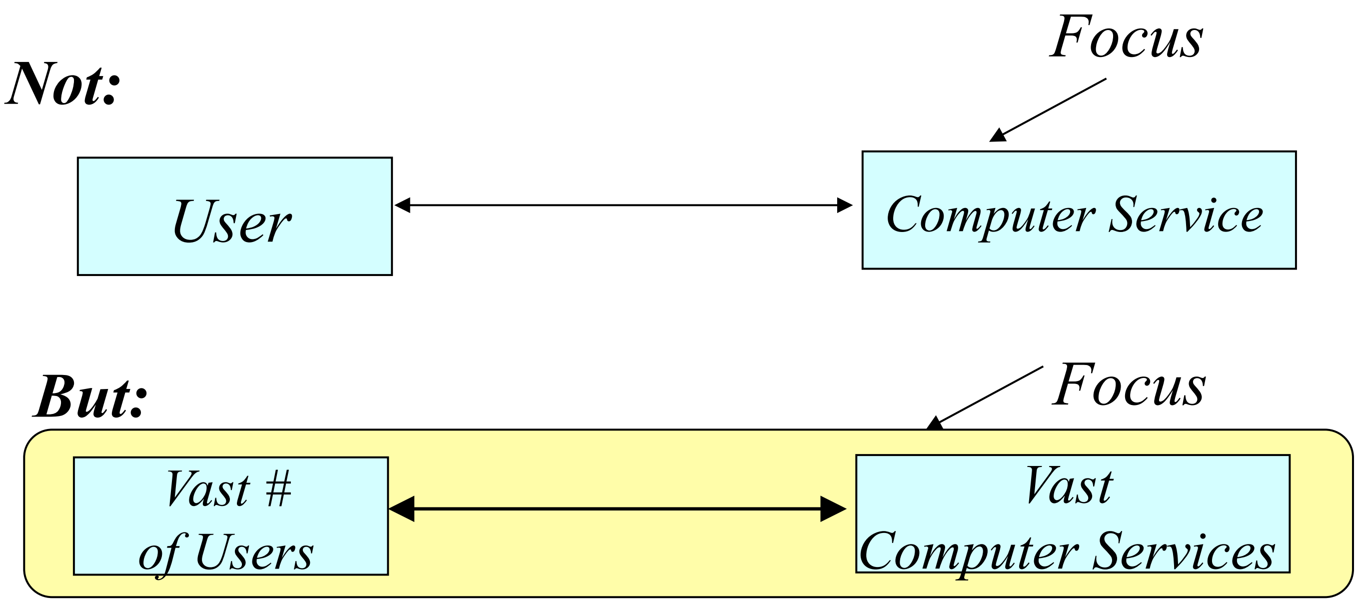
$0.303 CPA

108

$7,660,400

Google™

Key Change: Holistic Approach To Design



- **Implications**
  - Users and computers doing more than either could individually.
  - Virtuous circle from: ***Data and Processing, Reach, Feedback in a virtuous circle.***

How Do We Innovate…

- Commitment to advancing technology
  - Rich domain of work due to our mission
  - Some exciting, *grand challenge* problems
- Cultural agreement that getting a concept into production is often as challenging and fun as its initial invention
- Technical leverage
  - The Google Common Distributed System
  - A Focus on Services
  - Holistic Approach to Design

Google's Research Mission

To innovate, and to catalyze innovation, and to learn in ways that collectively help Google achieve its mission

Implications:

- Operation in areas relevant to Google
- Broad applicability of many areas of CS and related areas
- A diversified portfolio (various points on risk/reward curve)
- Strong relationship to academic community
- Strong emphasis on publication

Google’s Innovation Culture is Different

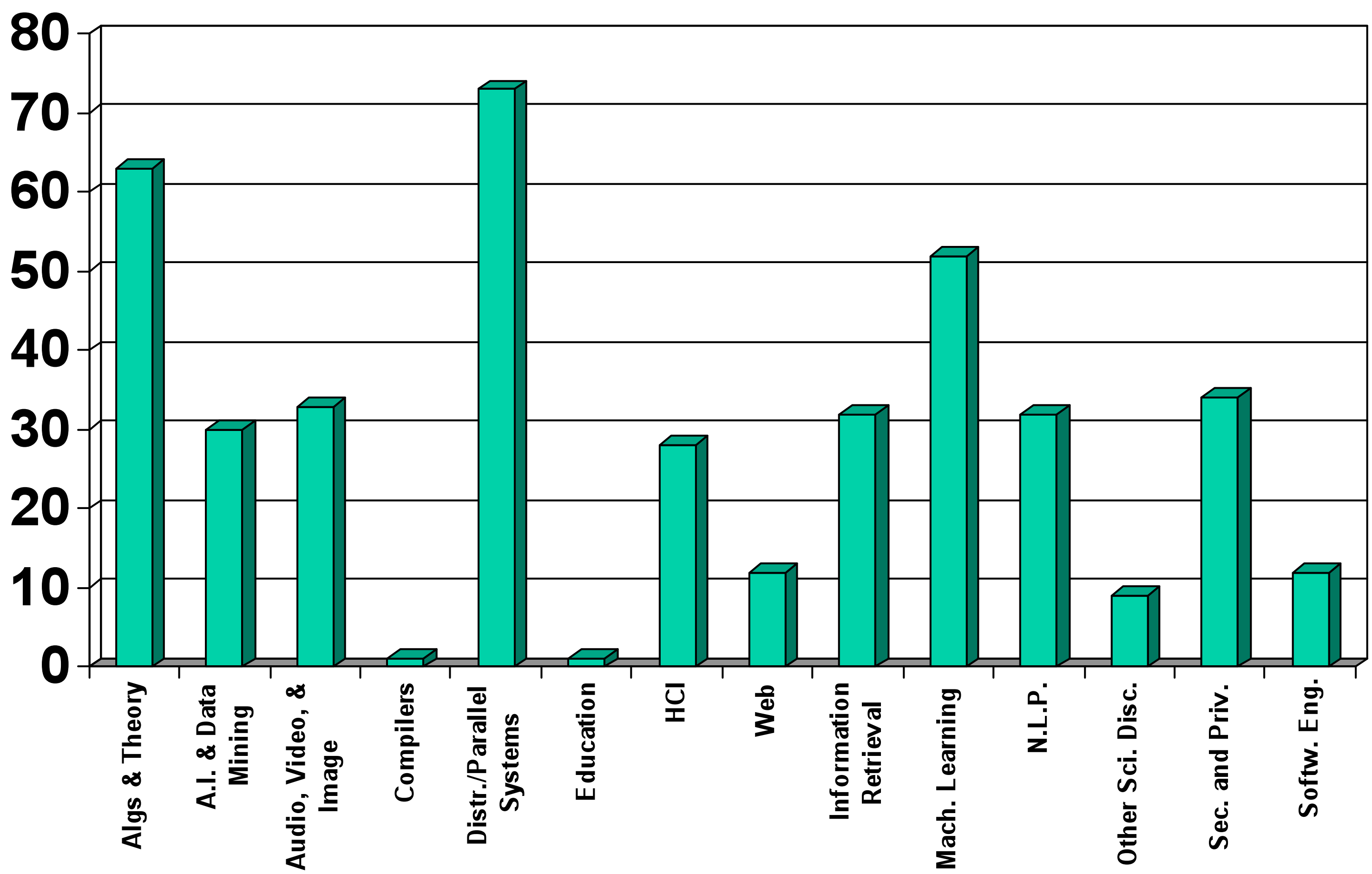
- Focus on **talent**
- **Distributed** across the organization:
  - Impacting Google necessitates broad, diverse involvement in science <u>and</u> engineering
  - Research is done both in our research team and in our engineering organization, organized opportunistically
- Teams benefit greatly:
  - From mutual talent
  - From Google’s comparative advantages to our scale and broad use
  - From service-based architecture (“ease” of working *in vivo*)

Publications

- Publishing/presenting is important
- Publications at www.research.google.com

Year	Number of Publications
2004	15
2005	54
2006	122
2007	197
2008	216
2009	250+ (p)

Google Publications as of Early 2008



From our web site:http://research.google.com/pubs/papers.html

Key Systems Papers, for Example

Google Cluster Infrastructure

Luiz Barroso, Jeffry Dean, Urs Hoelzle: Web Search for a Planet: The Google Cluster Infrastructure, IEEE Micro, Volume 23, Issue 2, March 2003.

GFS (Google File System)

Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung: The Google File System, 19th ACM Symposium on Operating System Principles, Lake George, NY, October 2003.

MapReduce Programming Model for generating & processing large data sets

Jeffrey Dean, Sanjay Ghemawat: MapReduce: Simplified Data Processing on Large Clusters, OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004.

BigTable: A Distributed Storage System for Structured Data

Chang, et al. OSDI 06.

- Many more papers in other areas, of course

Selected Publications

- ***Audiovisual Celebrity Recognition in Unconstrained Web Videos*, Sargin, Aradhye, Moreno, Zhao**
- *Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods,* Keshet, Bengio
- *Bid Optimization for Broad Match Ad Auctions, Even dar, Mansour,* Mirrokni, Muthukrishnan, Nadav
- *Dependency Parsing*, Kubler, McDonald, Nivre
- *Detecting The Origin Of Text Segments Efficiently,* Abdel-Hamid, Behzadi, Christoph, Henzinger
- *Domain Adaptation with Multiple Sources*, Mansour, Mohri
- *LSH Banding for Large-Scale Retrieval with Memory and Recall Constraints,* Covell, Baluja
- ***OpenFst: An Open-Source, Weighted Finite-State Transducer Library and its Applications to Speech and Language*, Riley, Allauzen, Jansche**

Selected Publications

- *Outclassing Wikipedia in Open-Domain Information Extraction: Weakly-Supervised Acquisition of Attributes over Conceptual Hierarchies*, Pasca
- *Solving Maximum Flow Problems on Real World Bipartite Graphs,* Negruseri, Pasoi, Stanley, Stein, Strat
- *The Unreasonable Effectiveness of Data,* Halevy, Norvig, Pereira
- *Using the Doubling Dimension to Analyze the Generalization of Learning Algorithms,* Bshouty, Li, Long
- ***Google's Deep-Web Crawl,* Madhavan, Ko, Kot, Ganapathy, Rasmussen, Halevy**
- *Webtables: Exploring the power of tables on the web,* Cafarella, Halevy, Wang, Zhang

Selected Publications

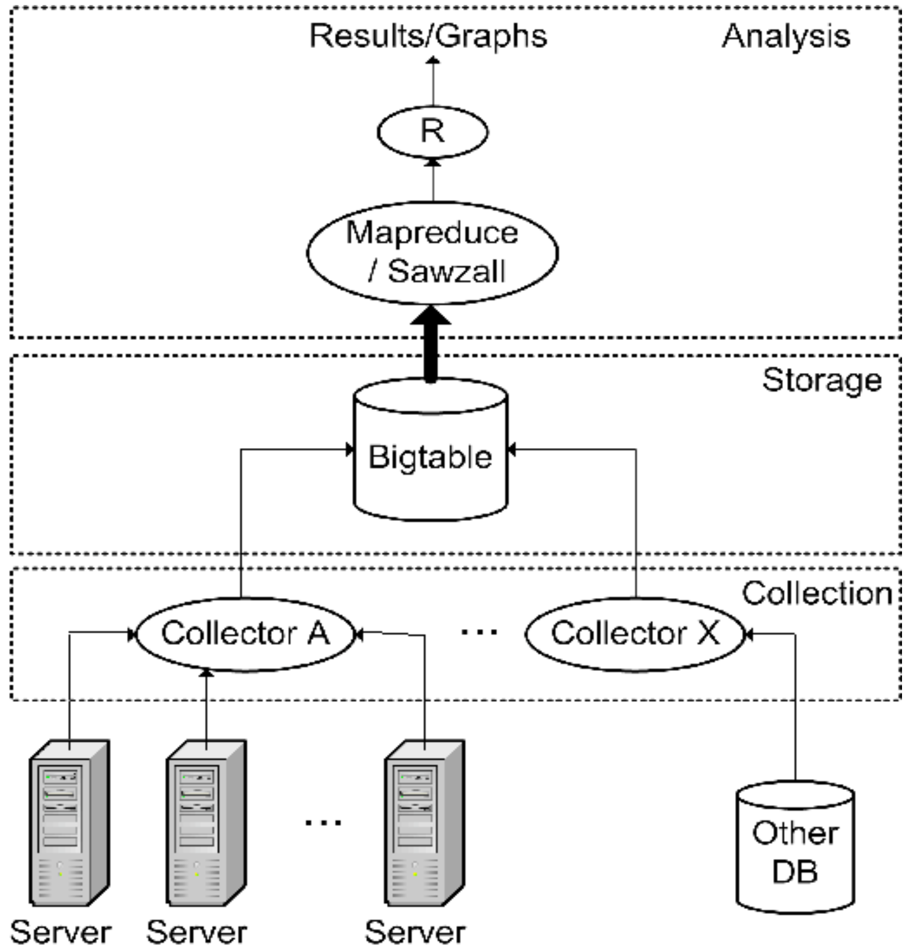
- *Cost-efficient Dragonfly Topology for Large-scale Systems*, Kim, Dally, Scott, Abts,
- ***Detecting influenza epidemics using search engine query data*, Ginsberg, Mohebbi,Patel,Brammer, Smolinski, Brilliant**
- ***Discriminating the relevance of web search results with measures of pupil size,* Oliveira, Aula, Russell**
- *A discriminative kernel-based model to rank images from text queries*, Grangier, Bengio
- ***Boosting with Structural Sparsity,* Duchi, Singer**
- *Affiliation Networks,* Lattanzi, Sivakumar
- *On Sampling-based Approximate Spectral Decomposition,* Kumar, Mohri, Talwalkar

Disk drive failures are a significant problem

- Datasheets do not tell whole story; MTTF is not enough
- Significant and valuable previous work, but insufficient
- Knowledge would help
- Conventional wisdom --- Is it true?
  - Typical disk drive failure rate: < 1% per year
  - Temperature: increases failures
  - Utilization: increases failures

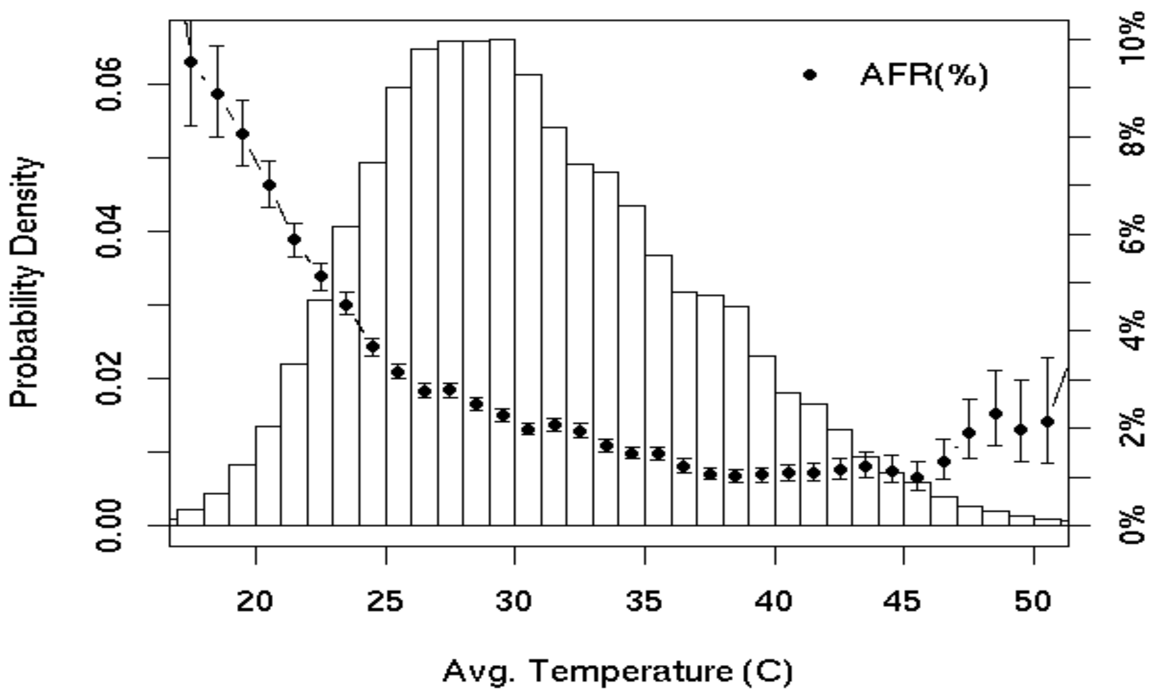
System Health Infrastructure

- Data collected on every machine, periodically
- Stored indefinitely
- Analysis done offline

