

Research @ Google™

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

Google™

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

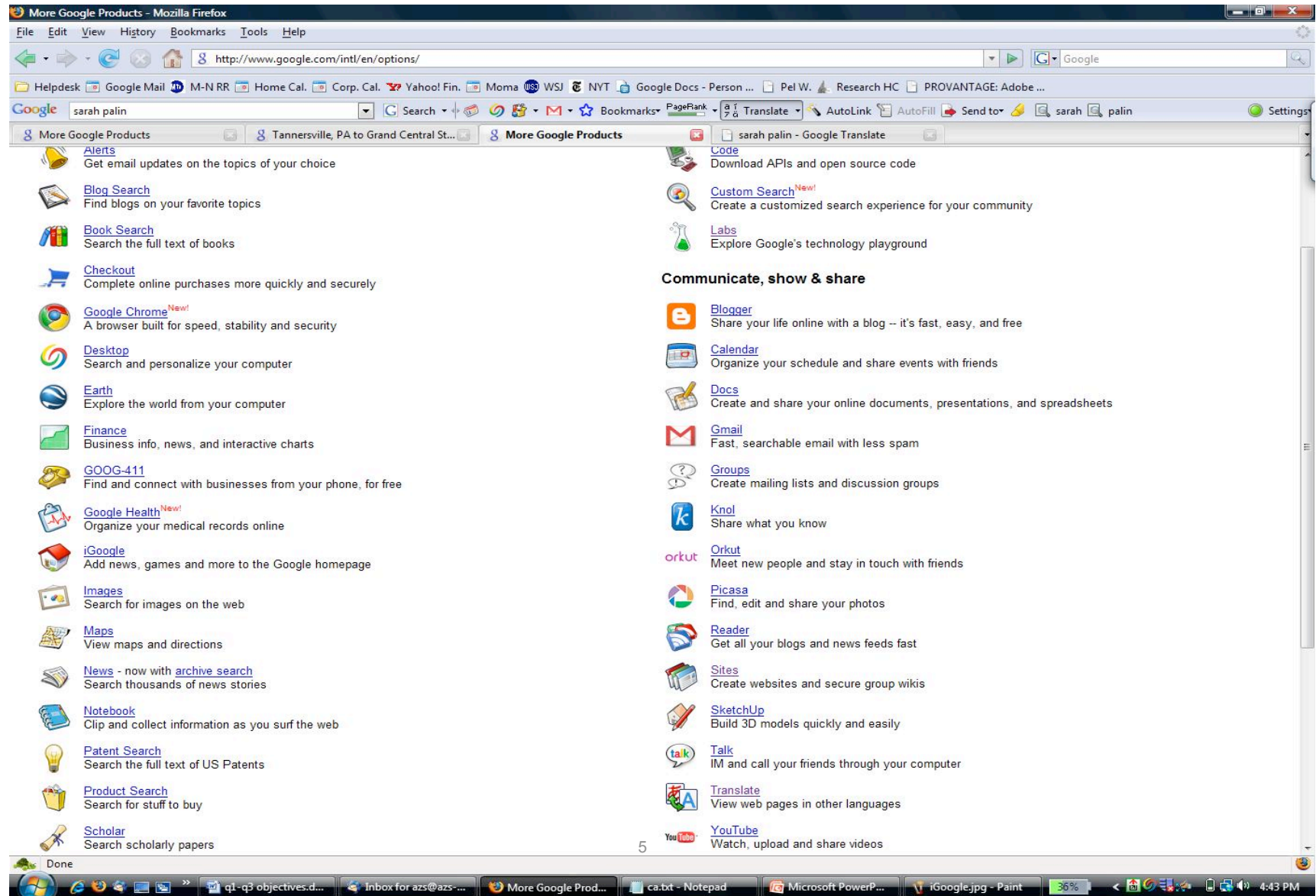
**Organizing the world's information and
Making it universally accessible and useful.**



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

How (3): Empiricism - Let Measurement & Feedback Rule

108 milliseconds *-0.54* *2.7M* *0.55060*

\$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*

\$0.303 CPA

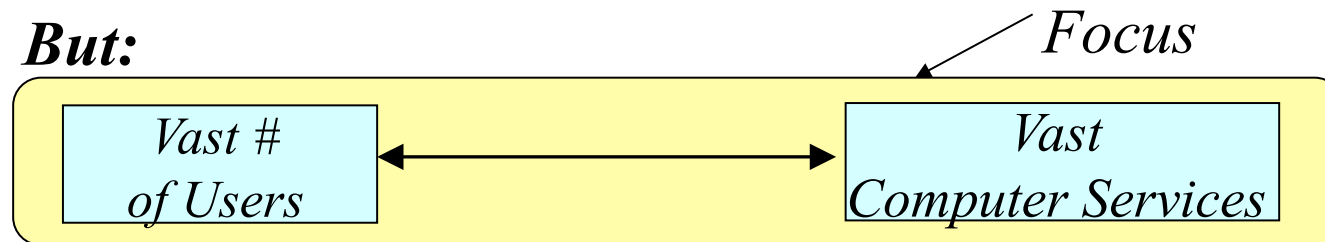
6.55

\$7,660,400

108



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 and 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*)

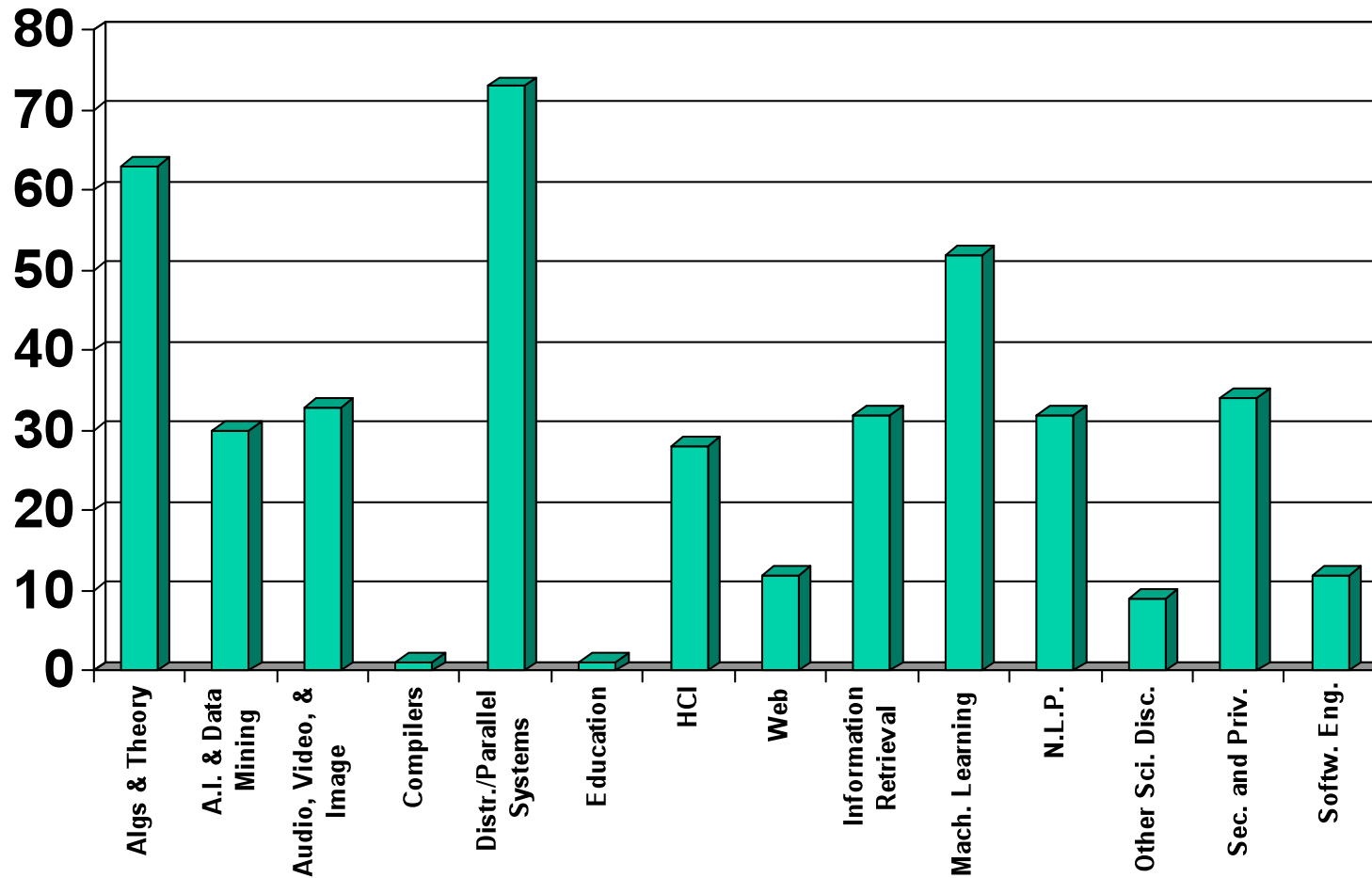
Publications

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

| <i>Year</i> | <i>Number of Publications</i> |
|-------------|-------------------------------|
| <i>2004</i> | <i>15</i> |
| <i>2005</i> | <i>54</i> |
| <i>2006</i> | <i>122</i> |
| <i>2007</i> | <i>197</i> |
| <i>2008</i> | <i>216</i> |
| <i>2009</i> | <i>250+ (p)</i> |



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, Jeffrey 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



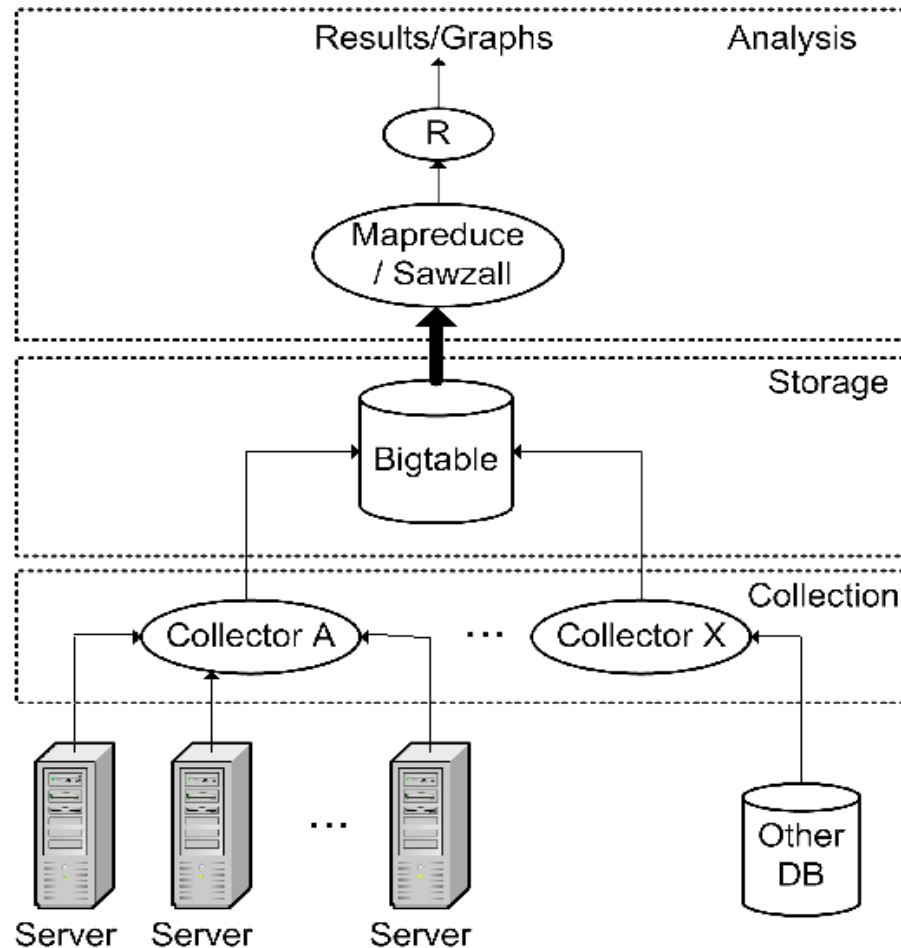
Disk drive failures are a significant problem