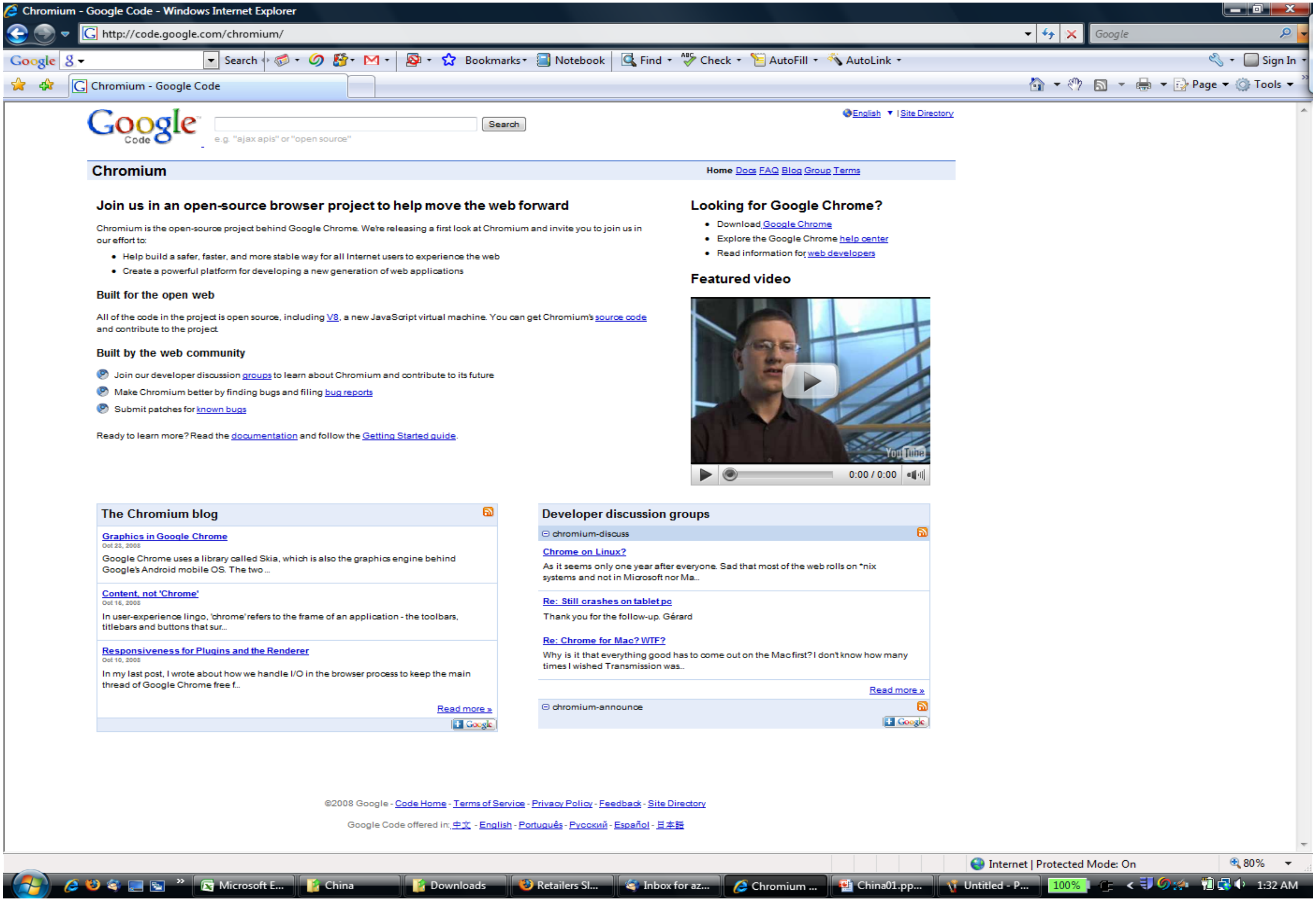


Temperature

Failure rates per average temperature



- No strong indication that high temp high failures.


Chromium - Google Code - Windows Internet Explorer
http://code.google.com/chromium/
Chromium - Google Code
English | Site Directory
Search
e.g. "ajax apis" or "open source"
Chromium
Home Docs FAQ Blog Group Terms
Join us in an open-source browser project to help move the web forward
Chromium is the open-source project behind Google Chrome. We're releasing a first look at Chromium and invite you to join us in our effort to:
Help build a safer, faster, and more stable way for all Internet users to experience the web
Create a powerful platform for developing a new generation of web applications
Built for the open web
All of the code in the project is open source, including V8, a new JavaScript virtual machine. You can get Chromium's source code and contribute to the project.
Built by the web community
Join our developer discussion groups to learn about Chromium and contribute to its future
Make Chromium better by finding bugs and filing bug reports
Submit patches for known bugs
Ready to learn more? Read the documentation and follow the Getting Started guide.
Looking for Google Chrome?
Download Google Chrome
Explore the Google Chrome help center
Read information for web developers
Featured video
0:00 / 0:00
The Chromium blog
Graphics in Google Chrome
Google Chrome uses a library called Skia, which is also the graphics engine behind Google's Android mobile OS. The two...
Content, not 'Chrome'
In user-experience lingo, 'chrome' refers to the frame of an application - the toolbars, titlebars and buttons that sur...
Responsiveness for Plugins and the Renderer
In my last post, I wrote about how we handle I/O in the browser process to keep the main thread of Google Chrome free f...
Read more »
Developer discussion groups
chromium-discuss
Chrome on Linux?
As it seems only one year after everyone. Sad that most of the web rolls on *nix systems and not in Microsoft nor Ma...
Re: Still crashes on tablet pc
Thank you for the follow-up. Gérard
Re: Chrome for Mac? WTF?
Why is it that everything good has to come out on the Mac first? I don't know how many times I wished Transmission was...
Read more »
chromium-announce
©2008 Google - Code Home - Terms of Service - Privacy Policy - Feedback - Site Directory
Google Code offered in: 中文 - English - Português - Русский - Español - 日本語
Internet | Protected Mode: On
80%
1:32 AM

ANDROID

open source project

How will you shape it?

Get the source code »

Market Updated!

Android Market is an open service that makes it easy for developers to distribute applications to handsets.

Learn more »

Developers

Android enables developers to easily create great mobile applications.

Learn more »

Open Source

Android Open Source Project provides access to the entire platform source and enables developers to contribute.

Learn more »

Featured Video

Introduction to Android Open Source Project.

Watch the video »

Timeline

Learn more about what's happened and where we're headed.

Android timeline »

In the News

T-Mobile G1 Now Available
T-Mobile

Android Open Source Project
Android

Open Handset Alliance announces the launch of Android Open Source Project
Open Handset Alliance

Site Terms of Service - Privacy Policy - Brand Guidelines

www.android.com

Google™

The Evolutionary Path Forward to New Accomplishments

- Application mix will continue to grow in unpredictable ways:
  - Four areas in flux today: *publishing, education, healthcare and government*
- Systems will evolve: ubiquitous high performance networking, distributed computing, new end-user devices, ...
- Three truly big results brewing:

1.Totally Transparent Processing

2.Ideal Distributed Computing

3.Hybrid, Not Artificial, Intelligence

Totally Transparent Processing

Totally Transparent Processing

$$\forall d \in D,\ \forall l \in L,\ \forall m \in M,\ \forall c \in C$$

D: The set of all end-user access devices

- Personal Computers
- Phone
- Media Players/Readers
- Telematics
- Set-top Boxes
- Appliances
- Health devices
- …

L: The set of all human languages

- Current languages
- Historical languages
- Other forms of human notation
- Possible language specialization
- Formal languages
- …

M: The set of all modalities

- Text
- Image
- Audio
- Video
- Graphics
- Other sensor-based data
- …

C: The set of all corpora

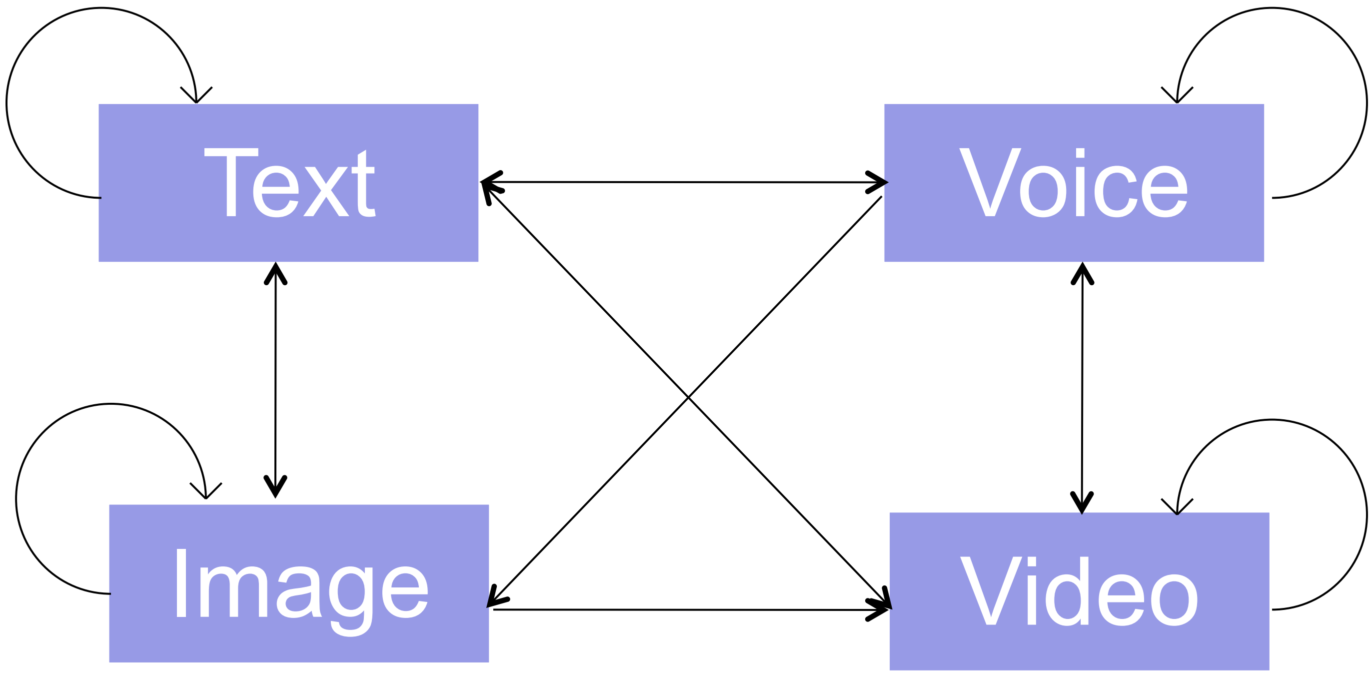
- The normal web
- The deep web
- Periodicals
- Books
- Catalogs
- Blogs
- <u>Universal</u> Geodata
- Scientific datasets
- Health data
- …

Types of Transparent Processing

- Search, of many forms
- Navigation and Suggestion
- Transformational Communication
- Information Fusion

- Some Google Examples:
  - Universal search
  - Voice Search
  - Find Similar, applied to images
  - Google Translate, particularly in mash-ups
  - Combining images and maps
  - Audio transcription
  - Images and 3d models

Fluidity Among the Modalities



Many New Modalities Here or Coming

The New Frontier of Web Search – Better/Faster Queries

Query completion before: *Used a fixed dictionary, e.g., in emacs, bash, T9, etc.*

Query suggestion today: *Model queries with query logs, serve them dynamically*

Technical challenges:

- *response-time, coverage, freshness, corpus dependency (YouTube, image, mobile)*
- *domain dependent:* `rea -> real madrid` *good suggestion in Spain*
- *diversity (danger of popularity), filtering out duplicates, inappropriate results, etc.*

Impact: *Made possible by scale,*

- *the richer the query log corpus, the better*
- *the faster the response time, the better*

Voice Search

Challenges and Rationale for Success

Technically this is very challenging:

- Huge vocabulary
- Variability in accent
- Background noise

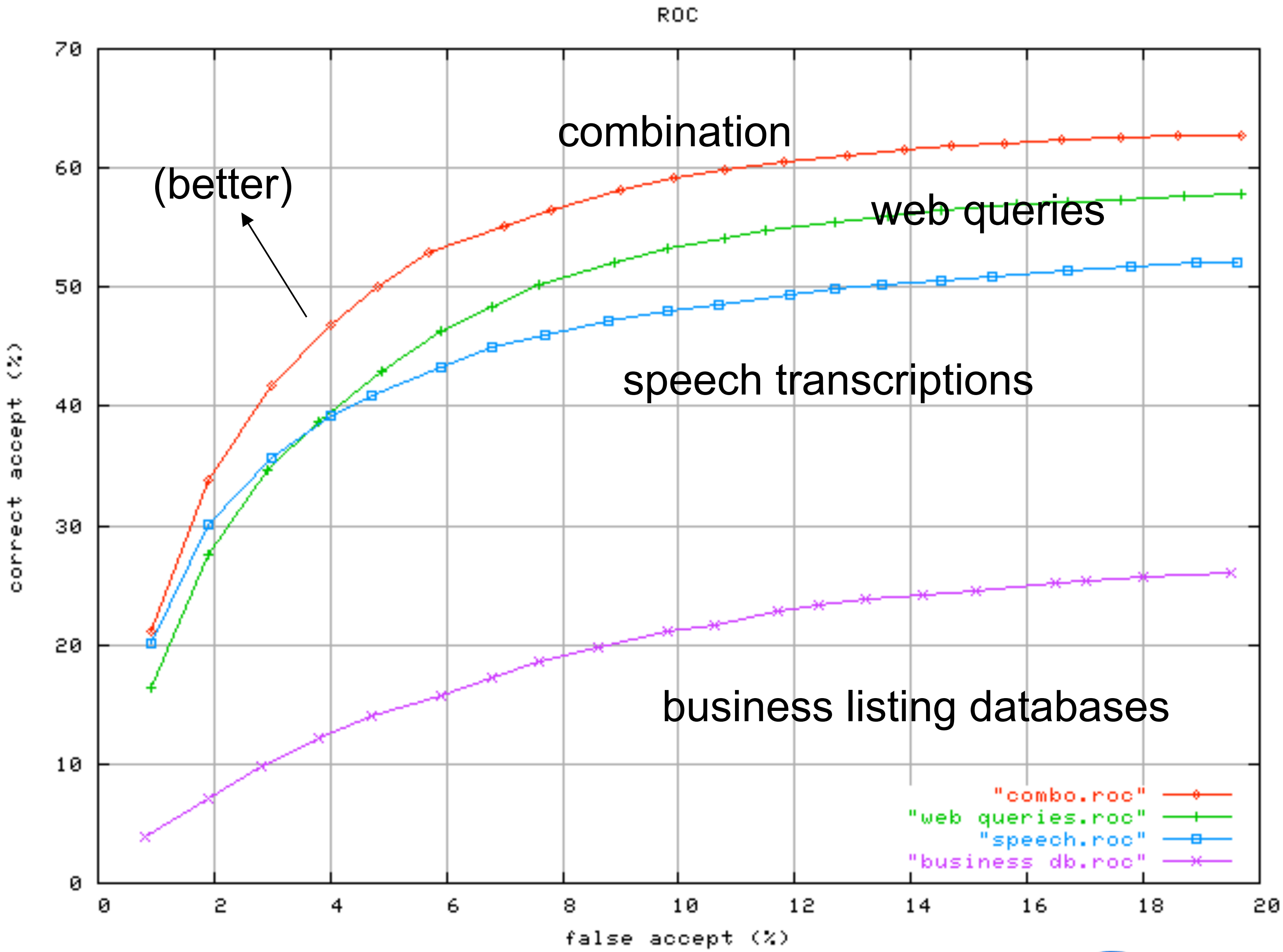
What makes this possible:

- Scalable technology
- Data inputs: Query logs, voice logs
- Compute power

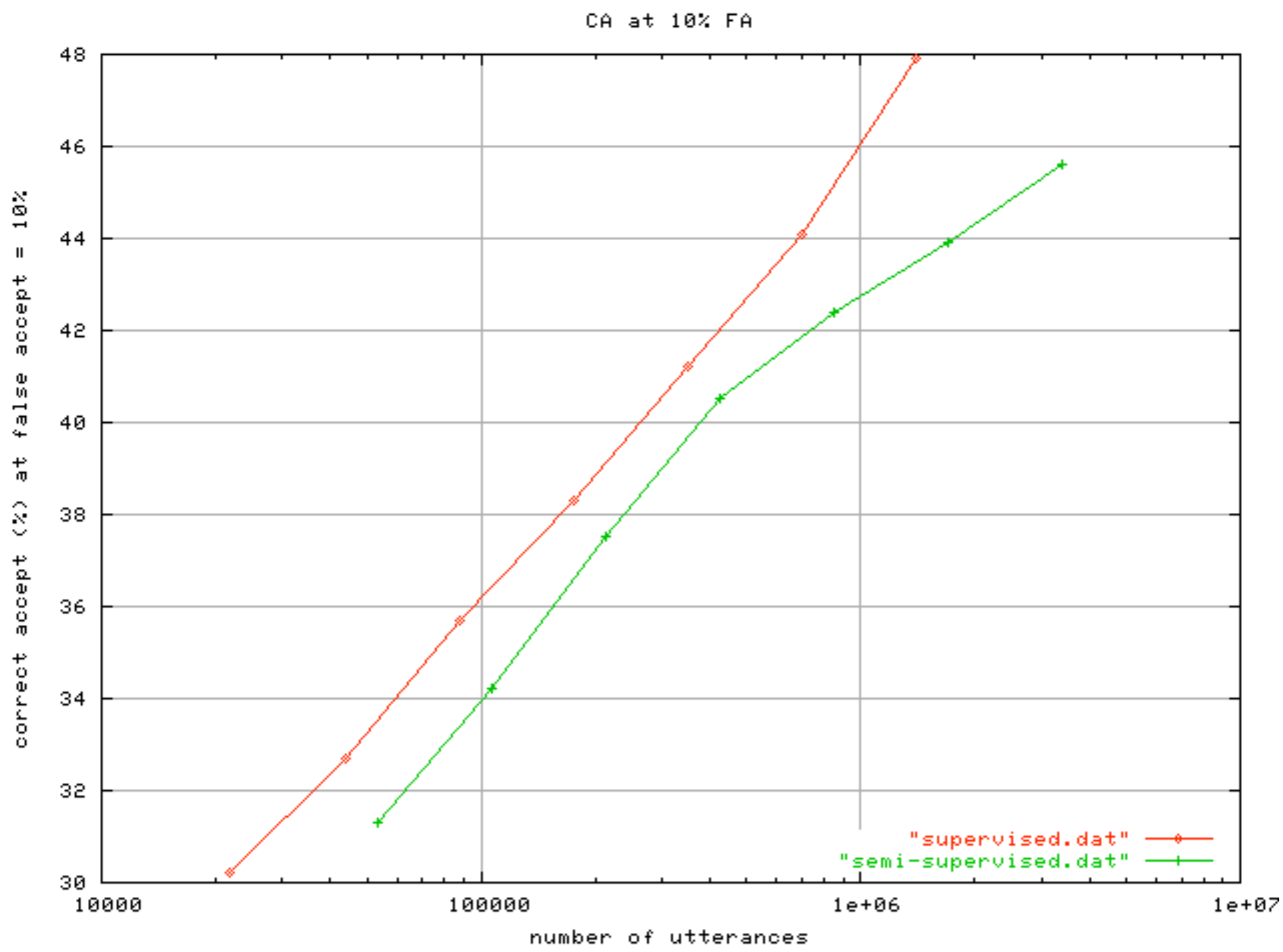
Choosing a data source

Web Queries	Speech Transcriptions	Business Listing Databases
huge N	tiny N	wide coverage
typed, not spoken	perfectly matched to task	little info for popularity
user in a different setting	slow, expensive, manual(!)	little info for how callers ask
Google advantage	chicken and egg	where most efforts like this start

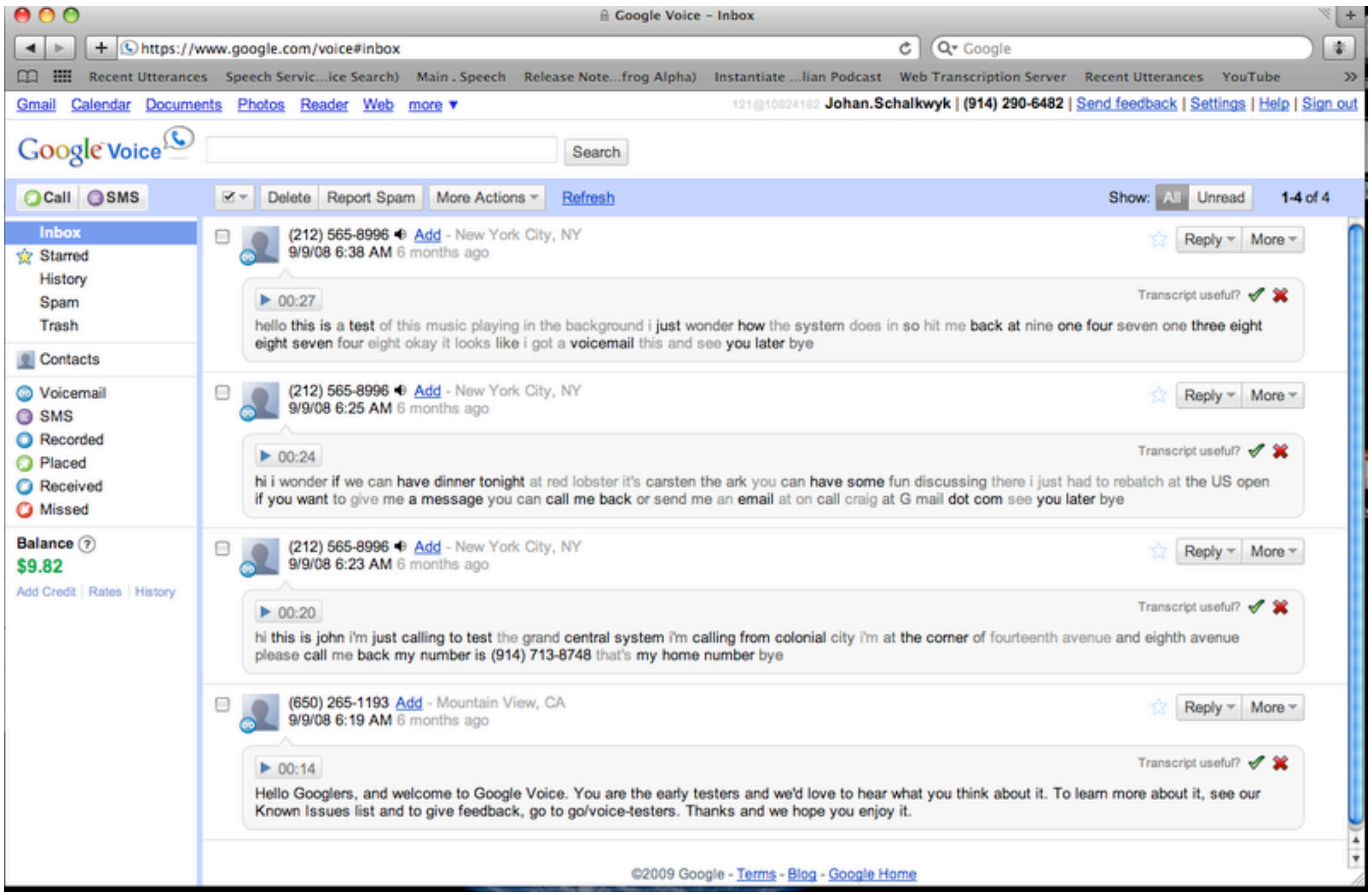
Combination works best: utterance ROC curves: incl. rejection)



The Benefits of Unsupervised Training



Transcriptions in Google Voice



Google Translate

RBMT *– Rules-based machine translation*

SMT *– Statistical (data-driven) machine translation*

2001-2004

RBMT – 3rd Party

2001

- *French*
- *Italian*
- *Spanish*
- *Portuguese*
- *German*

2004

- *Chinese*
- *Japanese*
- *Korean*

2006

SMT - Google

- *Chinese*
- *Arabic*
- *Russian*

RBMT – 3rd Party

- *French*
- *Italian*
- *Spanish*
- *Portuguese*
- *Japanese*
- *Korean*

2007

SMT - Google

- *Arabic*
- *Chinese (S)*
- *Chinese (T)*
- *Dutch*
- *French*
- *German*
- *Greek*
- *Italian*
- *Korean*
- *Japanese*
- *Russian*
- *Spanish*
- *Portuguese*

2008

SMT - Google

- *Arabic*
- *Bulgarian*
- *Catalan*
- *Chinese (S)*
- *Chinese (T)*
- *Croatian*
- *Czech*
- *Danish*
- *Dutch*
- *Filipino*
- *Finnish*
- *French*
- *German*
- *Greek*
- *Hebrew*
- *Hindi*
- *Indonesian*
- *Italian*
- *Japanese*
- *Korean*
- *Latvian*
- *Lithuanian*
- *Norwegian*
- *Polish*
- *Portuguese*
- *Romanian*
- *Russian*
- *Serbian*
- *Slovak*
- *Slovenian*
- *Spanish*
- *Swedish*
- *Ukrainian*
- *Vietnamese*

2009

SMT - Google

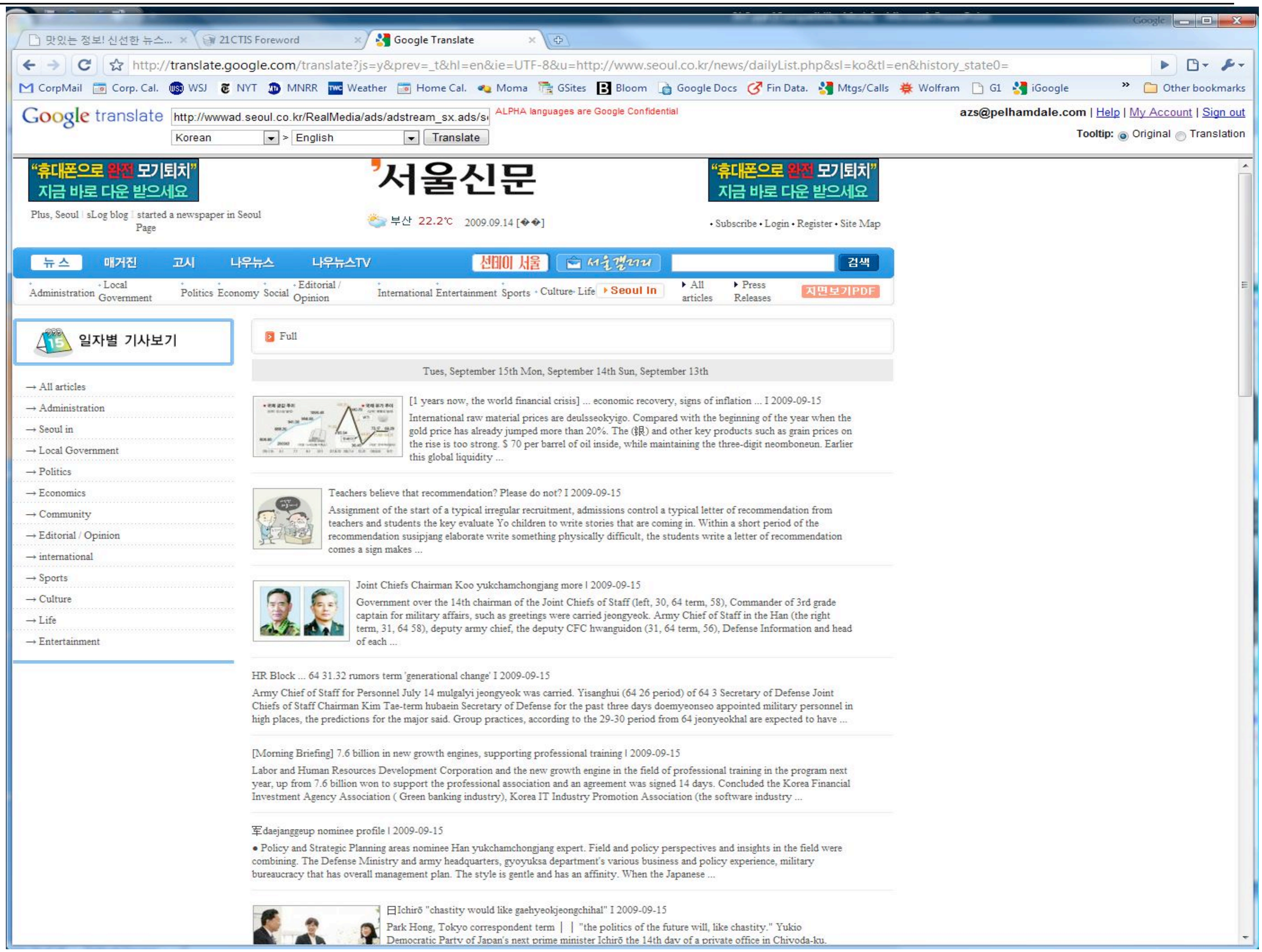
- *Albanian*
- *Arabic*
- *Bulgarian*
- *Catalan*
- *Chinese (S)*
- *Chinese (T)*
- *Croatian*
- *Czech*
- *Danish*
- *Dutch*
- *Estonian*
- *Filipino*
- *Finnish*
- *French*
- *Galician*
- *German*
- *Greek*
- *Hebrew*
- *Hungarian*
- *Hindi*
- *Indonesian*
- *Italian*
- *Japanese*
- *Korean*
- *Latvian*
- *Maltese*
- *Norwegian*
- *Polish*
- *Portuguese*
- *Romanian*
- *Russian*
- *Serbian*
- *Slovak*
- *Slovenian*
- *Spanish*
- *Swedish*
- *Thai*
- *Turkish*
- *Ukrainian*
- *Vietnamese*

NEW

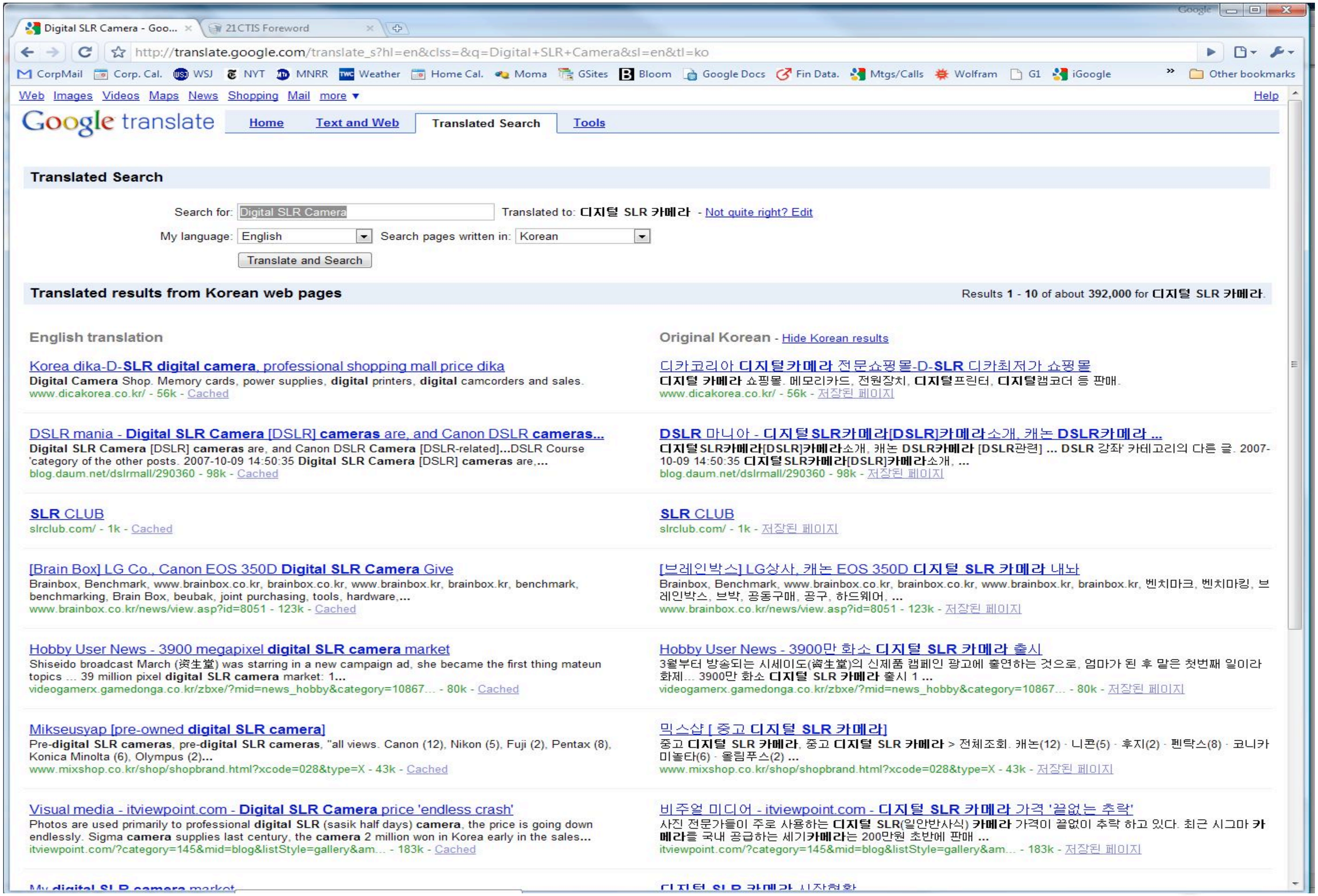
- *Afrikaans*
- *Belarusian*
- *Icelandic*
- *Irish*
- *Macedonia*
- *Malay*
- *Swahili*
- *Welsh*
- *Yiddish*