- 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



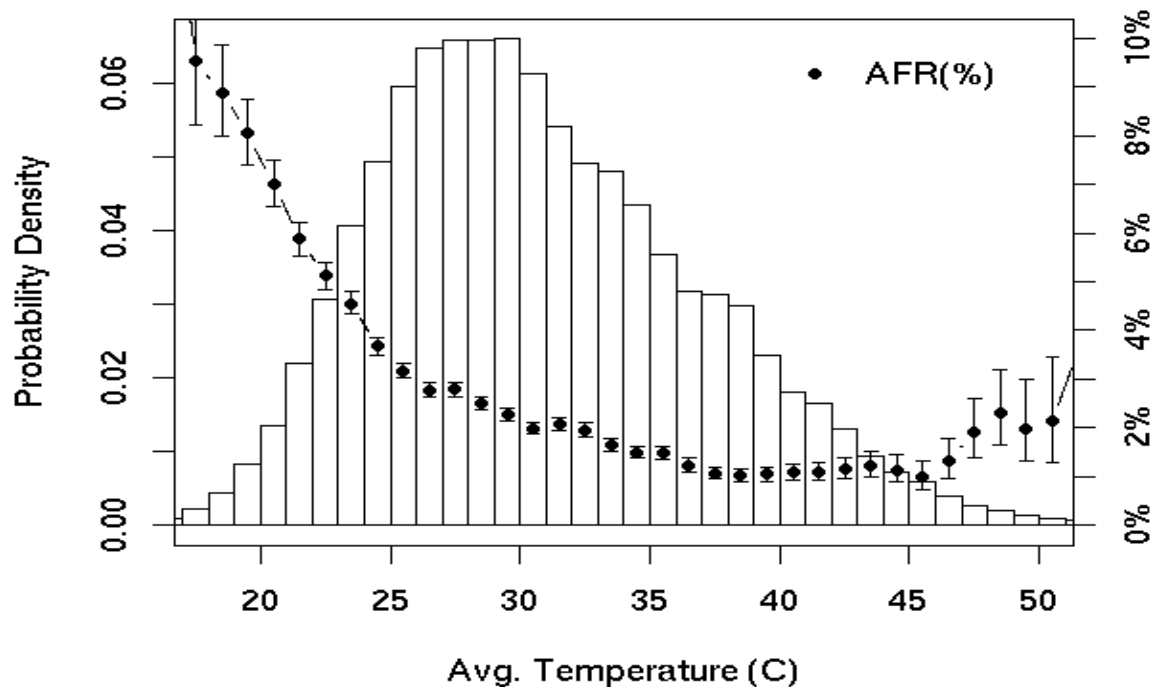
System Health Infrastructure

- 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/

Google Code
 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](#).

The Chromium blog

[Graphics in Google Chrome](#)
 Oct 22, 2008
 Google Chrome uses a library called Skia, which is also the graphics engine behind Google's Android mobile OS. The two ...

[Content, not 'Chrome'](#)
 Oct 16, 2008
 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](#)
 Oct 10, 2008
 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 Macfirst? 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%

Microsoft E... China Downloads Retailers SI... Inbox for az... Chromium ... China01.pp... Untitled - P... 100% 1:32 AM





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 »](#)

[Site Terms of Service](#) - [Privacy Policy](#) - [Brand Guidelines](#)

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

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

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

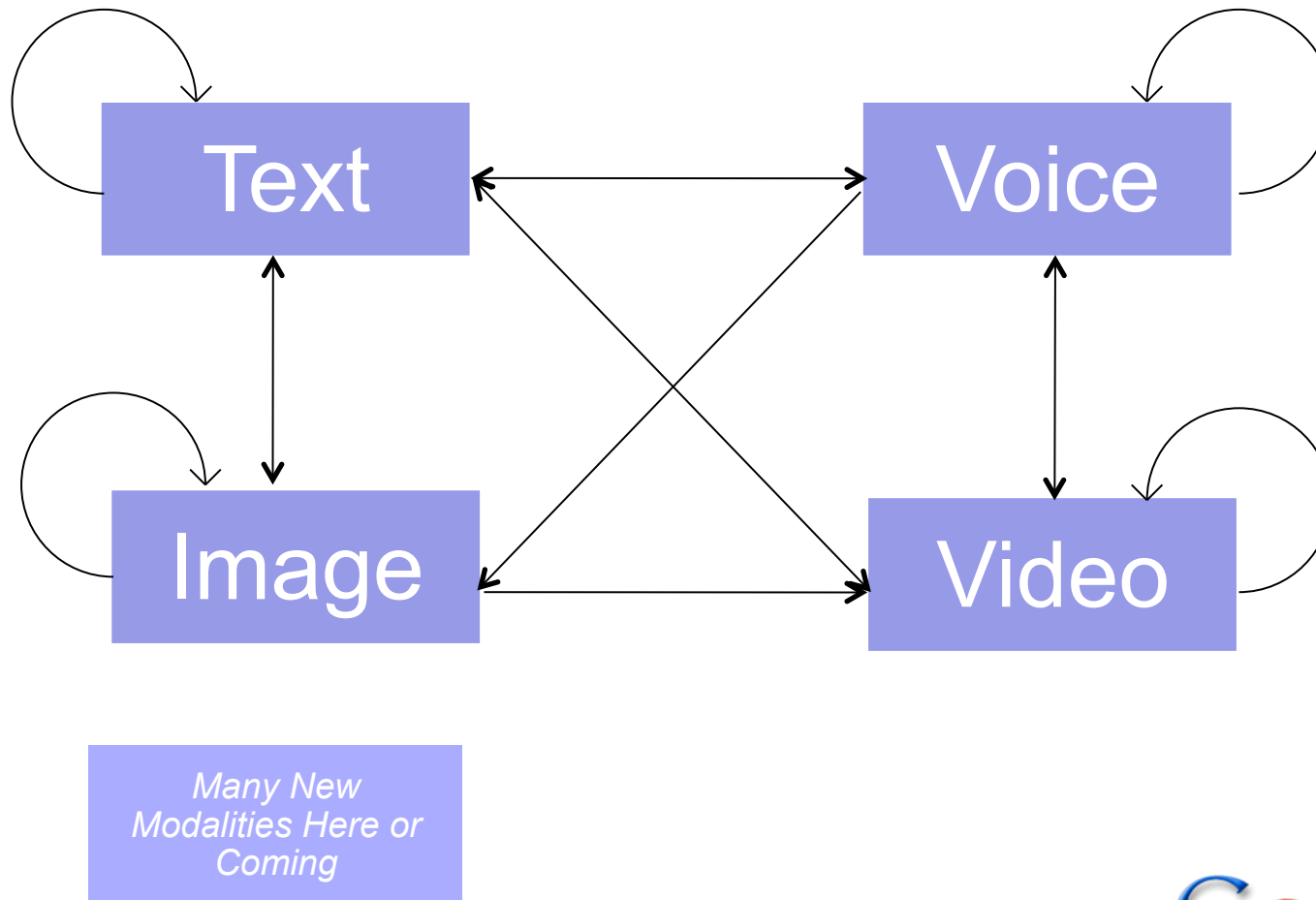
| D: The set of all end-user access devices | L: The set of all human languages | M: The set of all modalities | C: The set of all corpora |
|--|---|---|--|
| Personal Computers Phone Media Players/Readers Telematics Set-top Boxes Appliances Health devices ... | Current languages Historical languages Other forms of human notation Possible language specialization Formal languages ... | Text Image Audio Video Graphics Other sensor-based data ... | The normal web The deep web Periodicals Books Catalogs Blogs <u>Universal Geodata</u> Scientific datasets Health data ... |

Types of Transparent Processing

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



Fluidity Among the Modalities



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:

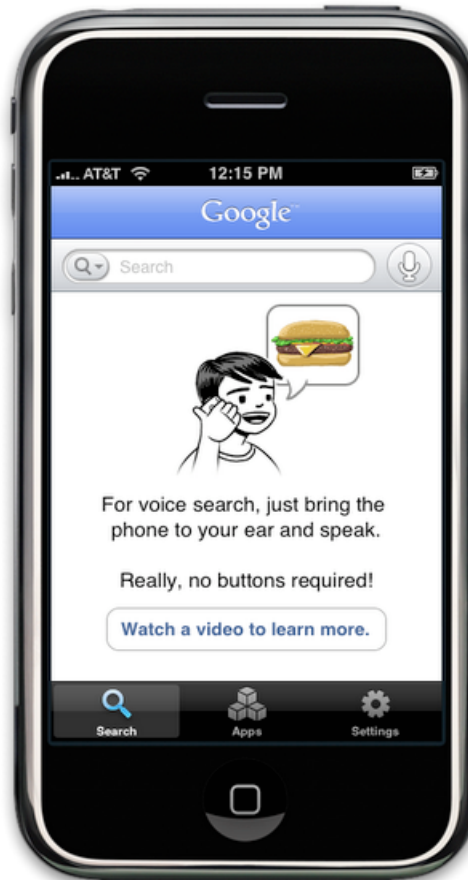
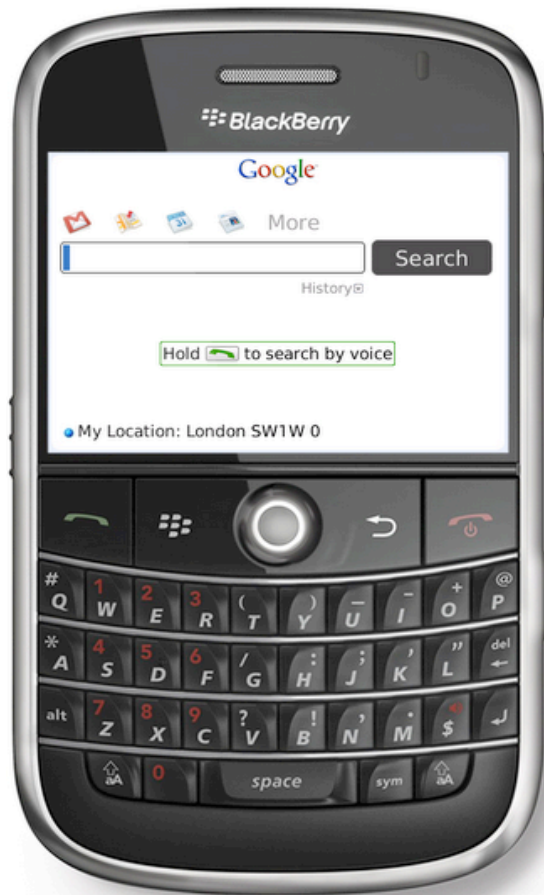
- *response-time, coverage, freshness, corpus dependency (YouTube, image, mobile)*
- *domain dependent: rea -> real madrid good suggestion in Spain*
- *diversity (danger of popularity), filtering out duplicates, inappropriate results, etc.*

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:

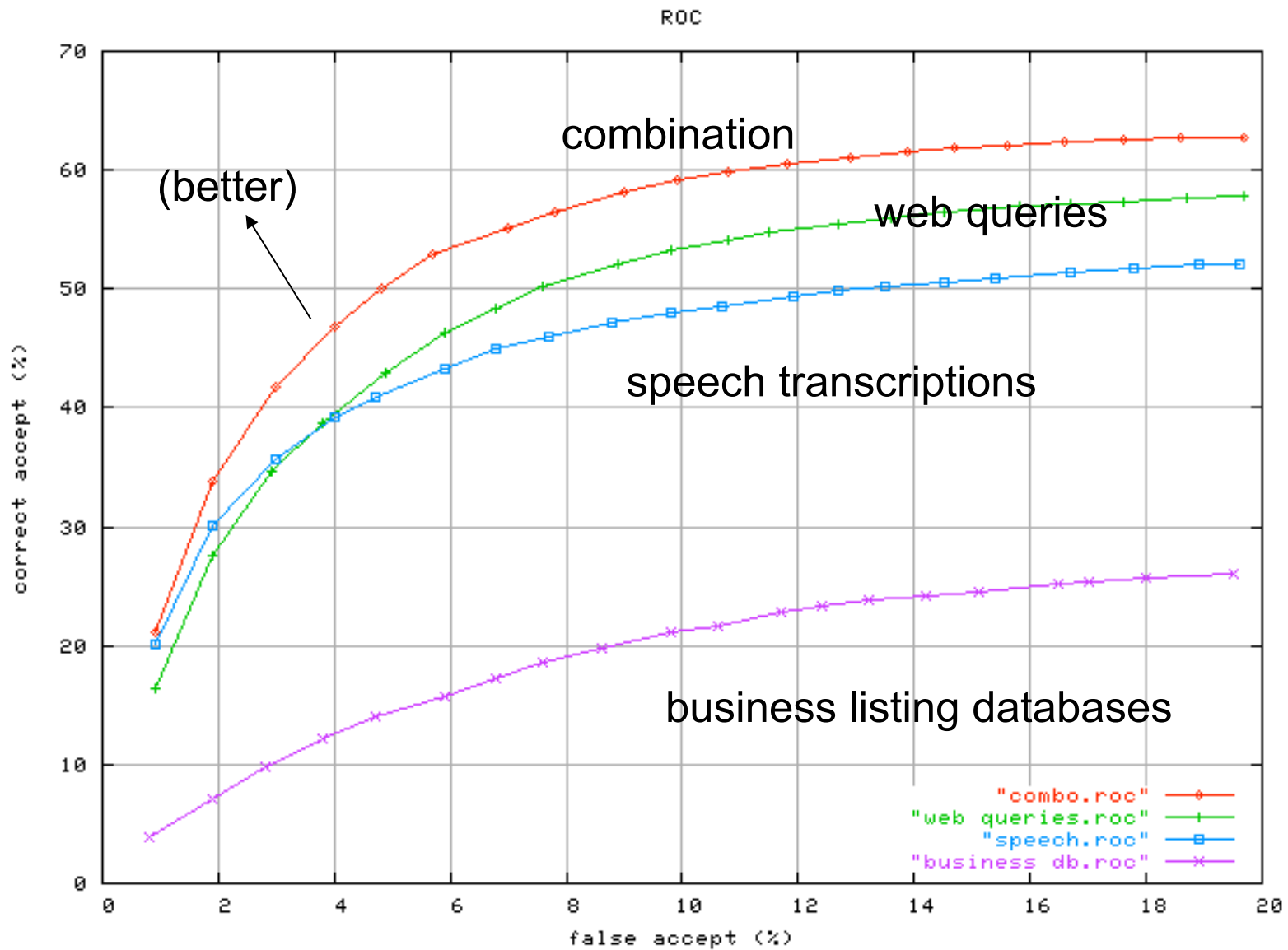
- Scalable technology
- Data inputs: Query logs, voice logs
- Compute power



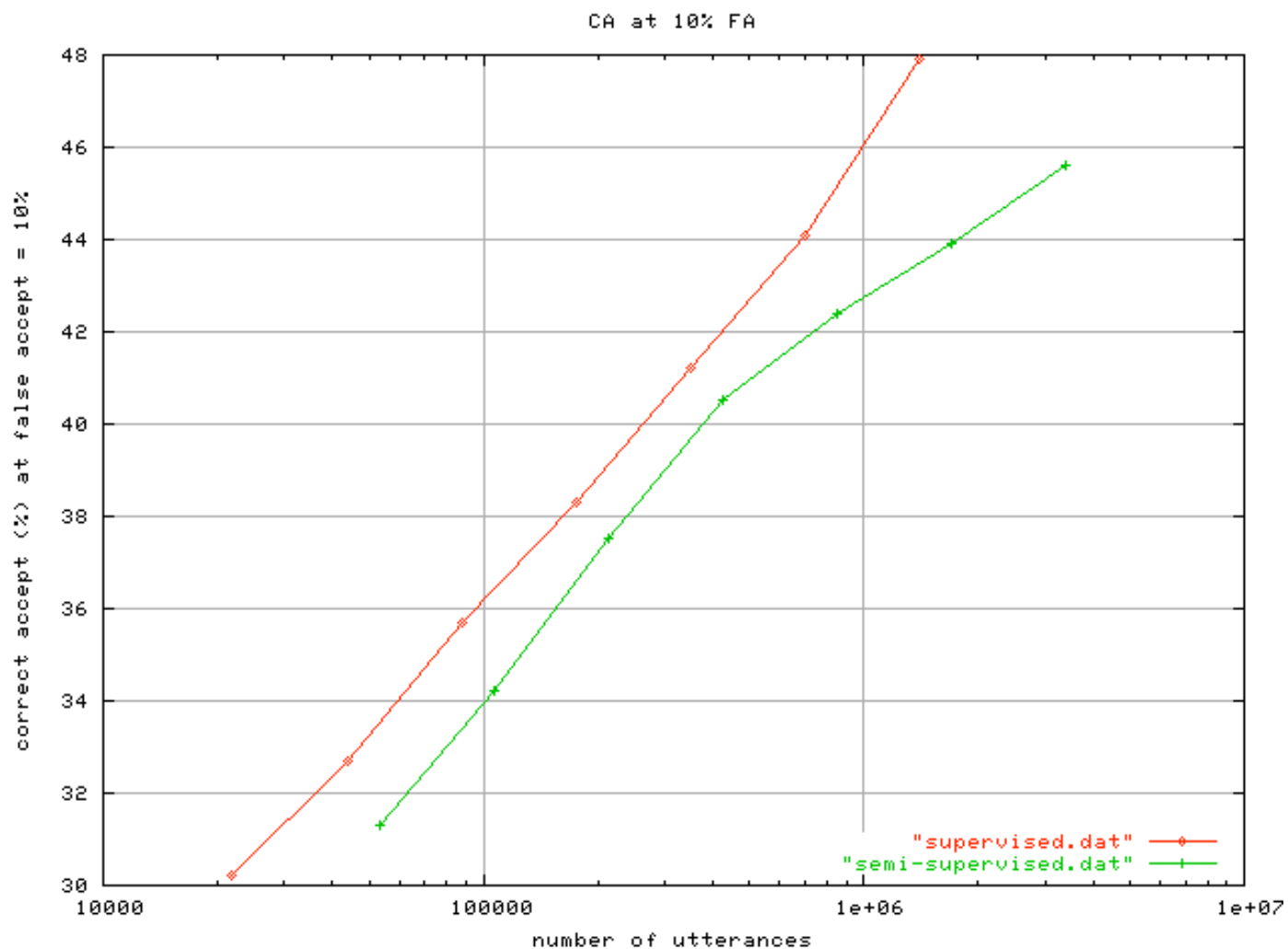
Choosing a data source

| 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

The screenshot displays the Google Voice web interface in a browser window. The address bar shows the URL <https://www.google.com/voice#inbox>. The page title is "Google Voice - Inbox". The user's name is "Johan.Schalkwyk | (914) 290-6482". The interface includes a search bar, navigation links (Call, SMS), and a list of messages. Each message entry shows the sender's name and phone number, the time received, a play button with a duration, and the transcribed text. A "Transcript useful?" checkbox is present for each message.

Message 1: (212) 565-8996 - New York City, NY, 9/9/08 6:38 AM. Transcript: "hello this is a test of this music playing in the background i just wonder how the system does in so hit me back at nine one four seven one three eight eight seven four eight okay it looks like i got a voicemail this and see you later bye".

Message 2: (212) 565-8996 - New York City, NY, 9/9/08 6:25 AM. Transcript: "hi i wonder if we can have dinner tonight at red lobster it's carsten the ark you can have some fun discussing there i just had to rebatch at the US open if you want to give me a message you can call me back or send me an email at on call craig at G mail dot com see you later bye".

Message 3: (212) 565-8996 - New York City, NY, 9/9/08 6:23 AM. Transcript: "hi this is john i'm just calling to test the grand central system i'm calling from colonial city i'm at the corner of fourteenth avenue and eighth avenue please call me back my number is (914) 713-8748 that's my home number bye".

Message 4: (650) 265-1193 - Mountain View, CA, 9/9/08 6:19 AM. Transcript: "Hello Googlers, and welcome to Google Voice. You are the early testers and we'd love to hear what you think about it. To learn more about it, see our Known Issues list and to give feedback, go to go/voice-testers. Thanks and we hope you enjoy it."