2001 - 2004 | *2006* | *2007* | *2008* | *2009*

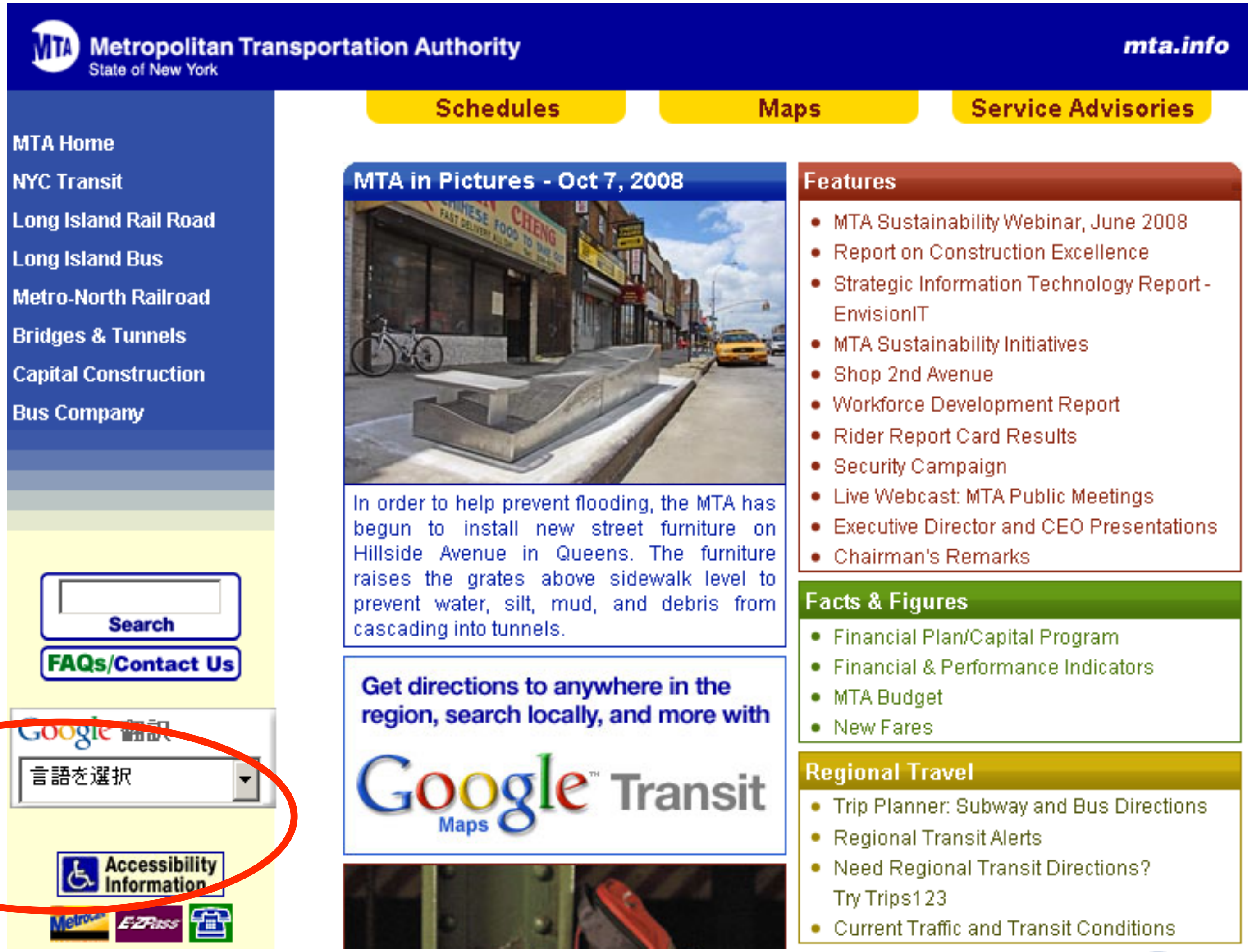
Web Translation



Cross-language search



Translate My Page Gadget



Google Desktop: America... | Translated version of htt...

http://www.translate.google.com/translate?prev=hp&hl=en&js=n&u=http://www.www2009.org%0D%0A%0D%0A&sl=en&tl=es

Customize Links | CorpMail | Corp. Cal. | Moma | WSJ | NYT | MNRR | Pel W. | Home Cal. | google sites | Bloomberg | Google Docs | iGoogle | Fin Data. | Other bookmarks

Google

This page was automatically translated from English.
View original web page or mouse over text to view original language.

Back to Translate

2009 MADRID!

WELCOME | ATTENDING | AUTHORS | COMMITTEES | PRESS ROOM | SPONSORS | STORE | REGISTRATION

CONTACTO

DAILY PROGRAM

PROCEEDINGS

Documentos y Presentaciones

Twitter etiqueta **# www2009**

Flickr etiqueta **WWW2009MADRID**

Grupo de Facebook

DISFRUTAR DE ESPAÑA!

MAPA: LUGAR Y MONUMENTOS

CAPTURAS DE LOS HECHOS DE MADRID VIDA!

TOURS 1 DÍA PARA: TOLEDO, EL ESCORIAL O SEGOVIA

EXPERIENCIA BULLET viajes en tren a: SEVILLA O BARCELONA (2 ½ horas)

Divertirse en SEVILLA Feria de Abril (4/28/2009 - 5/3/2009)

ENCONTRAR SU PREFERENCIA EN ONDA SURF MUNDAKA O Tarifa

Actualizaciones de Noticias

Noticias Anteriores actualizaciones ...

"Abril 22 - CEREMONIA DE INAUGURACIÓN"

La ceremonia se celebrará en el Auditorio A, justo después de la Web Grupo 20 º aniversario.

Con el fin de asistir a la ceremonia de inauguración, los participantes deben usar la Conferencia Insignia, de haber completado el proceso de inscripción en el registro de contador situado en el Palacio Municipal de Congresos del 20 de abril en adelante.

Por razones de seguridad siguientes restricciones se aplicarán el 22 de abril:

- Contra el registro el 22 de abril se abrirá a 07.45 hrs. A partir de las 10.00 a las 11.30 será cerrado
- Lugar de acceso, le será prohibido de 10.15 a 10.30 hrs. (incluso a los delegados el uso de una tarjeta de identificación WWW2009)

- **2009/04/07** relacionado de datos disponibles para WWW2009 documentos | Navegar ponencias y pósters en EPrints
- **2009/04/07** Mejor Papel / Cartel nominaciones anunciadas

WWW 2009 MADRID!SPAIN

WWW2009 Programa de un vistazo

20 de abril Lunes	21 de abril Martes	22 de abril Miércoles	23 de abril Jueves	24 de abril Viernes
Tutoriales		Ponencias		
		Documentos de referencia de pista		
Talleres		BOF	W3C pista	
		Vía web en Iberoamérica		
W4A		Paneles		
		Carteles | Desarrolladores de pista		

Pasado a ser patrocinador o expositor

PATROCINADORES PLATINO

Telefónica | Fundación

PATROCINADORES DE ORO

¡MADRID! | LA CATEDR

PATROCINADORES PLATA

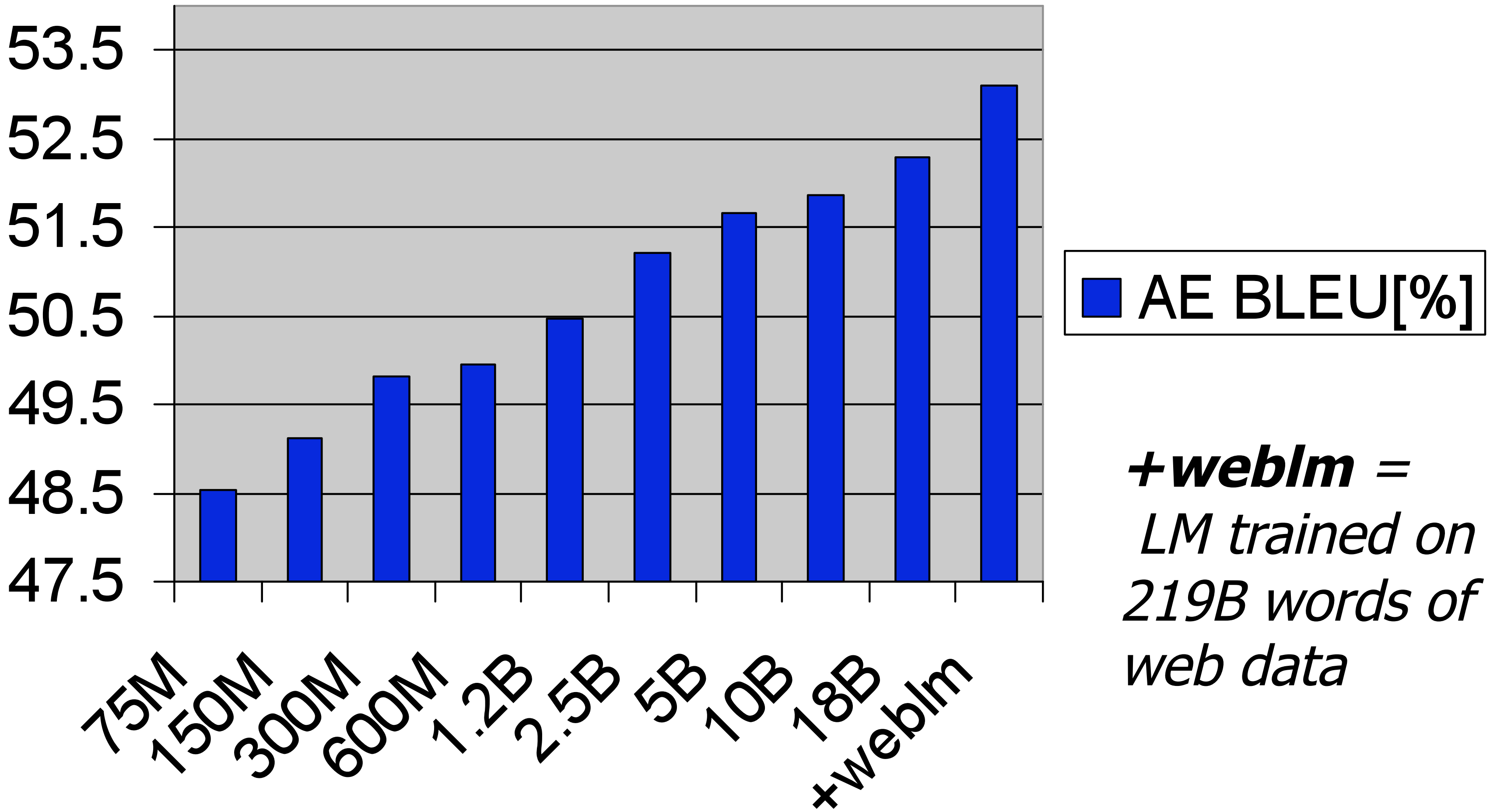
ontotext Semantic Technology Lab | M

PATROCINADORES DE BRONCE

DAEDALUS

YouTube Caption Translation

Impact of data - More data is better …



+weblm *=*
LM trained on 219B words of web data

News data

Challenges in Image processing

Visual Semantics

- Recognition (people, landmarks, objects)
- Machine Learning

Correspondence

- Matching images and videos
- Image mosaicing

Geometry

- Ego-motion estimation
- Multi-view stereo

Image Processing

- Maps from aerial imagery
- OCR in all the world's languages

Google™

Image Analysis in Image Search

- Image Search helps users find the image they want quickly.
- Understanding the actual content of an image is critical.
- We've been focusing more and more on analyzing images
- This has been rolling out over the last year.
  - Both as user visible filters
  - Behind the scenes in our back-ends.
- Genre filters like clip art / line drawings / color are great examples

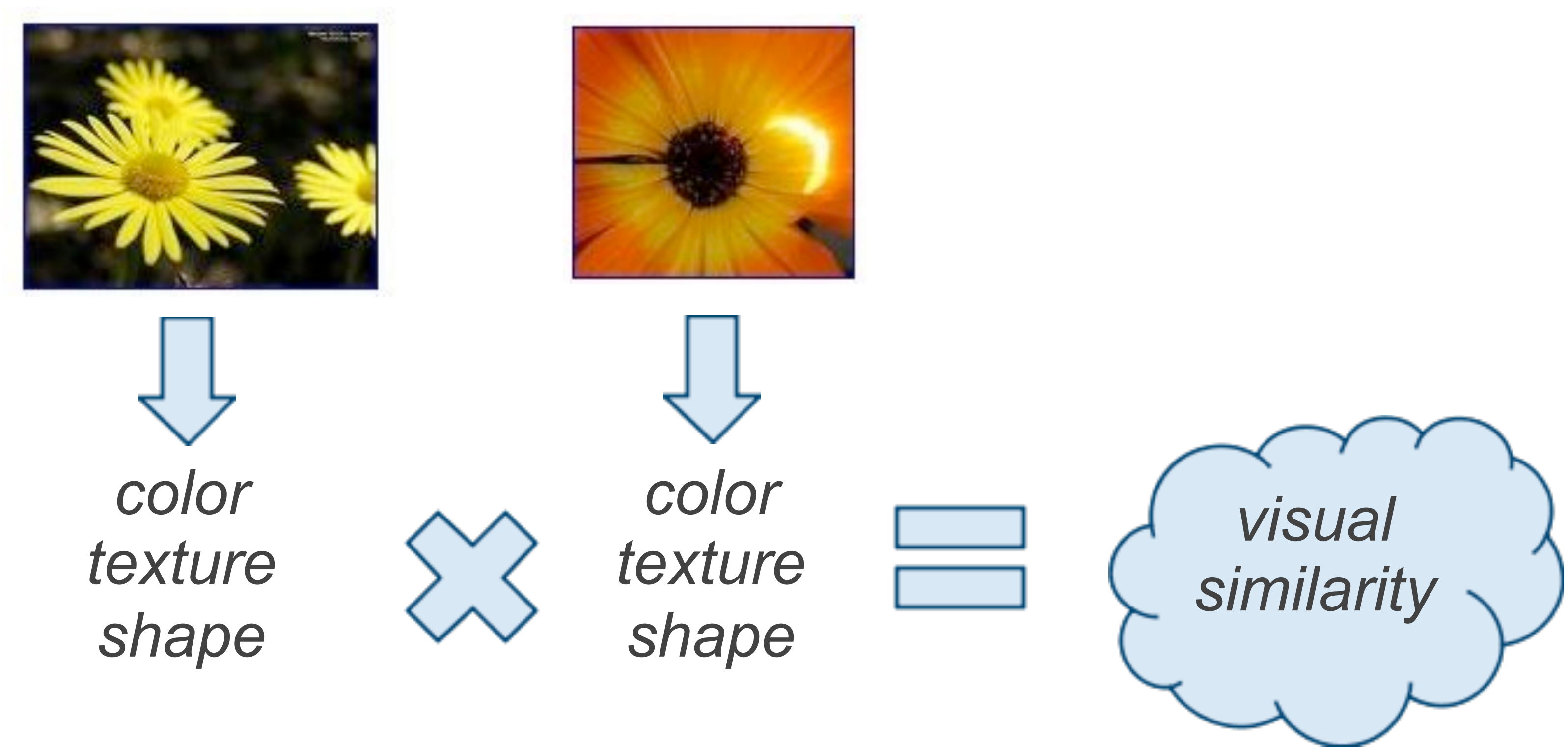
  - [flowers], line drawings, clip art, photo, face
  - [porsche] , red, green, yellow, orange, ...

Similar Images in Image Search

- Google has just launched a "Similar Images" feature.
- Accessed by clicking on the similar images link under an image.
- It can also be accessed via preview thumbnails in the result frame.
- We think this will create a major shift in how to search for images.
- Searching for images can now become a navigational experience, where the text (or voice) query acts as a starting point

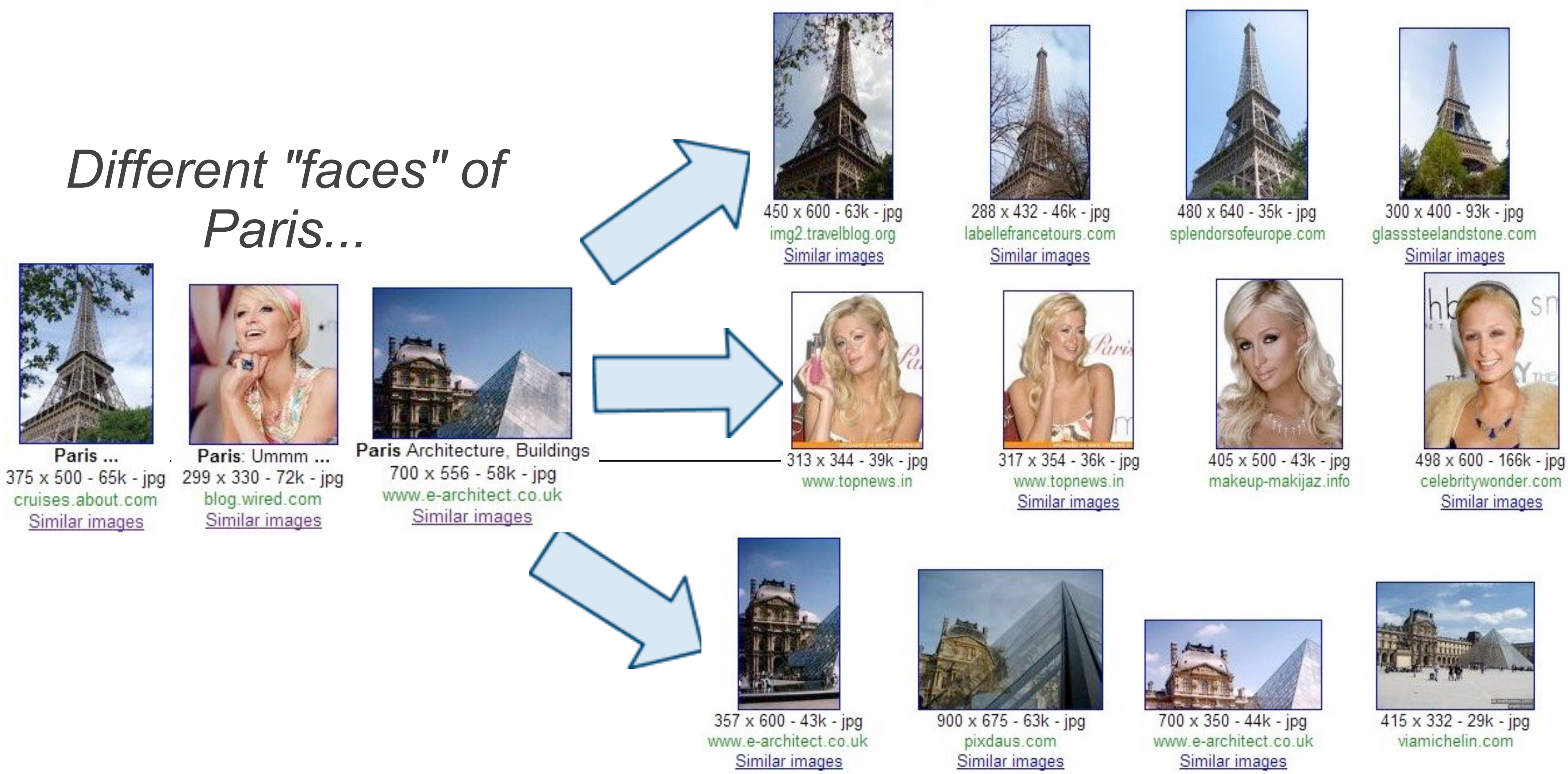
Similar Images in Image Search

- A variety of features are used to determine visual similarity.

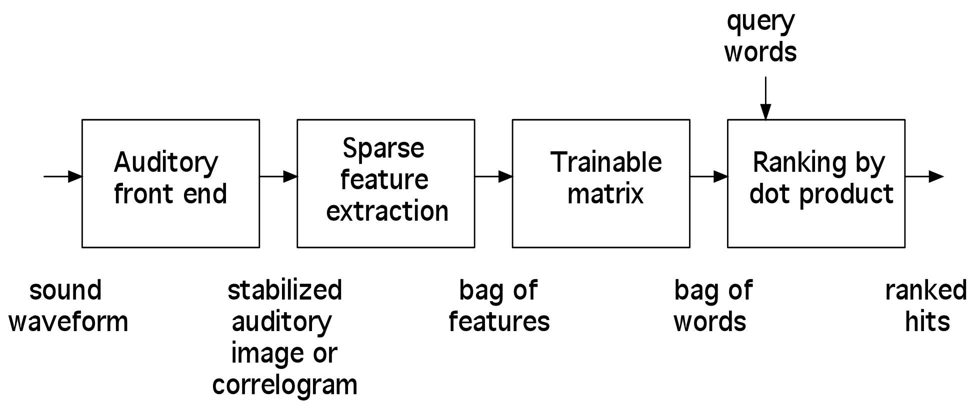


Similar Images in Image Search

***Refine by the content of a specific image*.**



Content-based retrieval of sound

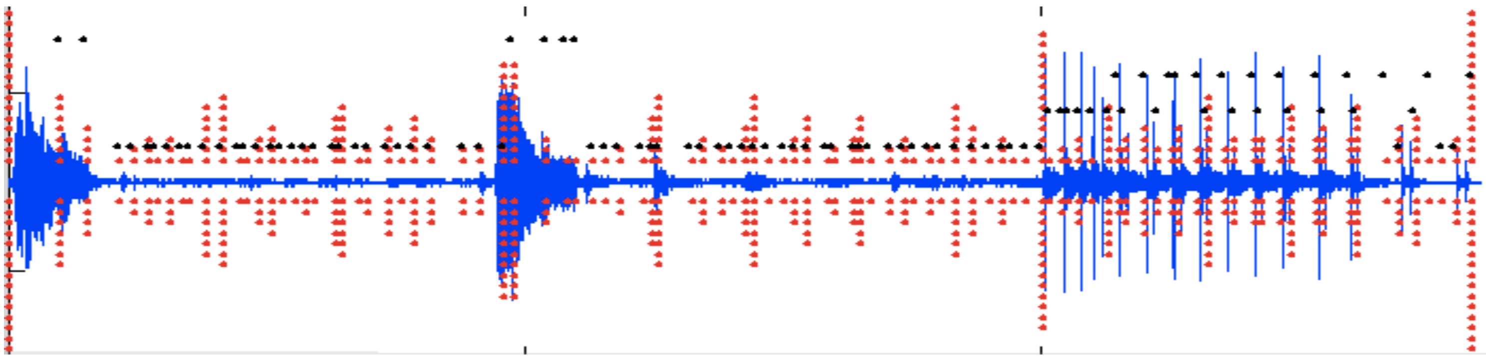


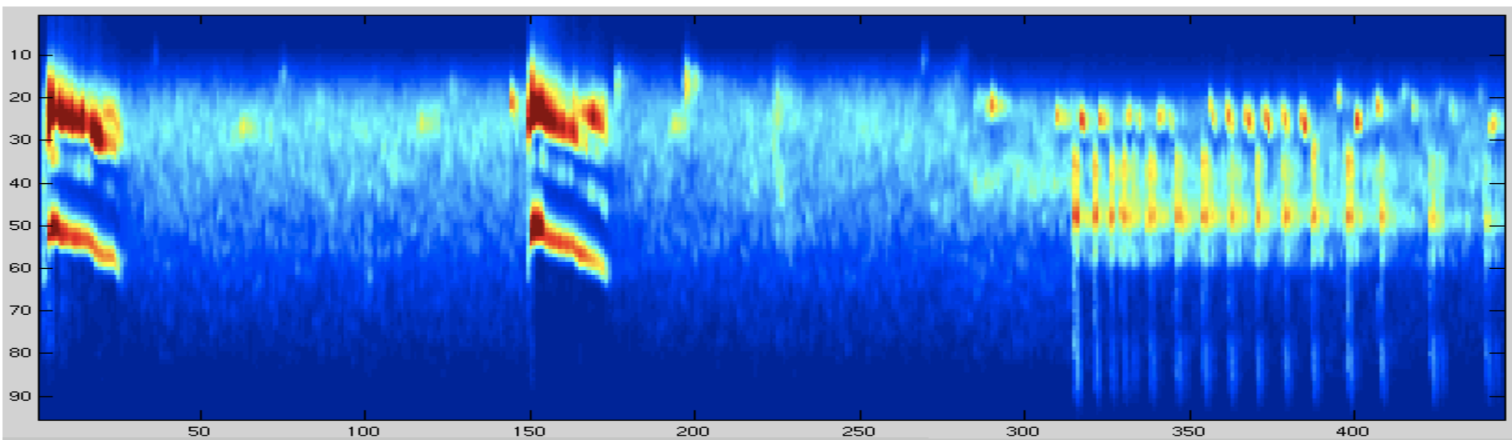
Auditory front end based on stable models from long ago, with new feature extraction ideas.

PAMIR multi-label retrieval (MLR) for the trainable back-end retrieval.

What about sound segmentation or separation?

“Sapsucker” (woodpecker) representations



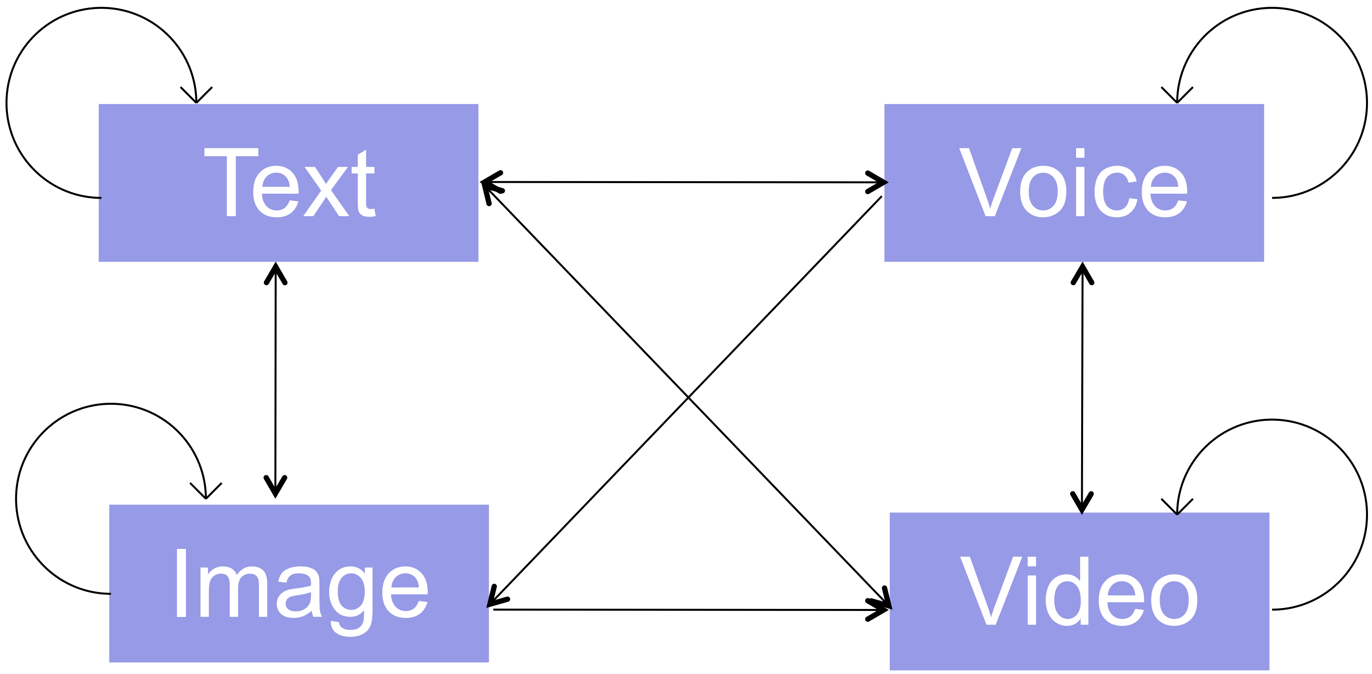


Newer Modalities

- 3D Graphics
- Maps and Geography
- Timelines
- Music
- Etc.