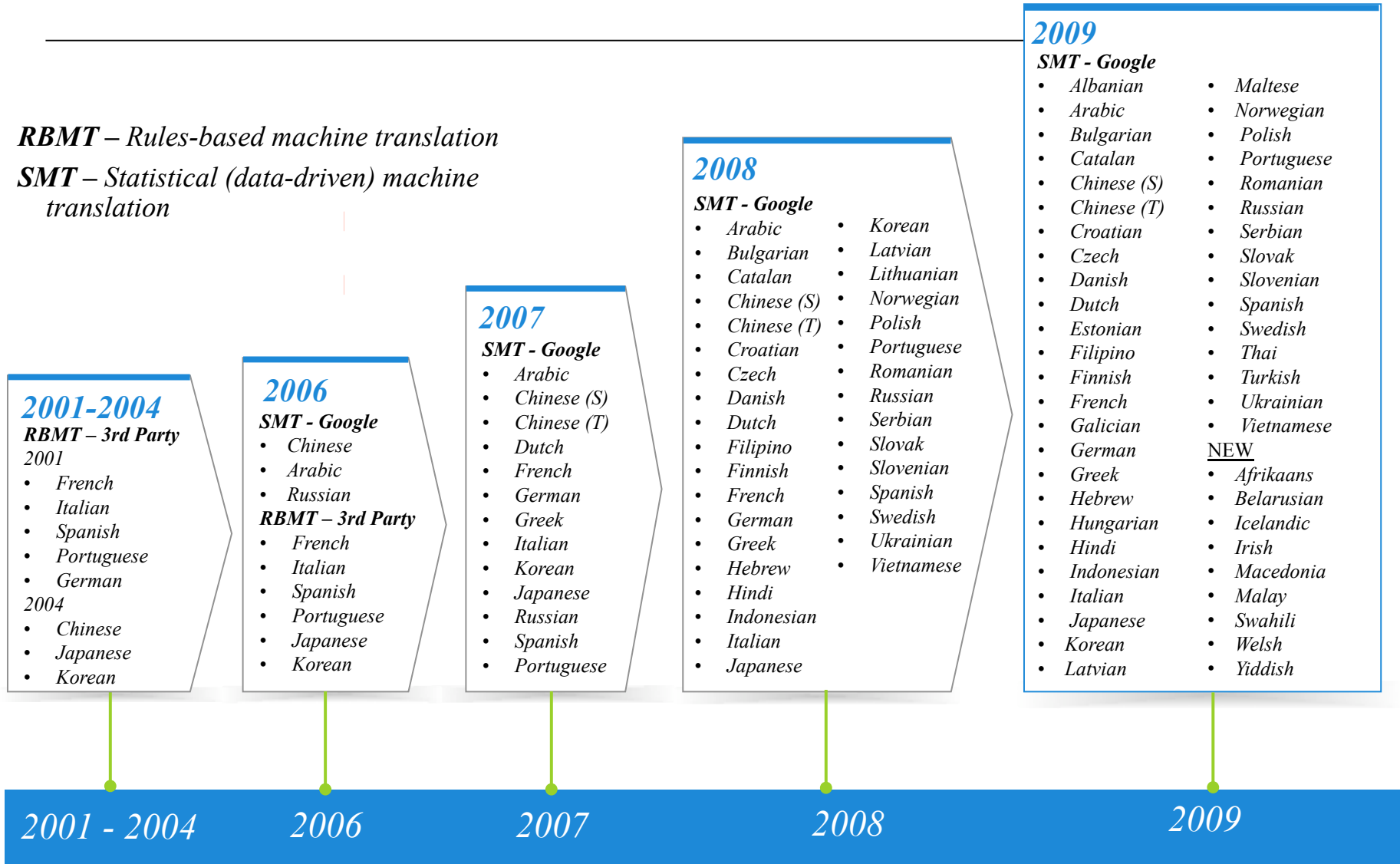
©2009 Google - [Terms](#) - [Blog](#) - [Google Home](#)



Google Translate

RBMT – Rules-based machine translation

SMT – Statistical (data-driven) machine translation



Web Translation

The screenshot shows a web browser window with Google Translate. The browser's address bar shows the URL: http://translate.google.com/translate?js=y&prev=_t&hl=en&ie=UTF-8&u=http://www.seoul.co.kr/news/dailyList.php&sl=ko&tl=en&history_state0=...

The page content is a news article from Seoul Shinmun. The article title is: **[1 years now, the world financial crisis] ... economic recovery, signs of inflation ... I 2009-09-15**

The article text includes: "International raw material prices are deulssseokyigo. Compared with the beginning of the year when the gold price has already jumped more than 20%. The (鎳) and other key products such as grain prices on the rise is too strong. \$ 70 per barrel of oil inside, while maintaining the three-digit neomboneun. Earlier this global liquidity ..."

Other articles listed include: "Teachers believe that recommendation? Please do not? I 2009-09-15", "Joint Chiefs Chairman Koo yukchamchongjang more I 2009-09-15", "HR Block ... 64 31.32 rumors term 'generational change' I 2009-09-15", "[Morning Briefing] 7.6 billion in new growth engines, supporting professional training I 2009-09-15", and "军dajejangseup nominee profile I 2009-09-15".

Cross-language search

The screenshot shows a Google Translate search interface. The search term is "Digital SLR Camera", which has been translated to Korean as "디지털 SLR 카메라". The search results are presented in two columns: English translation on the left and original Korean text on the right. The results include various web pages such as "Korea dika-D-SLR digital camera", "DSLR mania - Digital SLR Camera", "SLR CLUB", "[Brain Box] LG Co., Canon EOS 350D Digital SLR Camera Give", "Hobby User News - 3900 megapixel digital SLR camera market", "Mikseusyap [pre-owned digital SLR camera]", and "Visual media - itviewpoint.com - Digital SLR Camera price 'endless crash'". Each result includes a brief description and a link to the original page.

Translated Search

Search for: Digital SLR Camera Translated to: 디지털 SLR 카메라 - Not quite right? Edit

My language: English Search pages written in: Korean

Translate and Search

Translated results from Korean web pages Results 1 - 10 of about 392,000 for 디지털 SLR 카메라.

| English translation | Original Korean - Hide Korean results |
|---|--|
| Korea dika-D-SLR digital camera, professional shopping mall price dika Digital Camera Shop. Memory cards, power supplies, digital printers, digital camcorders and sales. www.dicakorea.co.kr/ - 56k - Cached | 디카코리아 디지털카메라 전문쇼핑몰-D.SLR 디카최저가 쇼핑몰 디지털 카메라 쇼핑몰 메모리카드, 전원장치, 디지털프린터, 디지털캠코더 등 판매. www.dicakorea.co.kr/ - 56k - 저장된 페이지 |
| DSLR mania - Digital SLR Camera [DSLR] cameras are, and Canon DSLR cameras... Digital SLR Camera [DSLR] cameras are, and Canon DSLR Camera [DSLR-related]...DSLR Course 'category of the other posts. 2007-10-09 14:50:35 Digital SLR Camera [DSLR] cameras are,... blog.daum.net/dslrmail/290360 - 98k - Cached | DSLR 마니아 - 디지털 SLR카메라[DSLR]카메라 소개, 캐논 DSLR카메라 ... 디지털 SLR카메라[DSLR]카메라소개, 캐논 DSLR카메라 [DSLR관련] ... DSLR 강좌 카테고리 의 다른 글. 2007-10-09 14:50:35 디지털 SLR카메라[DSLR]카메라소개, ... blog.daum.net/dslrmail/290360 - 98k - 저장된 페이지 |
| SLR CLUB slrclub.com/ - 1k - Cached | SLR CLUB slrclub.com/ - 1k - 저장된 페이지 |
| [Brain Box] LG Co., Canon EOS 350D Digital SLR Camera Give Brainbox, Benchmark, www.brainbox.co.kr , brainbox.co.kr , www.brainbox.kr , benchmark.kr , benchmarking, Brain Box, beubak, joint purchasing, tools, hardware,... www.brainbox.co.kr/news/view.asp?id=8051 - 123k - Cached | [브레인박스] LG상사, 캐논 EOS 350D 디지털 SLR 카메라 내보 Brainbox, Benchmark, www.brainbox.co.kr , brainbox.co.kr , www.brainbox.kr , benchmark.kr , 벤치마크, 벤치마킹, 브레인박스, 보박, 공동구매, 공구, 하드웨어, ... www.brainbox.co.kr/news/view.asp?id=8051 - 123k - 저장된 페이지 |
| Hobby User News - 3900 megapixel digital SLR camera market Shiseido broadcast March (資生堂) was starring in a new campaign ad, she became the first thing mateoun topics ... 39 million pixel digital SLR camera market: 1... videogamerx.gamedonga.co.kr/zbxel/?mid=news_hobby&category=10867 ... - 80k - Cached | Hobby User News - 3900만 화소 디지털 SLR 카메라 출시 3월부터 방송되는 시세이도(資生堂)의 신제품 캠페인 광고에 출연하는 것으로, 엄마가 된 후 말은 첫번째 일이라 화제... 3900만 화소 디지털 SLR 카메라 출시 1 ... videogamerx.gamedonga.co.kr/zbxel/?mid=news_hobby&category=10867 ... - 80k - 저장된 페이지 |
| Mikseusyap [pre-owned digital SLR camera] Pre-digital SLR cameras, pre-digital SLR cameras, "all views. Canon (12), Nikon (5), Fuji (2), Pentax (8), Konica Minolta (6), Olympus (2)... www.mixshop.co.kr/shop/shopbrand.html?xcode=028&type=X - 43k - Cached | 믹스샵 [중고 디지털 SLR 카메라] 중고 디지털 SLR 카메라, 중고 디지털 SLR 카메라 > 전체조회. 캐논(12) · 니콘(5) · 후지(2) · 펜탁스(8) · 코니카 미놀타(6) · 올림푸스(2) ... www.mixshop.co.kr/shop/shopbrand.html?xcode=028&type=X - 43k - 저장된 페이지 |
| Visual media - itviewpoint.com - Digital SLR Camera price 'endless crash' Photos are used primarily to professional digital SLR (sasik half days) camera, the price is going down endlessly. Sigma camera supplies last century, the camera 2 million won in Korea early in the sales... itviewpoint.com/?category=145&mid=blog&listStyle=gallery&am... - 183k - Cached | 비주얼 미디어 - itviewpoint.com - 디지털 SLR 카메라 가격 '끝없는 추락' 사진 전문가들이 주로 사용하는 디지털 SLR(일안반사식) 카메라 가격이 끝없이 추락 하고 있다. 최근 시그마 카메라를 국내 공급하는 세기카메라는 200만원 초반에 판매 ... itviewpoint.com/?category=145&mid=blog&listStyle=gallery&am... - 183k - 저장된 페이지 |

Translate My Page Gadget

The screenshot shows the MTA website interface. On the left is a blue navigation menu with links: MTA Home, NYC Transit, Long Island Rail Road, Long Island Bus, Metro-North Railroad, Bridges & Tunnels, Capital Construction, and Bus Company. Below the menu are a search box, a 'Search' button, and a 'FAQs/Contact Us' button. A Google Translate widget is highlighted with a red circle; it features the Google logo, the text 'Google 翻訳', and a dropdown menu labeled '言語を選択'. Below the widget are icons for 'Accessibility Information', 'Metropass', and 'EZPass'. The main content area has a blue header with the MTA logo and 'State of New York', and the URL 'mta.info'. Navigation tabs for 'Schedules', 'Maps', and 'Service Advisories' are visible. The main content includes a photo of a street with a sign for 'CHENG' and a yellow taxi, with text explaining the installation of new street furniture to prevent flooding. Below this is a 'Google Transit' logo. On the right, there are three sections: 'Features' with a list of items like 'MTA Sustainability Webinar, June 2008', 'Report on Construction Excellence', and 'Strategic Information Technology Report - EnvisionIT'; 'Facts & Figures' with items like 'Financial Plan/Capital Program', 'Financial & Performance Indicators', 'MTA Budget', and 'New Fares'; and 'Regional Travel' with items like 'Trip Planner: Subway and Bus Directions', 'Regional Transit Alerts', 'Need Regional Transit Directions? Try Trips123', and 'Current Traffic and Transit Conditions'.