Maps/Earth as a Modality

Totally Transparent Processing In-Process…



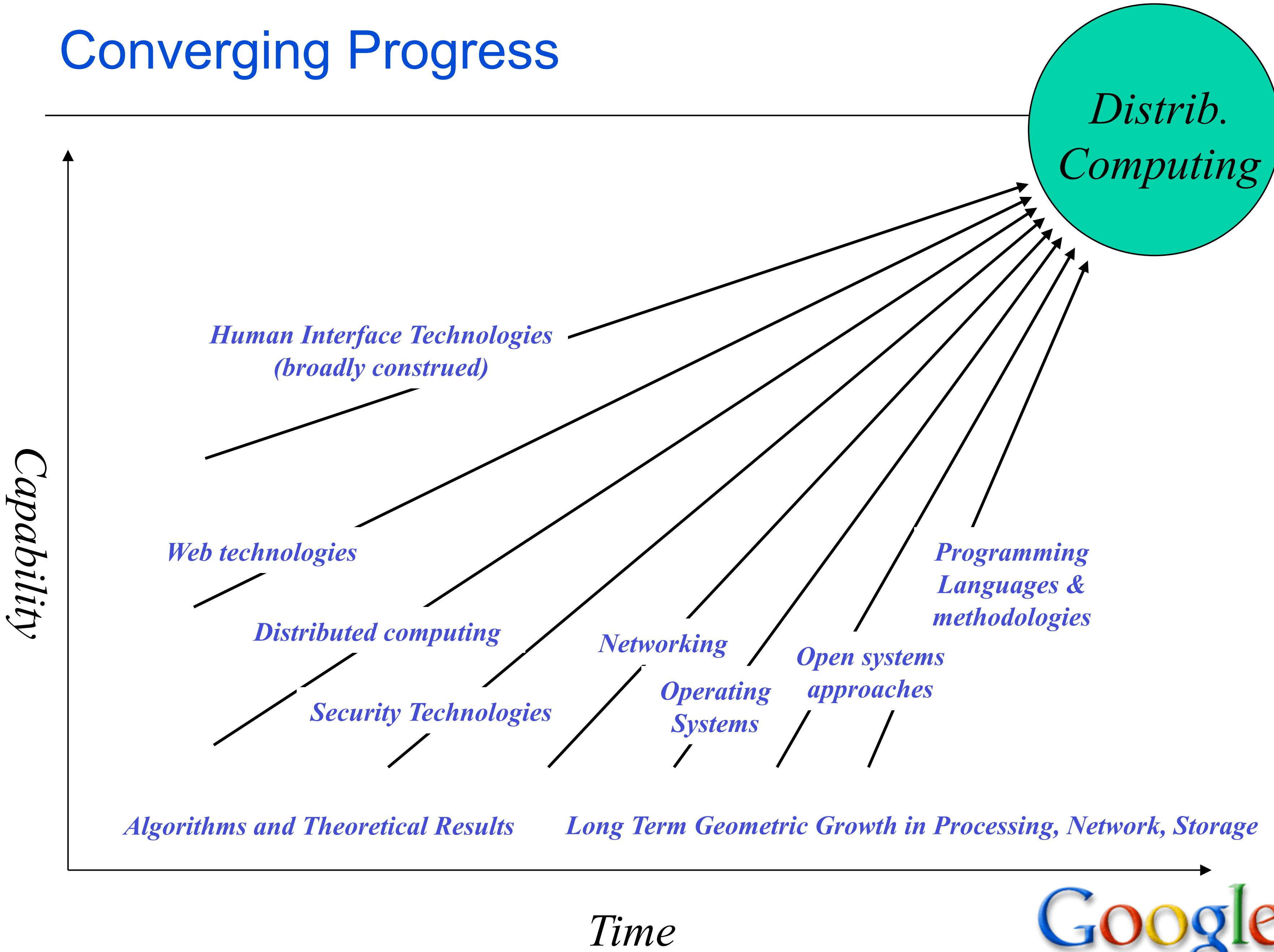
Many New Modalities Here or Coming

Ideal Distributed Computing

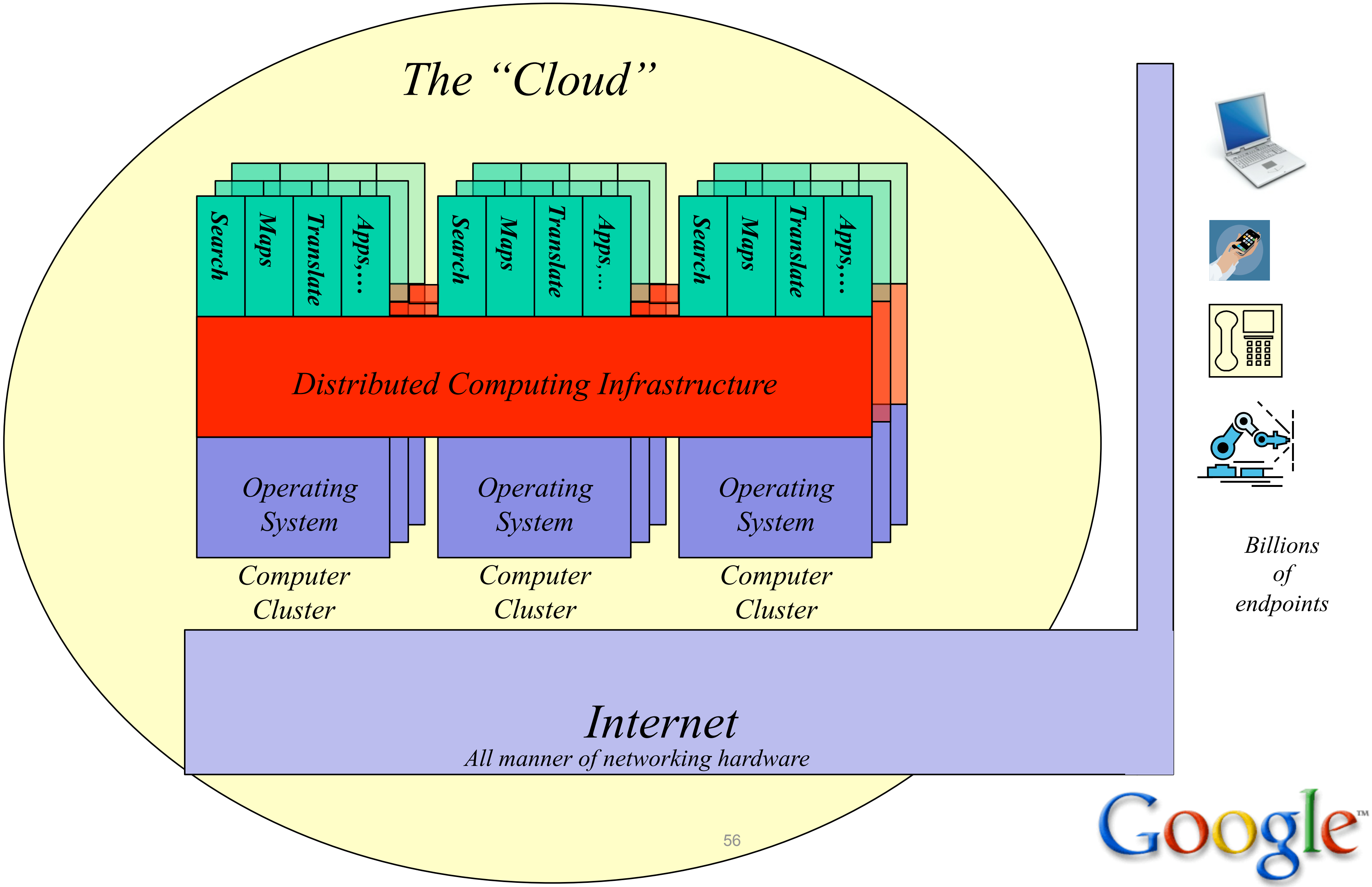
Distributed Computing is 30 years old. But, not very deeply understood until recently

- The application mix
- The true nature of global, open systems:
  - Implications on systems, applications, mix and match.
- The implications of *operations* at true scale
  - E.g., work on programming & runtimes predominated system mgmt.
- The complexity of the architecture that would result
  - We tend to assume, *if we can conceive it, it's okay.*
- The collection of further abstractions that would build on fundamentals then known
- In summary there was a limitation of understanding of (truly) large-scale, open integrated distributed systems

Converging Progress



Cloud Computing Architecture



Optimization Applications

Vast opportunities for applying optimization in large systems:

- There are two traditional fields:
  1. Long Term Planning:
     - Network planning, capacity planning.
  2. Short Term Operational Decisions:
     - Scheduling of tasks, dynamic allocation.
- We want to introduce new types of applications:
  1. On Line Optimization.
  2. Data Checking:
     - Verify consistency and validate data/config file.
  3. Dynamic repair:
     - Find the closest feasible solution after an incident (computer broke down).
     - Continual optimization (i.e. replan after a driver missed a turn).

Fusion Tables: a collaborative database in the cloud

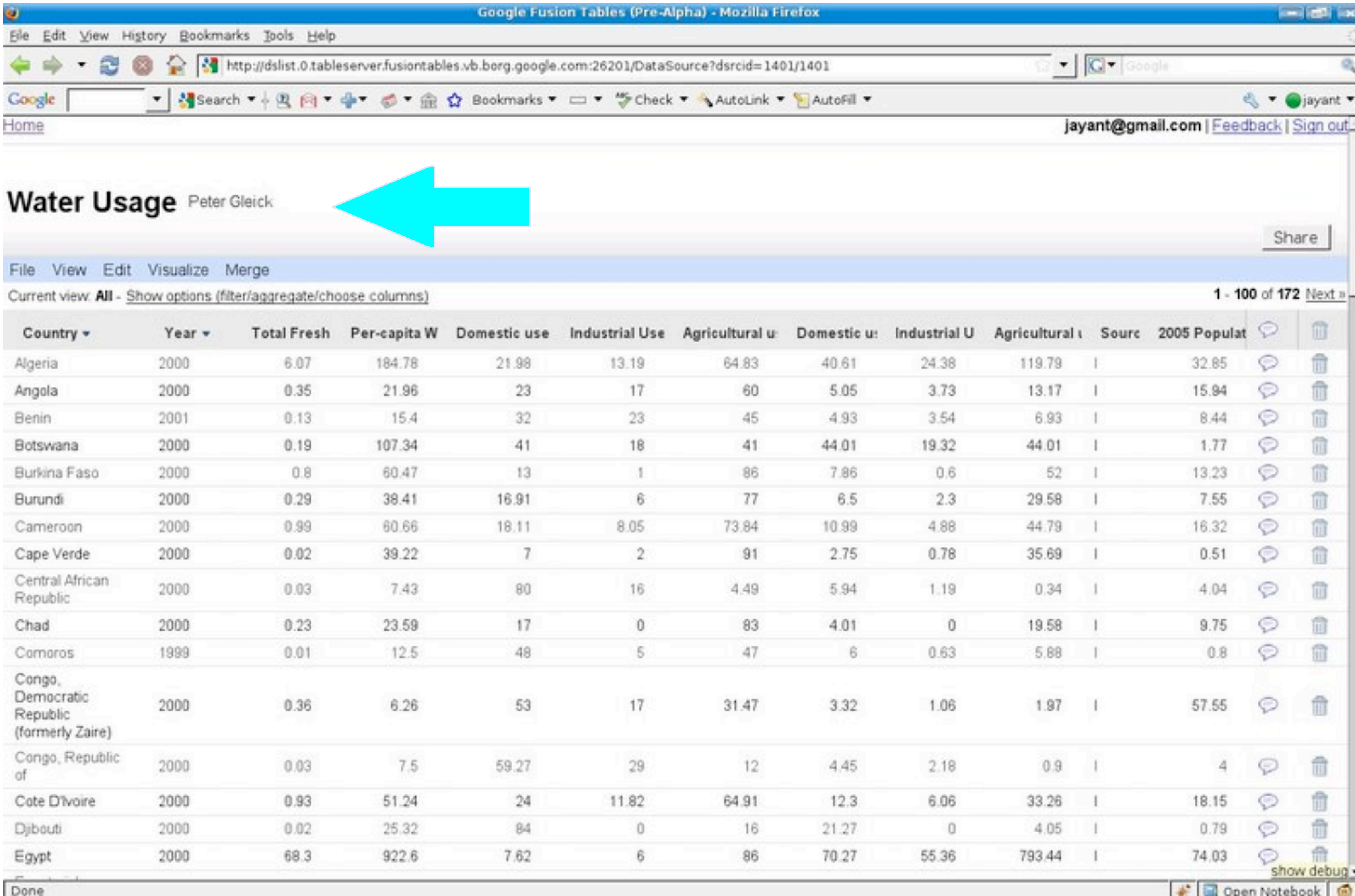
An online database that will make it easier for users to

share data with others	*uploading tables and inviting collaborators*
explore their data	*using filters & aggregates and visualizing the data on maps, timelines, chars, etc.*
combine datasets	*creating merged tables from multiple base tables*
discuss their data	*participating in threaded discussions*
publish as web properties	*making public datasets and embedding visualization on external pages*

- leverage the reliability and performance of the Google infrastructure for data management applications

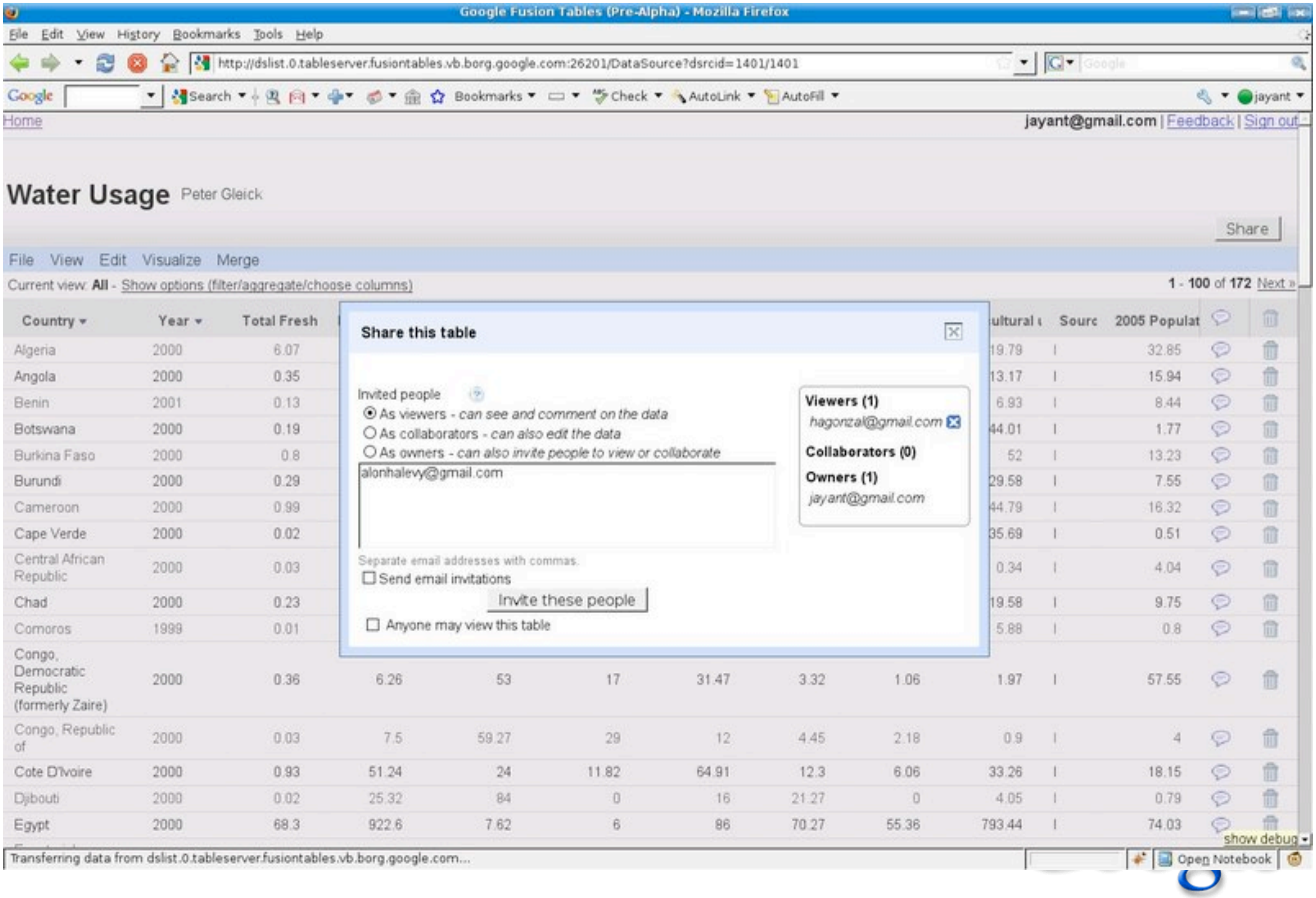
Google™

Fusion Tables Example (1)

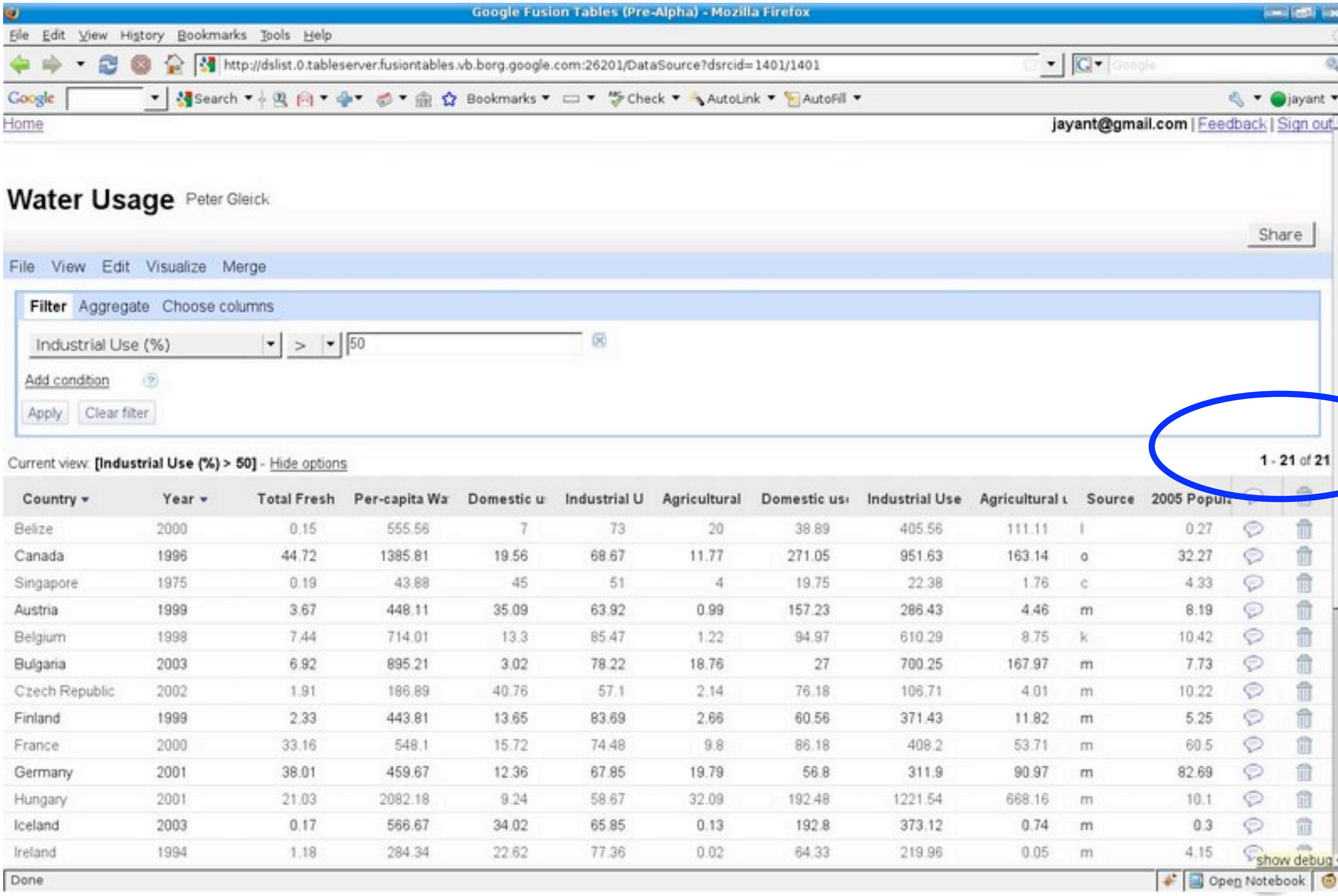


Country	Year	Total Fresh	Per-capita W	Domestic use	Industrial Use	Agricultural u	Domestic u	Industrial U	Agricultural u	Sourc	2005 Populat
Algeria	2000	6.07	184.78	21.98	13.19	64.83	40.61	24.38	119.79	I	32.85
Angola	2000	0.35	21.96	23	17	60	5.05	3.73	13.17	I	15.94
Benin	2001	0.13	15.4	32	23	45	4.93	3.54	6.93	I	8.44
Botswana	2000	0.19	107.34	41	18	41	44.01	19.32	44.01	I	1.77
Burkina Faso	2000	0.8	60.47	13	1	86	7.86	0.6	52	I	13.23
Burundi	2000	0.29	38.41	16.91	6	77	6.5	2.3	29.58	I	7.55
Cameroon	2000	0.99	60.66	18.11	8.05	73.84	10.99	4.88	44.79	I	16.32
Cape Verde	2000	0.02	39.22	7	2	91	2.75	0.78	35.69	I	0.51
Central African Republic	2000	0.03	7.43	80	16	4.49	5.94	1.19	0.34	I	4.04
Chad	2000	0.23	23.59	17	0	83	4.01	0	19.58	I	9.75
Comoros	1999	0.01	12.5	48	5	47	6	0.63	5.88	I	0.8
Congo, Democratic Republic (formerly Zaire)	2000	0.36	6.26	53	17	31.47	3.32	1.06	1.97	I	57.55
Congo, Republic of	2000	0.03	7.5	59.27	29	12	4.45	2.18	0.9	I	4
Cote D'Ivoire	2000	0.93	51.24	24	11.82	64.91	12.3	6.06	33.26	I	18.15
Djibouti	2000	0.02	25.32	84	0	16	21.27	0	4.05	I	0.79
Egypt	2000	68.3	922.6	7.62	6	86	70.27	55.36	793.44	I	74.03

Fusion Tables Example (2)

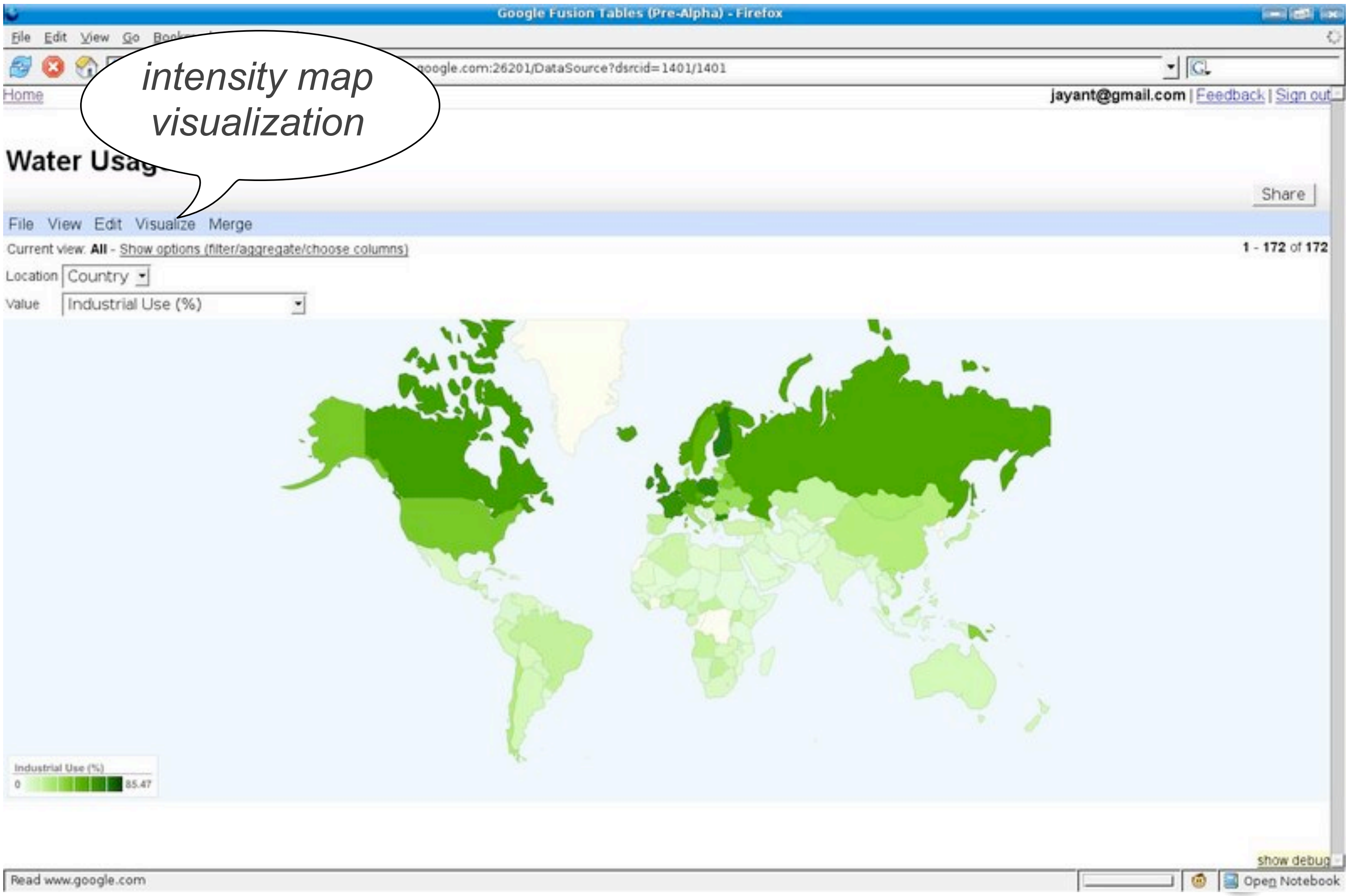


Fusion Tables Example (3)

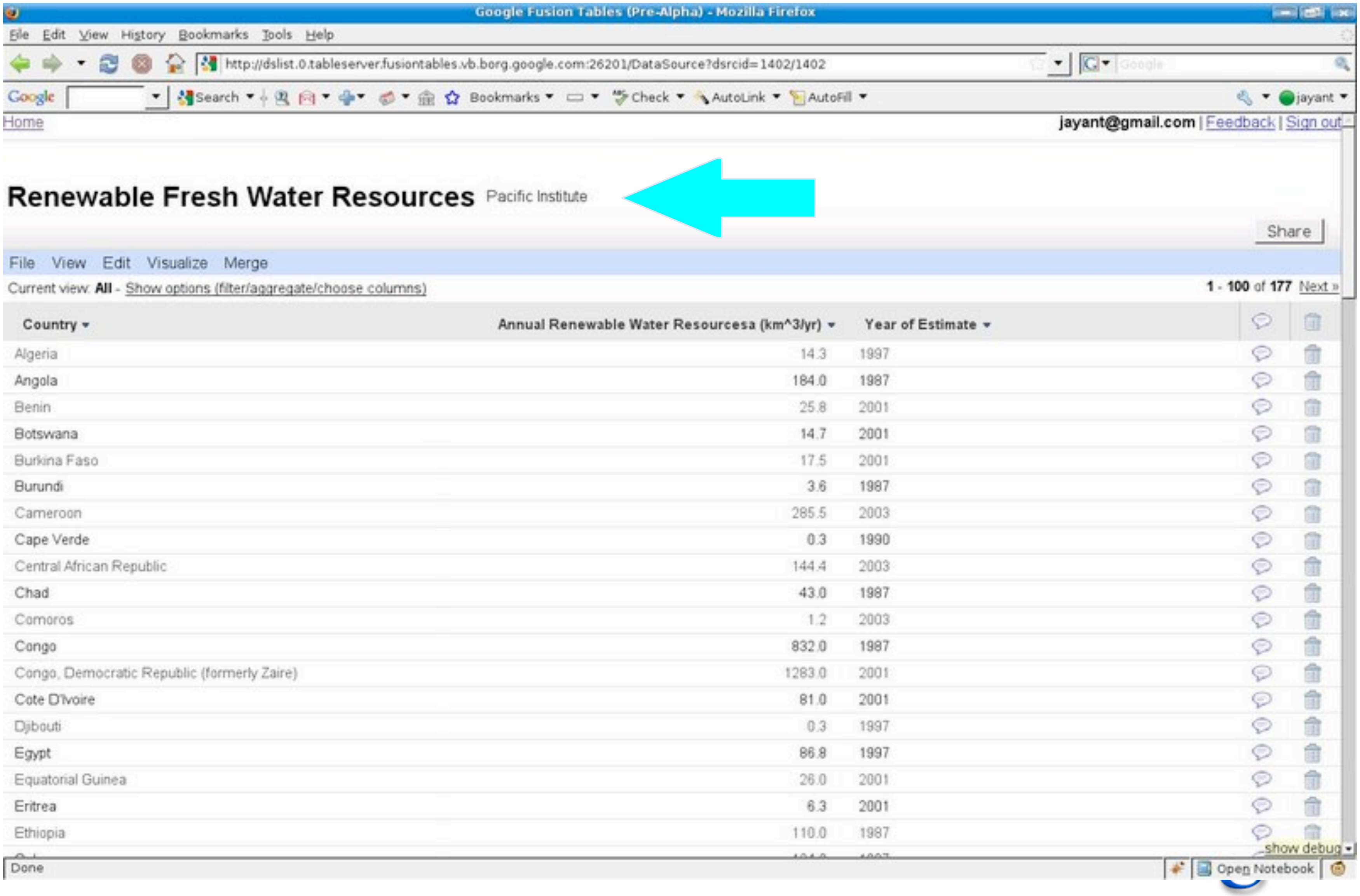


Country	Year	Total Fresh	Per-capita Wa	Domestic u	Industrial U	Agricultural	Domestic us	Industrial Use	Agricultural u	Source	2005 Popul
Belize	2000	0.15	555.56	7	73	20	38.89	405.56	111.11	l	0.27
Canada	1996	44.72	1385.81	19.56	68.67	11.77	271.05	951.63	163.14	o	32.27
Singapore	1975	0.19	43.88	45	51	4	19.75	22.38	1.76	c	4.33
Austria	1999	3.67	448.11	35.09	63.92	0.99	157.23	286.43	4.46	m	8.19
Belgium	1998	7.44	714.01	13.3	85.47	1.22	94.97	610.29	8.75	k	10.42
Bulgaria	2003	6.92	895.21	3.02	78.22	18.76	27	700.25	167.97	m	7.73
Czech Republic	2002	1.91	186.89	40.76	57.1	2.14	76.18	106.71	4.01	m	10.22
Finland	1999	2.33	443.81	13.65	83.69	2.66	60.56	371.43	11.82	m	5.25
France	2000	33.16	548.1	15.72	74.48	9.8	86.18	408.2	53.71	m	60.5
Germany	2001	38.01	459.67	12.36	67.85	19.79	56.8	311.9	90.97	m	82.69
Hungary	2001	21.03	2082.18	9.24	58.67	32.09	192.48	1221.54	668.16	m	10.1
Iceland	2003	0.17	566.67	34.02	65.85	0.13	192.8	373.12	0.74	m	0.3
Ireland	1994	1.18	284.34	22.62	77.36	0.02	64.33	219.96	0.05	m	4.15

Fusion Tables Example (4)

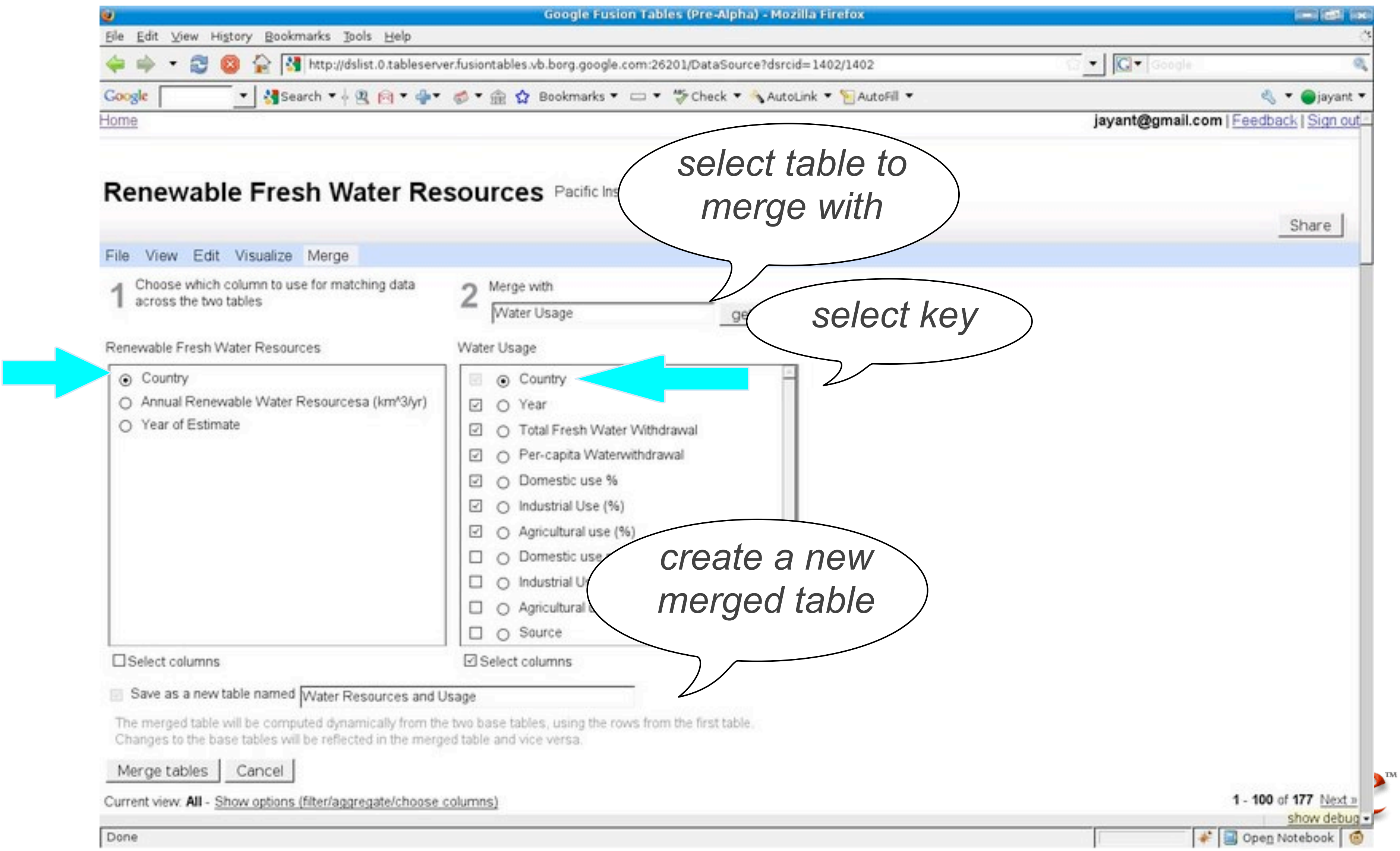


Fusion Tables Example (5)