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

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 20º 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



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 | | |

CONTACTO

Pasado a ser patrocinador o expositor

PATROCINADORES PLATINO

PATROCINADORES DE ORO

PATROCINADORES PLATA

PATROCINADORES DE BRONCE

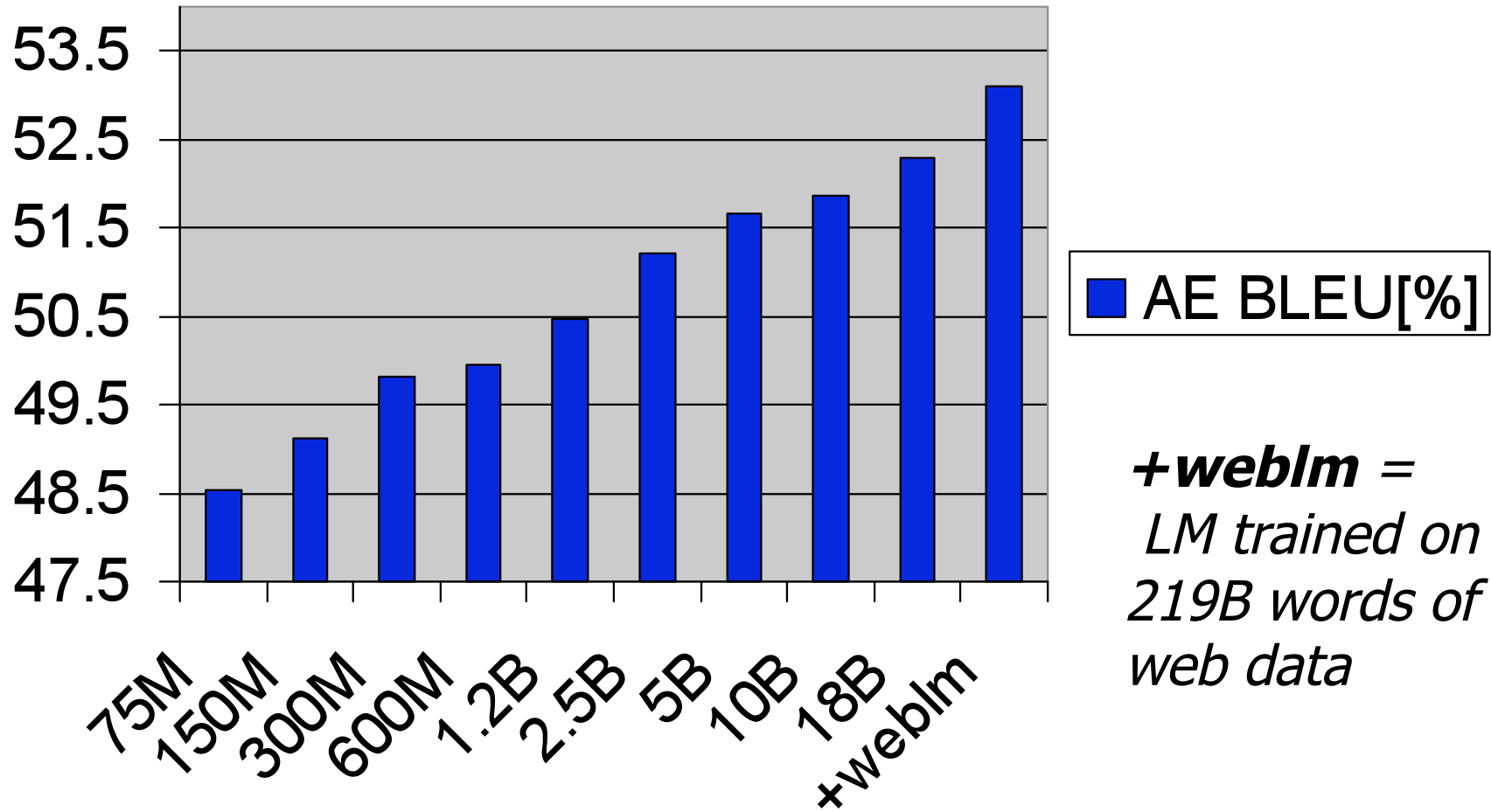


YouTube Caption Translation



The image is a screenshot of a YouTube video player. At the top left, the YouTube logo is visible with the text "Broadcast Yourself™ Worldwide | English". To the right of the logo, the text "Si" is partially visible. Below the logo is a navigation bar with buttons for "Home", "Videos", "Channels", and "Community". A search bar is located to the right of these buttons. The video title is "2/28/09: Your Weekly Address". The video frame shows Barack Obama in a dark suit and red tie, speaking in a room with bookshelves and an American flag. Arabic subtitles are displayed at the bottom of the video frame: "في القضاء على برامج اننا لسنا في حاجة ل". The video player interface at the bottom includes a progress bar, a play/pause button, a volume icon, an HD icon, and a full screen icon. The time displayed is 2:36 / 4:53.

Impact of data - More data is better ...



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

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

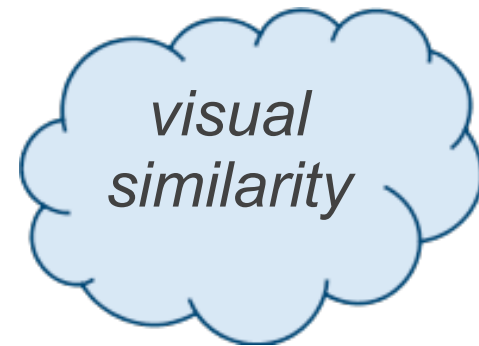
- A variety of features are used to determine visual similarity.



color
texture
shape



color
texture
shape



Similar Images in Image Search

Refine by the content of a specific image.

Different "faces" of Paris...



Paris ...
375 x 500 - 65k - jpg
[cruises.about.com](#)
[Similar images](#)



Paris: Ummm ...
299 x 330 - 72k - jpg
[blog.wired.com](#)
[Similar images](#)



Paris Architecture, Buildings
700 x 556 - 58k - jpg
[www.e-architect.co.uk](#)
[Similar images](#)



450 x 600 - 63k - jpg
[img2.travelblog.org](#)
[Similar images](#)



288 x 432 - 46k - jpg
[labellefrancetours.com](#)
[Similar images](#)



480 x 640 - 35k - jpg
[splendorsofeurope.com](#)



300 x 400 - 93k - jpg
[glassteelandstone.com](#)
[Similar images](#)



313 x 344 - 39k - jpg
[www.topnews.in](#)



317 x 354 - 36k - jpg
[www.topnews.in](#)
[Similar images](#)



405 x 500 - 43k - jpg
[makeup-makijaz.info](#)



498 x 600 - 166k - jpg
[celebritywonder.com](#)
[Similar images](#)



357 x 600 - 43k - jpg
[www.e-architect.co.uk](#)
[Similar images](#)



900 x 675 - 63k - jpg
[pixdaus.com](#)
[Similar images](#)

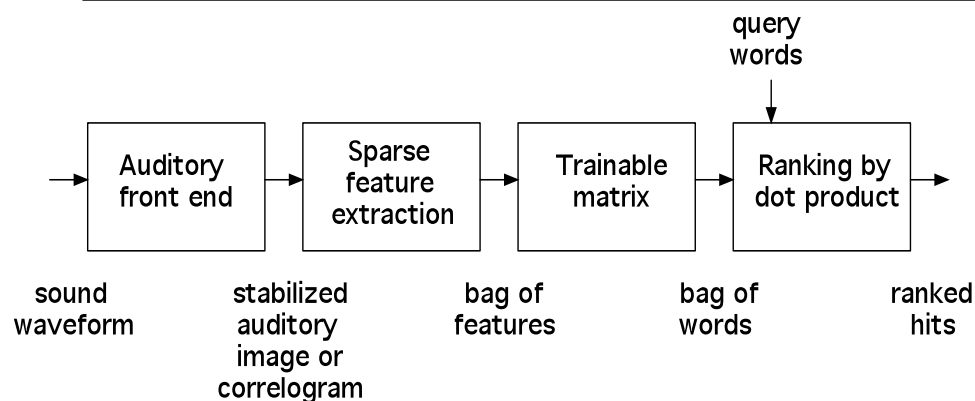


700 x 350 - 44k - jpg
[www.e-architect.co.uk](#)
[Similar images](#)



415 x 332 - 29k - jpg
[viamichelin.com](#)

Content-based retrieval of sound

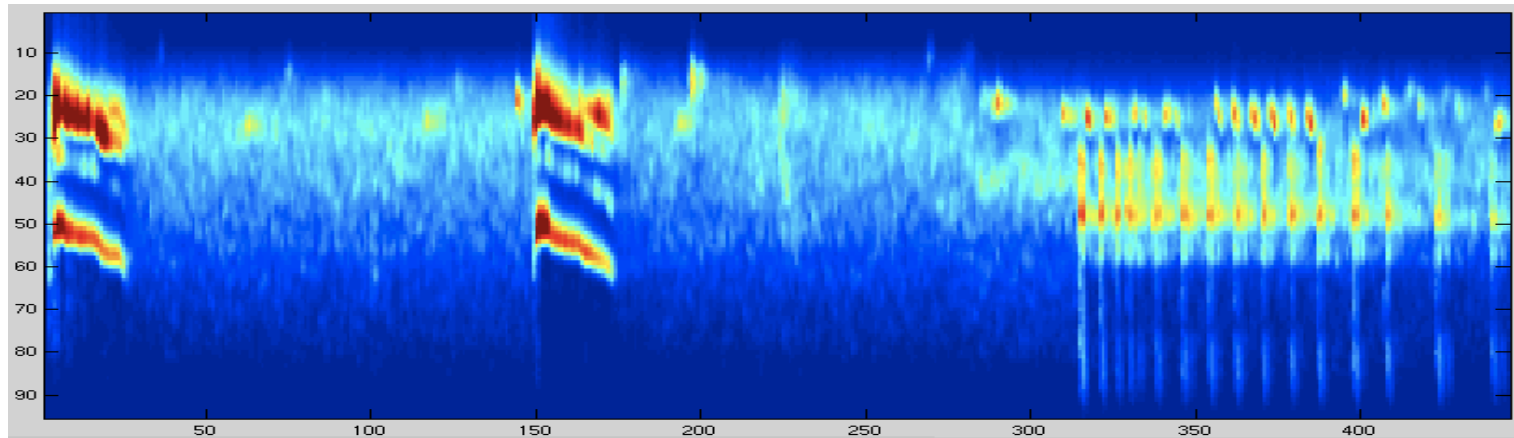
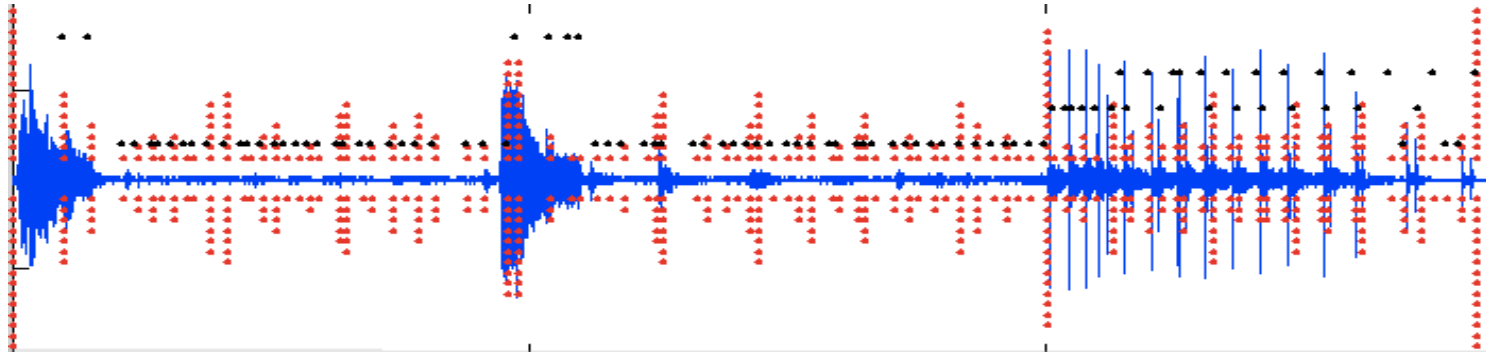


Auditory front end based on stable models from long ago, with new feature extraction ideas.

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

What about sound segmentation or separation?

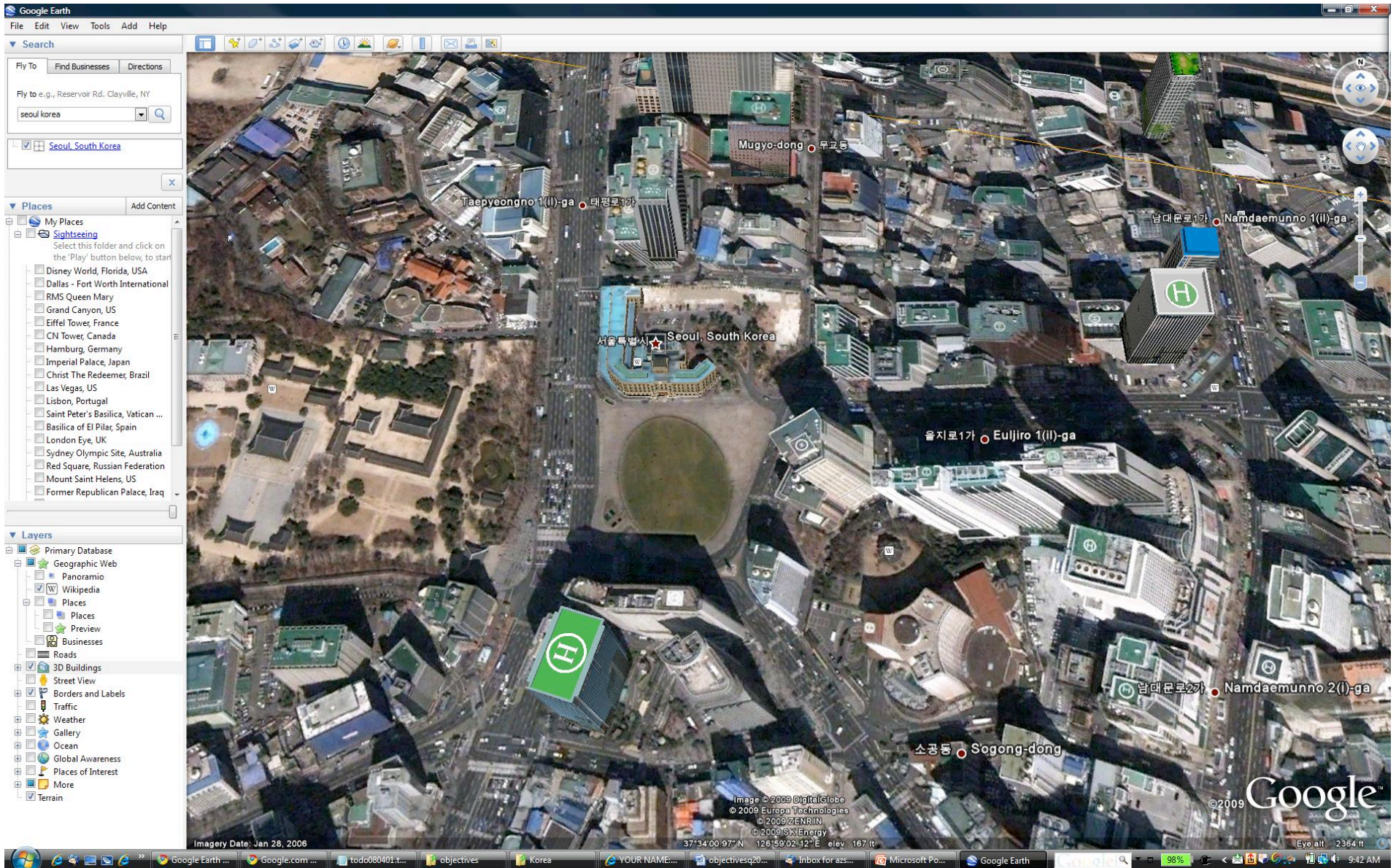
“Sapsucker” (woodpecker) representations



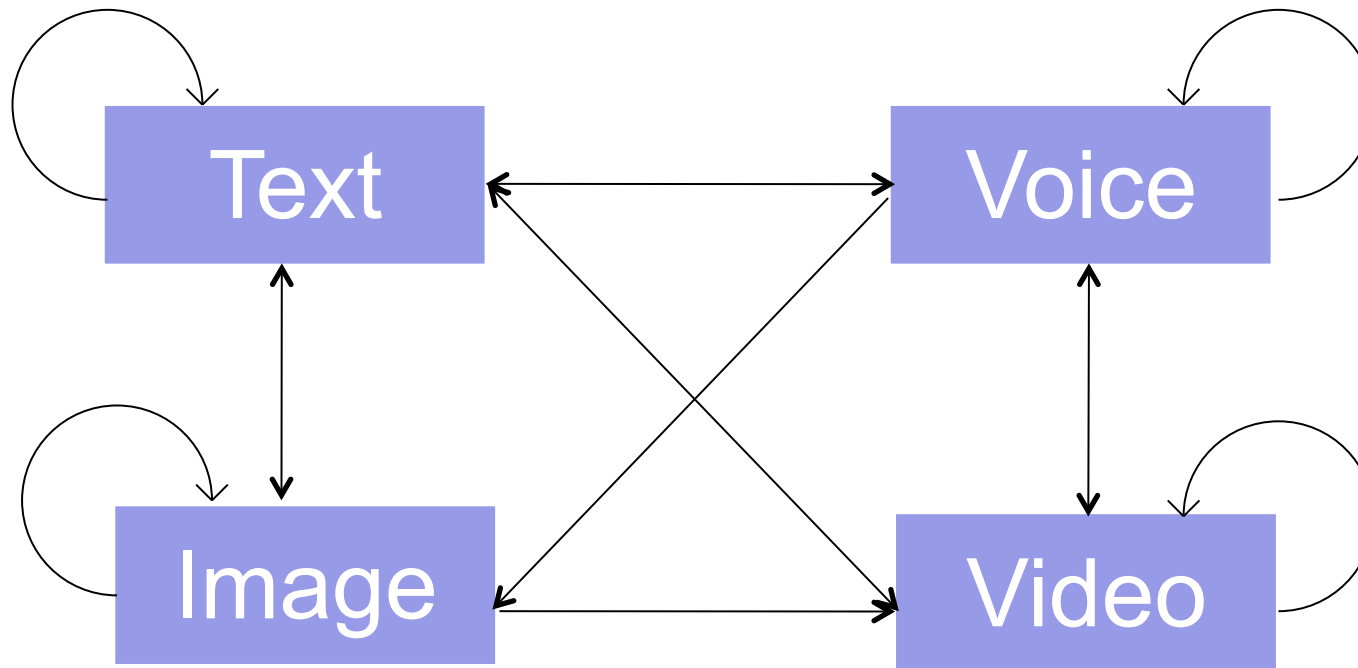
Newer Modalities

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

Maps/Earth as a Modality



Totally Transparent Processing In-Process...