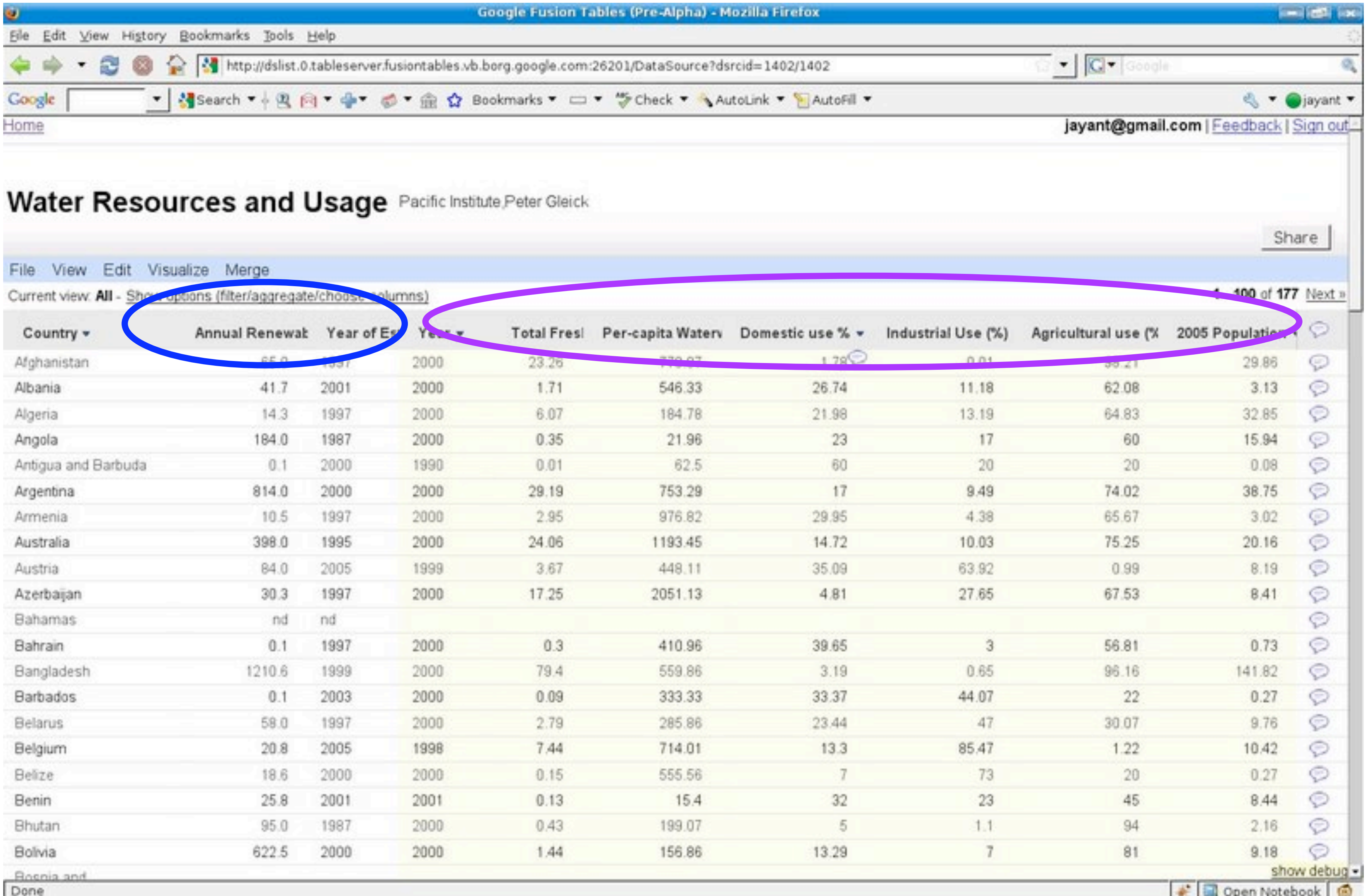


Country	Annual Renewable Water Resourcesa (km^3/yr)	Year of Estimate
Algeria	14.3	1997
Angola	184.0	1987
Benin	25.8	2001
Botswana	14.7	2001
Burkina Faso	17.5	2001
Burundi	3.6	1987
Cameroon	285.5	2003
Cape Verde	0.3	1990
Central African Republic	144.4	2003
Chad	43.0	1987
Comoros	1.2	2003
Congo	832.0	1987
Congo, Democratic Republic (formerly Zaire)	1283.0	2001
Cote D'Ivoire	81.0	2001
Djibouti	0.3	1997
Egypt	86.8	1997
Equatorial Guinea	26.0	2001
Eritrea	6.3	2001
Ethiopia	110.0	1987

Fusion Tables Example (6)

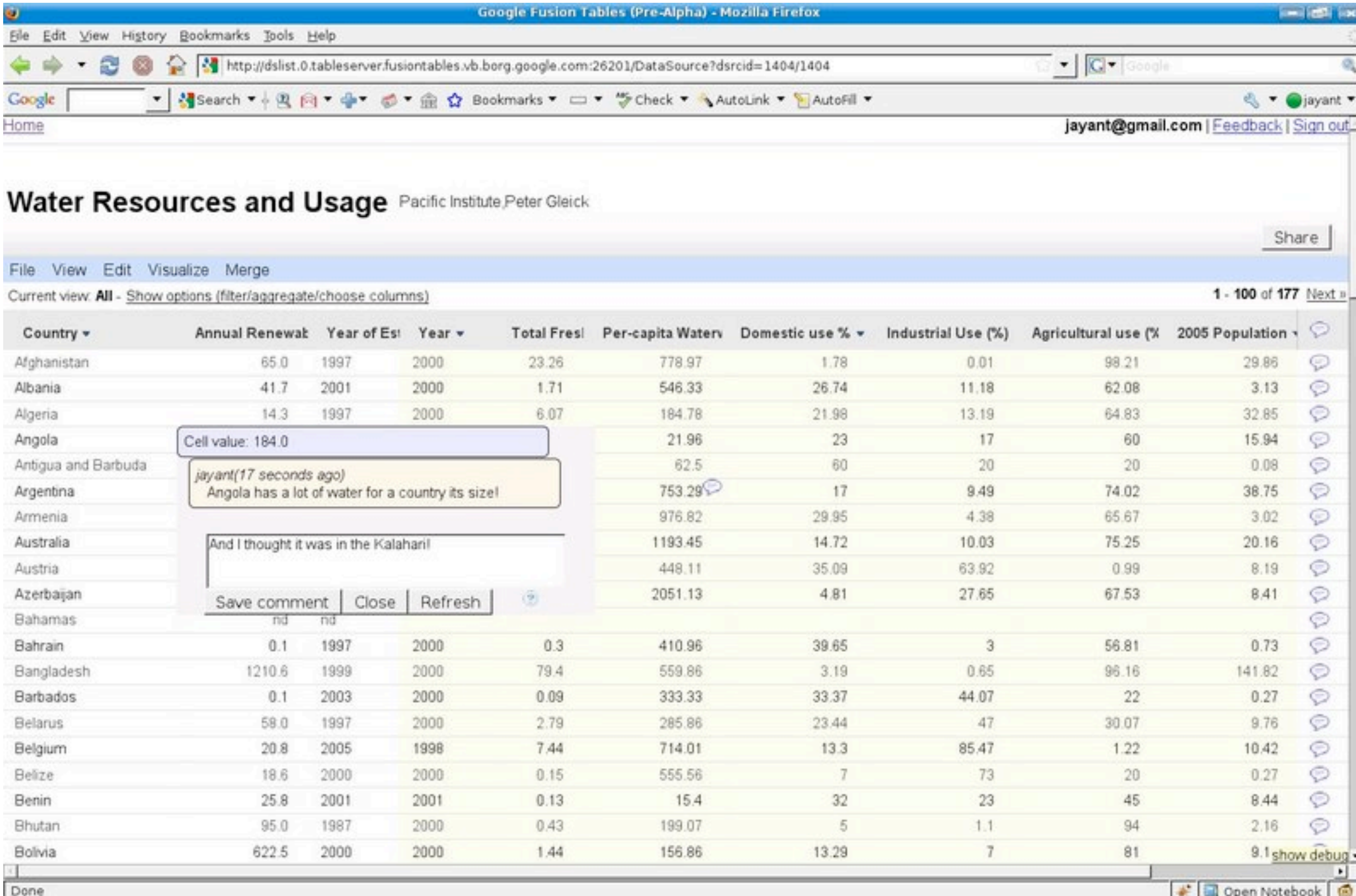


Fusion Tables Example (7)



Country	Annual Renewal	Year of Es	Year	Total Fresl	Per-capita Waterv	Domestic use %	Industrial Use (%)	Agricultural use (%	2005 Population
Afghanistan	[illegible]	[illegible]	2000	23.26	[illegible]	1.78	[illegible]	[illegible]	29.86
Albania	41.7	2001	2000	1.71	546.33	26.74	11.18	62.08	3.13
Algeria	14.3	1997	2000	6.07	184.78	21.98	13.19	64.83	32.85
Angola	184.0	1987	2000	0.35	21.96	23	17	60	15.94
Antigua and Barbuda	0.1	2000	1990	0.01	62.5	60	20	20	0.08
Argentina	814.0	2000	2000	29.19	753.29	17	9.49	74.02	38.75
Armenia	10.5	1997	2000	2.95	976.82	29.95	4.38	65.67	3.02
Australia	398.0	1995	2000	24.06	1193.45	14.72	10.03	75.25	20.16
Austria	84.0	2005	1999	3.67	448.11	35.09	63.92	0.99	8.19
Azerbaijan	30.3	1997	2000	17.25	2051.13	4.81	27.65	67.53	8.41
Bahamas	nd	nd							
Bahrain	0.1	1997	2000	0.3	410.96	39.65	3	56.81	0.73
Bangladesh	1210.6	1999	2000	79.4	559.86	3.19	0.65	96.16	141.82
Barbados	0.1	2003	2000	0.09	333.33	33.37	44.07	22	0.27
Belarus	58.0	1997	2000	2.79	285.86	23.44	47	30.07	9.76
Belgium	20.8	2005	1998	7.44	714.01	13.3	85.47	1.22	10.42
Belize	18.6	2000	2000	0.15	555.56	7	73	20	0.27
Benin	25.8	2001	2001	0.13	15.4	32	23	45	8.44
Bhutan	95.0	1987	2000	0.43	199.07	5	1.1	94	2.16
Bolivia	622.5	2000	2000	1.44	156.86	13.29	7	81	9.18
Bosnia and									

Fusion Tables Example (8)



Country	Annual Renewal	Year of Es	Year	Total Fres	Per-capita Water	Domestic use %	Industrial Use (%)	Agricultural use (%	2005 Population
Afghanistan	65.0	1997	2000	23.26	778.97	1.78	0.01	98.21	29.86
Albania	41.7	2001	2000	1.71	546.33	26.74	11.18	62.08	3.13
Algeria	14.3	1997	2000	6.07	184.78	21.98	13.19	64.83	32.85
Angola					21.96	23	17	60	15.94
Antigua and Barbuda					62.5	60	20	20	0.08
Argentina					753.29	17	9.49	74.02	38.75
Armenia					976.82	29.95	4.38	65.67	3.02
Australia					1193.45	14.72	10.03	75.25	20.16
Austria					448.11	35.09	63.92	0.99	8.19
Azerbaijan					2051.13	4.81	27.65	67.53	8.41
Bahamas	nd	nd							
Bahrain	0.1	1997	2000	0.3	410.96	39.65	3	56.81	0.73
Bangladesh	1210.6	1999	2000	79.4	559.86	3.19	0.65	96.16	141.82
Barbados	0.1	2003	2000	0.09	333.33	33.37	44.07	22	0.27
Belarus	58.0	1997	2000	2.79	285.86	23.44	47	30.07	9.76
Belgium	20.8	2005	1998	7.44	714.01	13.3	85.47	1.22	10.42
Belize	18.6	2000	2000	0.15	555.56	7	73	20	0.27
Benin	25.8	2001	2001	0.13	15.4	32	23	45	8.44
Bhutan	95.0	1987	2000	0.43	199.07	5	1.1	94	2.16
Bolivia	622.5	2000	2000	1.44	156.86	13.29	7	81	9.1

Where is the research?

Motivation: commercial databases have become far too intimidating for common users

- Not being used even when there is a fit

Redefining how structured data management can be done (on the Web)

- Focusing on supporting common user activities
- Making data management primitives easily available
- Blurring the notion of database boundaries

Excitement in Distributed Systems

- Size of user community
- Storage Scale (requiring various characteristics)
  - E.g., security, privacy, availability,
- Processing Scale
  - High performance batch processing
  - High throughput
  - Low latency
- Rapid dynamics
- Highly variable end-user devices
- Communication Scale
  - Bandwidth
  - Endpoints
- Efficiency
  - Equipment
  - Communication
  - Power
  - Management
- Extensibility
- Compliance
- *And more to come, no doubt*

Ideal Distributed Computing

Large networked clusters grow in a fully distributed world

- Arbitrarily high volume transactions
- And, various, partitionable batch process for learning, fusion, etc.
- Network
  - Response-time and bandwidth as needed
- Cluster Processing, or “Cloud Computing” growing ever larger
  - Massive parallelism to hit sweet spot of capital & operating efficiency
- Distributed computing
  - Data sharing, function shipping, as needed
  - Connected and disconnected operation, as seamless as possible
  - Auto balancing of loads between client device and cloud elements
  - Emphasis on manageability (newly, to handle consumers’ many endpoints)
- Significant efficiency gains

Hybrid, Not Artificial, Intelligence

Hybrid, not Artificial, Intelligence

- “Artificial Intelligence” aimed at having computers as capable as people, often in very broad problem domains
- It has proven more useful for computers rather:
  - To extend the capability of people, not in isolation
  - To focus on more specific problem areas
- Aggregation of user responses has proven extremely valuable in learning
- Examples
  - Feedback in Information Retrieval; e.g., in ranking or spelling correction
  - Machine learning; e.g., image content analysis, speech recognition with semi-supervised learning
- Another example of bottom up successes

Google™

My Long-held View on Semantics, Syntax, & Learning

- Large scale learning has proven surprisingly effective
- Learning is occurring over increasingly variegated features:
  - Both Semantic
  - And Syntactic, and generated in multiple ways
- In my WWW 2002 (Architecting Knowledge Middleware) and Semantic Web 2005 Keynotes, I referred to this as *The Combination Hypothesis*
- Today, I would refine this as the combination of approaches *and* learning from people.

Google™

Fantastic Opportunities Abound

Just a Few Opportunities

- New Interfaces and applications with mass customization
  – Implications on every vertical
  – Examples: health, government, entertainment
- Virtually unlimited data storage
- Ever improving system “understanding”
- Increasingly fluid partnership between people and computation
- Fundamental changes in the methods of science
- Opportunities for optimization in many more domains

There are no real limits

We Desire A Stronger Relationship with Academe (1)

Google tremendously values talented people and education

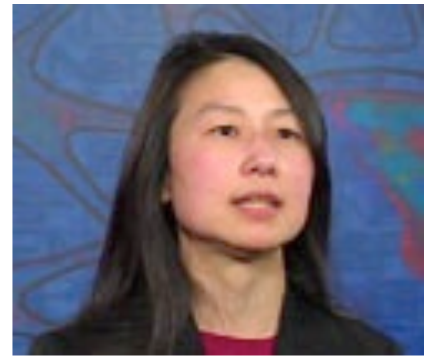
- We have preferred a bottom-up approach:
  - Google collaborating with faculty and students for *mutual* benefit
    - Google: Knowledge of challenging problems, skilled employees, opportunities for internship/sabbatical, money
    - University: Faculty and student skills, breadth/depth of perspective
  - Internships
  - Visiting Faculty
- University Research Grants with moderate funding
  - Roughly 100 worldwide
  - Recommend proposals developed with advice/perspective of Google employees

Google™

Technology Leadership: NSF CLuE Program

National Science Foundation

WHERE DISCOVERIES BEGIN



Jeannette Wing

CluE in the Search for Data-Intensive (Cloud) Computing

Unique relationship between Google, IBM and NSF allows academic computing research community to access large-scale computer cluster for cloud computing.

Google™

Google Relationship with Academe (cont.)

- Check out growing sites:
  - //research.google.com
  - //research.google.com/university
  - //code.google.com
  - //code.google.com/edu
- Technology Round Table Videos (new!)
  - http://research.google.com/roundtable/

Summary: Innovation at Google

- Strong commitment to broadly advancing technology due to our mission: some *grand challenge* problems
- **Leverage**
  - **Scale in processing and information, and Usage**
  - **Learning and Empiricism**
- We try to minimize the distance between Research and Development
  - Recognize that putting ideas into production is often as challenging and fun as inventing the idea
- Google will attempt to use its capabilities beneficially to foster research, education, and advancement broadly
- Go to http://research.google.com for more information and resources

Google™

Thank you!