*Many New
Modalities Here or
Coming*

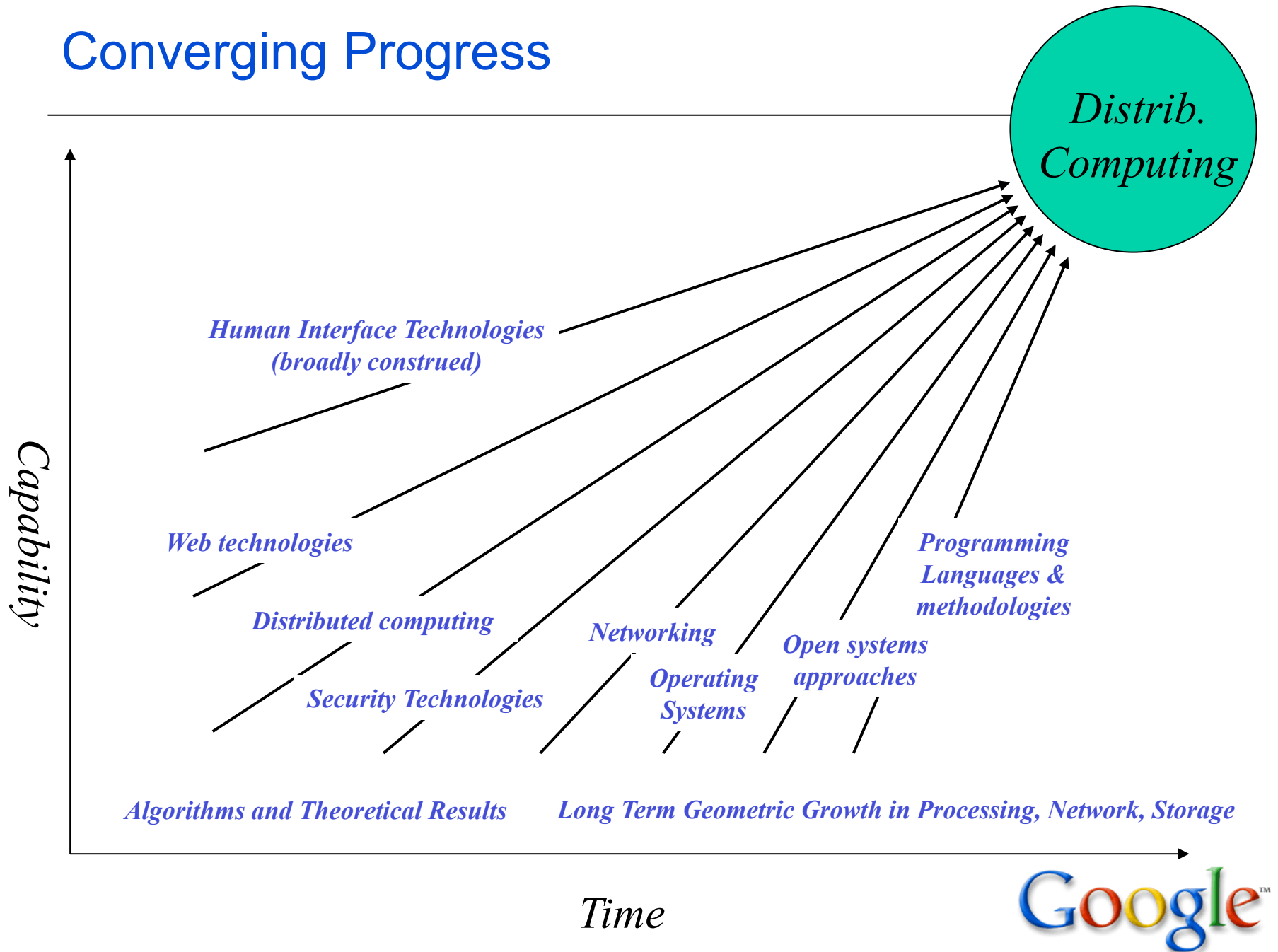
Ideal Distributed Computing



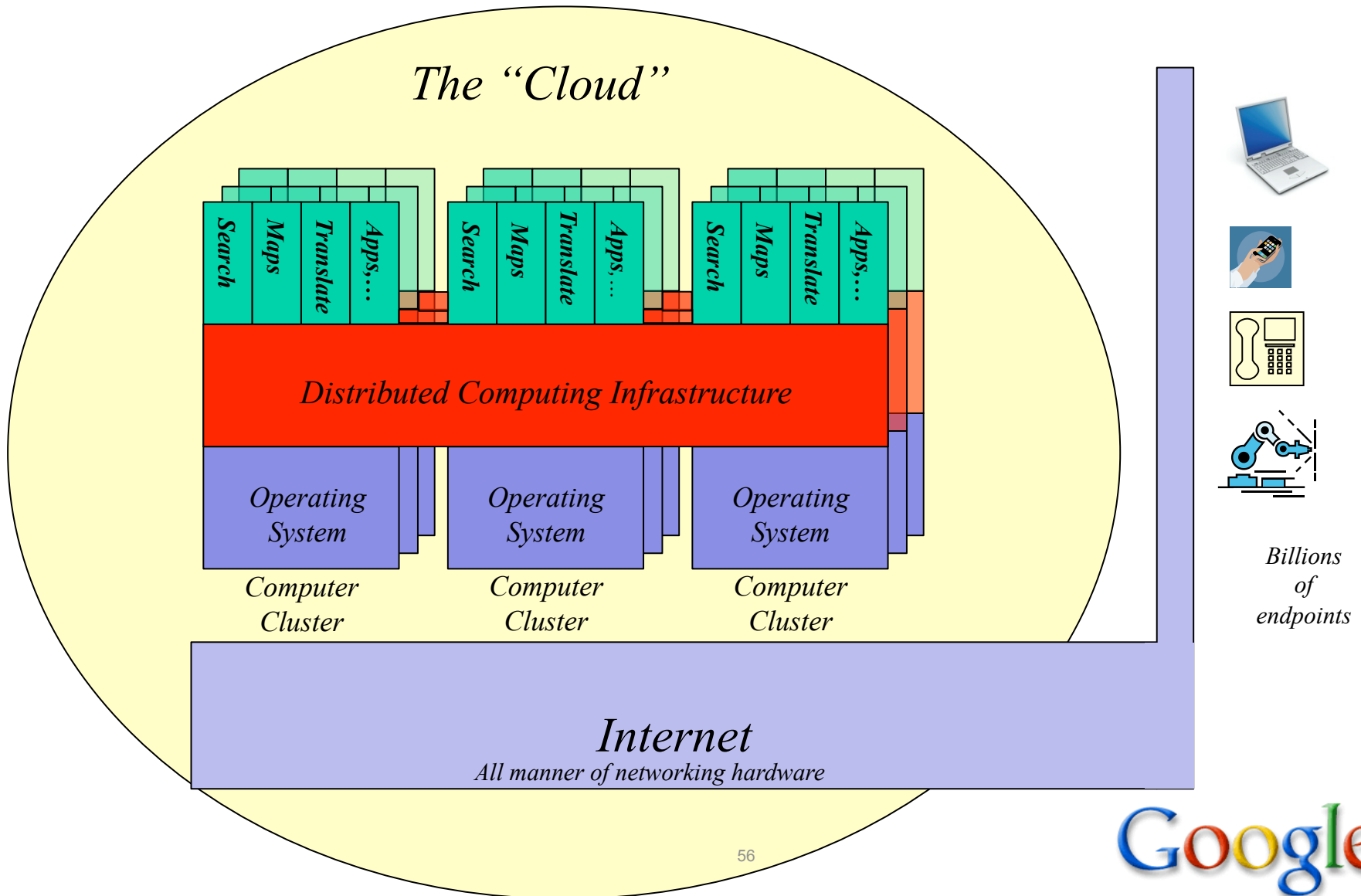
Distributed Computing is 30 years old. But, not very deeply understood until recently

- 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

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

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



Fusion Tables Example (1)

The screenshot shows the Google Fusion Tables interface in a Mozilla Firefox browser. The page title is "Water Usage" by Peter Gleick. A red arrow points to the title. The interface includes a menu bar (File, View, Edit, Visualize, Merge) and a toolbar with options like "All", "filter/aggregate/choose columns", and "Share". The data table below lists water usage metrics for 15 countries from 1993 to 2001. The table has 12 columns: Country, Year, Total Fresh, Per-capita W, Domestic use, Industrial Use, Agricultural u, Domestic u, Industrial U, Agricultural t, Sourc, and 2005 Populat. Each row includes a trash icon for deletion and a comment icon.

| Country | Year | Total Fresh | Per-capita W | Domestic use | Industrial Use | Agricultural u | Domestic u | Industrial U | Agricultural t | Sourc | 2005 Populat |
|---|------|-------------|--------------|--------------|----------------|----------------|------------|--------------|----------------|-------|--------------|
| Algeria | 2000 | 6.07 | 184.78 | 21.98 | 13.19 | 64.83 | 40.61 | 24.38 | 119.79 | | 32.85 |
| Angola | 2000 | 0.35 | 21.96 | 23 | 17 | 60 | 5.05 | 3.73 | 13.17 | | 15.94 |
| Benin | 2001 | 0.13 | 15.4 | 32 | 23 | 45 | 4.93 | 3.54 | 6.93 | | 8.44 |
| Botswana | 2000 | 0.19 | 107.34 | 41 | 18 | 41 | 44.01 | 19.32 | 44.01 | | 1.77 |
| Burkina Faso | 2000 | 0.8 | 60.47 | 13 | 1 | 86 | 7.86 | 0.6 | 52 | | 13.23 |
| Burundi | 2000 | 0.29 | 38.41 | 16.91 | 6 | 77 | 6.5 | 2.3 | 29.58 | | 7.55 |
| Cameroon | 2000 | 0.99 | 60.66 | 18.11 | 8.05 | 73.84 | 10.99 | 4.88 | 44.79 | | 16.32 |
| Cape Verde | 2000 | 0.02 | 39.22 | 7 | 2 | 91 | 2.75 | 0.78 | 35.69 | | 0.51 |
| Central African Republic | 2000 | 0.03 | 7.43 | 80 | 16 | 4.49 | 5.94 | 1.19 | 0.34 | | 4.04 |
| Chad | 2000 | 0.23 | 23.59 | 17 | 0 | 83 | 4.01 | 0 | 19.58 | | 9.75 |
| Comoros | 1993 | 0.01 | 12.5 | 48 | 5 | 47 | 6 | 0.63 | 5.88 | | 0.8 |
| Congo, Democratic Republic (formerly Zaire) | 2000 | 0.36 | 6.26 | 53 | 17 | 31.47 | 3.32 | 1.06 | 1.97 | | 57.55 |
| Congo, Republic of | 2000 | 0.03 | 7.5 | 59.27 | 29 | 12 | 4.45 | 2.18 | 0.9 | | 4 |
| Cote D'Ivoire | 2000 | 0.93 | 51.24 | 24 | 11.82 | 64.91 | 12.3 | 6.06 | 33.26 | | 18.15 |
| Djibouti | 2000 | 0.02 | 25.32 | 84 | 0 | 16 | 21.27 | 0 | 4.05 | | 0.79 |
| Egypt | 2000 | 68.3 | 922.6 | 7.62 | 6 | 86 | 70.27 | 55.36 | 793.44 | | 74.03 |

Fusion Tables Example (2)

Google Fusion Tables (Pre-Alpha) - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://dslis0.tableserv.fusiontables.vb.borg.google.com:26201/DataSource?dsrclid=1401/1401

Home jayant@gmail.com | Feedback | Sign out

Water Usage Peter Gleick

Share

File View Edit Visualize Merge

Current view: All - Show options (filter/aggregate/choose columns) 1 - 100 of 172 Next

| Country | Year | Total Fresh | Cultural i | Sourc | 2005 Populat | | |
|---|------|-------------|------------|-------|--------------|-------|----------------|
| Algeria | 2000 | 6.07 | 19.79 | | 32.85 | | |
| Angola | 2000 | 0.35 | 13.17 | | 15.94 | | |
| Benin | 2001 | 0.13 | 6.93 | | 8.44 | | |
| Botswana | 2000 | 0.19 | 44.01 | | 1.77 | | |
| Burkina Faso | 2000 | 0.8 | 52 | | 13.23 | | |
| Burundi | 2000 | 0.29 | 29.58 | | 7.55 | | |
| Cameroon | 2000 | 0.99 | 44.79 | | 16.32 | | |
| Cape Verde | 2000 | 0.02 | 35.69 | | 0.51 | | |
| Central African Republic | 2000 | 0.03 | 0.34 | | 4.04 | | |
| Chad | 2000 | 0.23 | 19.58 | | 9.75 | | |
| Comoros | 1999 | 0.01 | 5.88 | | 0.8 | | |
| Congo, Democratic Republic (formerly Zaire) | 2000 | 0.36 | 6.26 | 53 | 17 | 31.47 | 3.32 1.06 |
| Congo, Republic of | 2000 | 0.03 | 7.5 | 59.27 | 29 | 12 | 4.45 2.18 |
| Cote D'Ivoire | 2000 | 0.93 | 51.24 | 24 | 11.82 | 64.91 | 12.3 6.06 |
| Djibouti | 2000 | 0.02 | 25.32 | 84 | 0 | 16 | 21.27 0 |
| Egypt | 2000 | 68.3 | 922.6 | 7.62 | 6 | 86 | 70.27 55.36 |
| | | | | | | | 793.44 74.03 |

Share this table

Invited people

- As viewers - can see and comment on the data
- As collaborators - can also edit the data
- As owners - can also invite people to view or collaborate

alohalevy@gmail.com

Separate email addresses with commas.

Send email invitations

Anyone may view this table

Viewers (1)
hagonza@gmail.com

Collaborators (0)

Owners (1)
jayant@gmail.com

Invite these people

Transferring data from dslis0.tableserv.fusiontables.vb.borg.google.com...

show debug

Open Notebook

Fusion Tables Example (3)

The screenshot shows the Google Fusion Tables interface in a Mozilla Firefox browser. The page title is "Water Usage" by Peter Gleick. A filter is applied to the "Industrial Use (%)" column, showing values greater than 50. The table displays 21 rows of data, with the first 12 rows visible. The filter is circled in blue.

Water Usage Peter Gleick

File View Edit Visualize Merge

Filter Aggregate Choose columns

Industrial Use (%) > 50

Add condition ?

Apply Clear filter

Current view: [Industrial Use (%) > 50] - Hide options

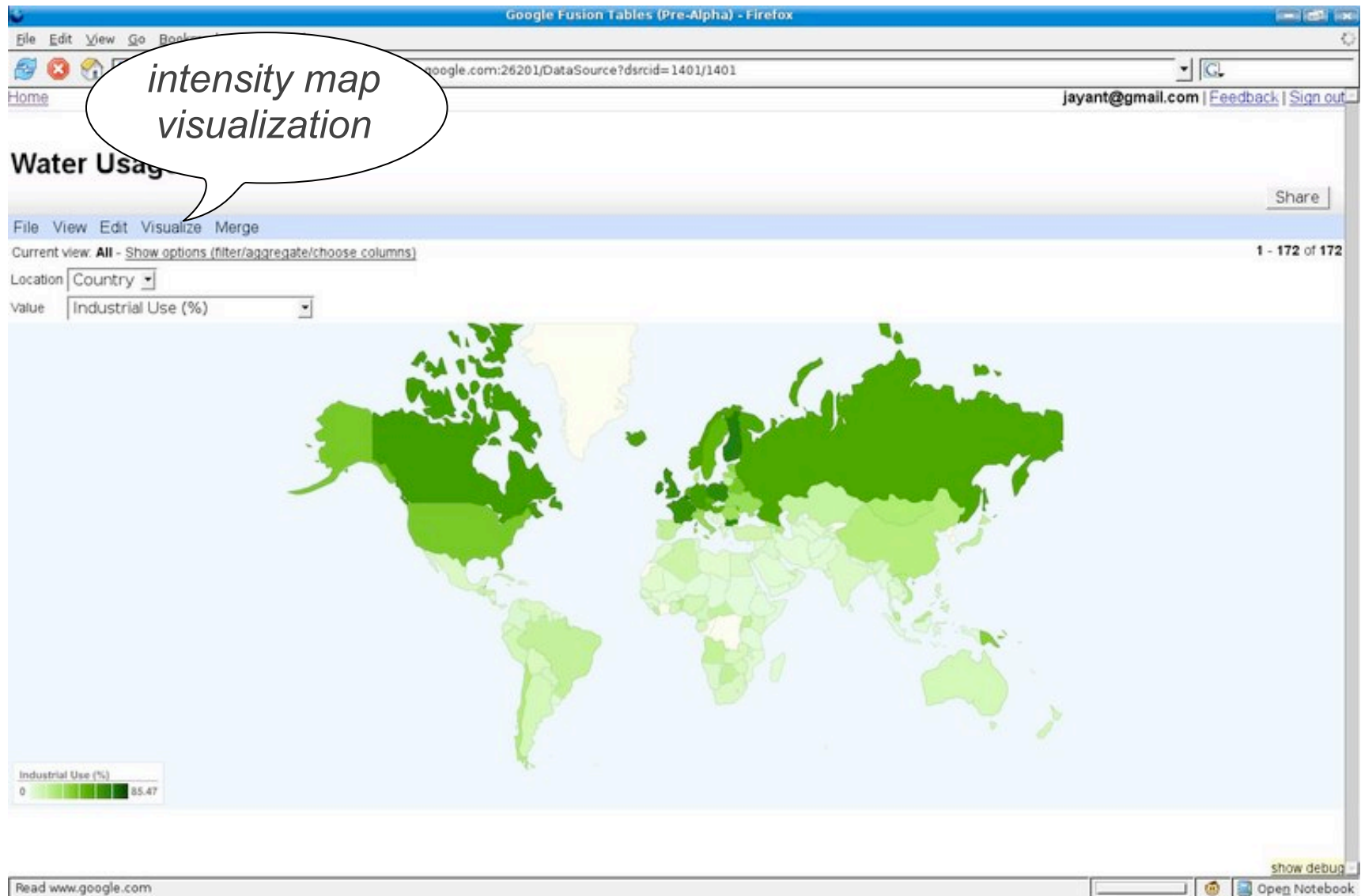
1 - 21 of 21

| Country | Year | Total Fresh | Per-capita Wa | Domestic u | Industrial U | Agricultural | Domestic us | Industrial Use | Agricultural t | 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 | 196.89 | 40.76 | 57.1 | 2.14 | 76.18 | 106.71 | 4.01 | m | 10.22 | | |
| Finland | 1999 | 2.33 | 443.81 | 13.65 | 83.89 | 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 | | |

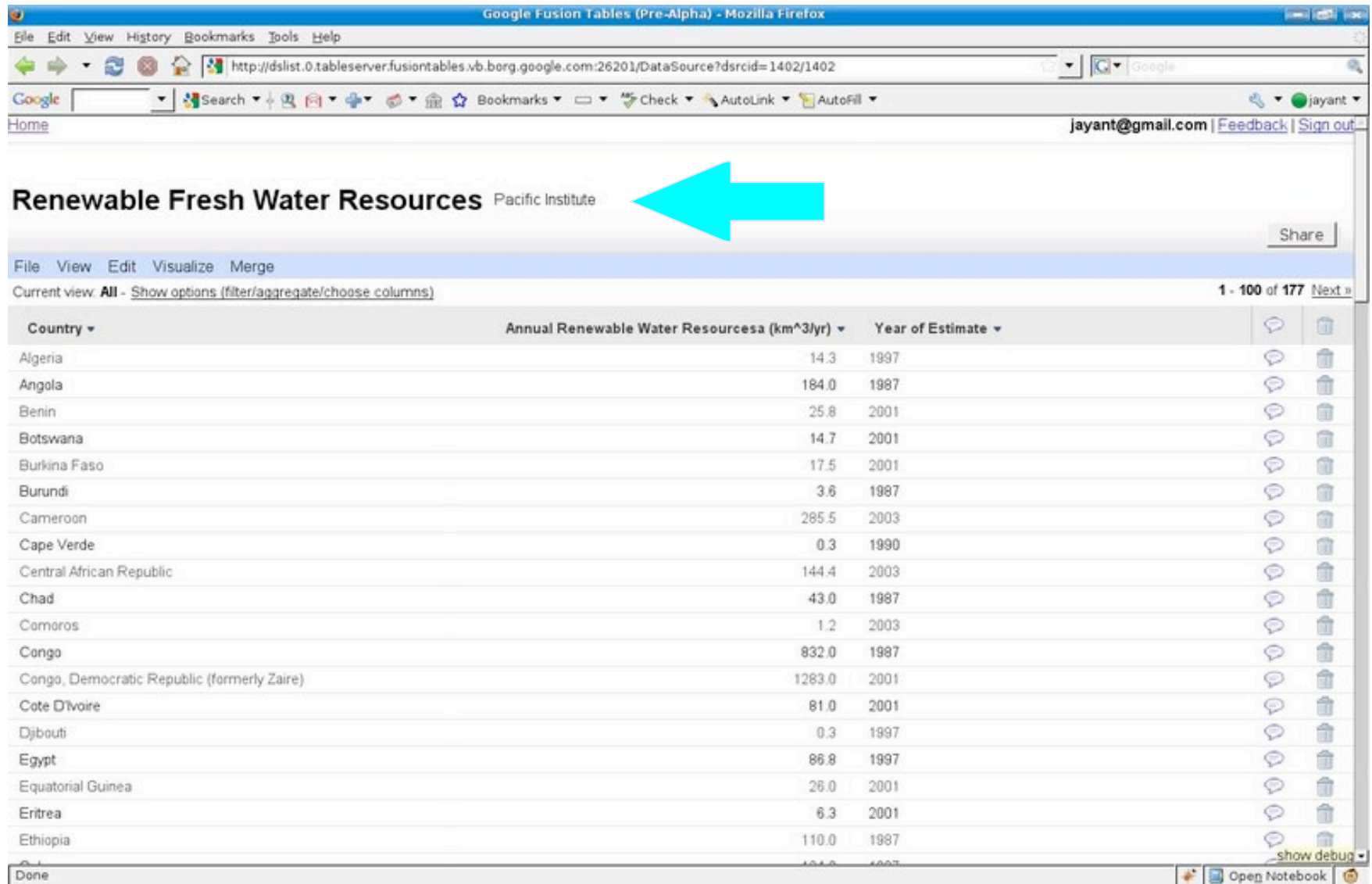
Done

Open Notebook

Fusion Tables Example (4)



Fusion Tables Example (5)



The screenshot shows the Google Fusion Tables interface in a Mozilla Firefox browser. The page title is "Renewable Fresh Water Resources" by Pacific Institute. A red arrow points to this title. Below the title is a menu with options: File, View, Edit, Visualize, Merge. The current view is set to "All" with a link to "Show options (filter/aggregate/choose columns)". The table displays data for 177 countries, showing the Annual Renewable Water Resources in km³/yr and the Year of Estimate. The table is sorted by Annual Renewable Water Resources in descending order.

| Country | Annual Renewable Water Resources (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)

The screenshot shows the Google Fusion Tables Merge interface. The browser title is "Google Fusion Tables (Pre-Alpha) - Mozilla Firefox". The URL is "http://dslist.0.tableserver.fusiontables.vb.borg.google.com:26201/DataSource?dsrclid=1402/1402". The user is logged in as "jayant@gmail.com".

The interface is titled "Renewable Fresh Water Resources" and "Water Usage". It has a menu bar with "File", "View", "Edit", "Visualize", and "Merge".

Step 1: "Choose which column to use for matching data across the two tables". The "Renewable Fresh Water Resources" table has three options: "Country" (selected), "Annual Renewable Water Resourcesa (km³/yr)", and "Year of Estimate". A cyan arrow points to the "Country" option.

Step 2: "Merge with". The "Water Usage" table has several options: "Country" (selected), "Year", "Total Fresh Water Withdrawal", "Per-capita Waterwithdrawal", "Domestic use %", "Industrial Use (%)", "Agricultural use (%)", "Domestic use", "Industrial U", "Agricultural", and "Source". A cyan arrow points to the "Country" option. A speech bubble says "select key".

Below the tables, there are checkboxes for "Select columns" and a text input field "Save as a new table named" with the value "Water Resources and Usage". A speech bubble says "create a new merged table".

At the bottom, there are buttons for "Merge tables" and "Cancel". The status bar shows "Current view: All - Show options (filter/aggregate/choose columns)" and "1 - 100 of 177 Next a show debug".

Fusion Tables Example (8)

Google Fusion Tables (Pre-Alpha) - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://dslist.0.tableserver.fusiontables.vb.borg.google.com:26201/DataSource?dsrclid=1404/1404

Google Search

Home jayant@gmail.com | Feedback | Sign out

Water Resources and Usage Pacific Institute, Peter Gleick

Share

File View Edit Visualize Merge

Current view: All - Show options (filter/aggregate/choose columns) 1 - 100 of 177 Next

| Country | Annual Renewal | Year of Est. | Year | Total Fresh | 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 | 184.0 | | | | 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 |

Cell value: 184.0

jayant(17 seconds ago)
Angola has a lot of water for a country its size!

And I thought it was in the Kalahari!

Save comment | Close | Refresh

Done Open Notebook

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



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.

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

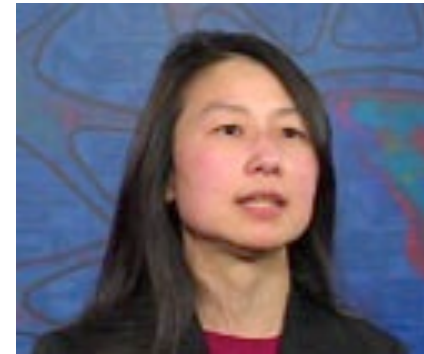
Technology Leadership: NSF CLuE Program



National Science Foundation
WHERE DISCOVERIES BEGIN

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.**



Jeannette Wing

Google Relationship with Academe (cont.)

- Check out growing sites:
 - [//research.google.com](http://research.google.com)
 - [//research.google.com/university](http://research.google.com/university)
 - [//code.google.com](http://code.google.com)
 - [//code.google.com/edu](http://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

Thank you!

