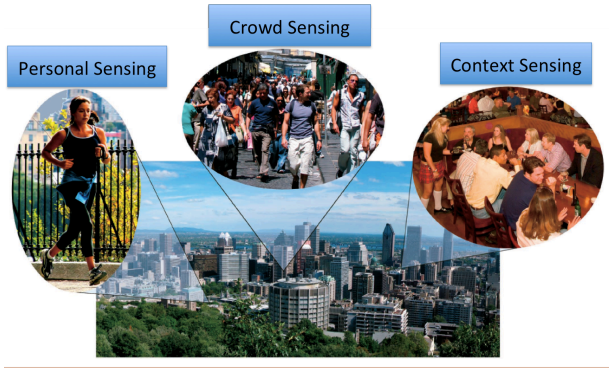
Social Sensing and Cyber-Physical Systems



http://www3.nd.edu/~dwang5/courses/spring15/

CSE 40437/60437-Spring 2015

Prof. Dong Wang

Logistics

- **Instructor**

  Dong Wang, Office: 214 B Cushing, Tel: 631-3749

  Office Hours: Tue, 3:15-5:15 PM

  Email:dwang5@nd.edu

- **TA**

  Chao Huang, Office: 254 Fitzpatrick Hall

  Office Hours: Mon: 4-6 PM; Thu: 4-6 PM.

  Email: chuang7@nd.edu

A Little About Me

- Ph.D. in *Quantifying the QoI in Social Sensing*, Department of Computer Science, University of Illinois at Urbana-Champaign, 2012.
- 2014-present: Assistant Professor, Department of Computer Science and Engineering, University of Notre Dame
- Research Interests: Big Data Analytics, Social Sensing and Cyber-Physical Systems, Trust Analysis in Information and Social Networks
- Homepage: http://www3.nd.edu/~dwang5/

Where and When

- **Lecture Times and Location**

  Tuesday and Thursday, 2:00 - 3:15pm

  DeBartolo Hall 117

Honor Code

- The graded work you do in this class must be your own. In the case where you collaborate with other students make sure to fairly attribute their contribution to your project.

Prerequisite

- Senior or Graduate Standing
- Proficiency in Python Programming
- Experiences with Big Data Analytics (e.g., Data Mining) and Processing (e.g., Cloud Computing) will be a PLUS!
- Experiences with Real-time, Mobile Sensing, Online Social Media will be a PLUS!

Grading

- Participation: 10%
  - Assigned for individuals' class participation, and discussion of papers (*individual based*).
- Assignments: 30%
  - Three assignments, 10% for each (*individual based*).
- Presentation: 10%
  - Assigned for an in-class presentation (*group based*).
- Project: 50%
  - Determined by a group course project (*group based*).

Grading

- Project: 50%
  - Updates and Discussion: 5%
  - Proposal: 5%
  - Mid-term presentation: 5%
  - Mid-term report: 10%
  - Final presentation: 10%
  - Final project paper: 15%

Assignments

- There will be *three* assignments, each accounts for 10% of the final grade.
- They are centered around building a simplified data crawling and analysis engine based on Twitter.
- Most of the coding can be done in Python.
- The assignments are individual based. You can discuss the problems with others but you are expected to **finish the assignment by yourself**.

In-class Paper Presentation

- Form a group of up to three people for a in-class presentation of a selected technical paper (the group can be the same as your project group).
- Exercise your critical thinking ability, understand how to present and challenge other's work, and learn how to lead technical discussions.
- It contains a presentation part and Q&A part (about 25 min).
- It is scheduled in the second half of the semester. Detailed instructions on preparation and sign-up will be announced later.

Collaborative Projects
(up to 3 people per group)

- It is a group based collaborative project. You can form your group of *up to three people* in the first couple of weeks. Feel free to use Piazza to find your partner.
- The only *restriction* is the the project should use the ***online social media data*** (e.g., Twitter, Flickr, Foursquare, Facebook, etc.) to solve some interesting problem in social sensing.
- Some example ideas are on the course website, and you are always **encouraged to come up with your own ideas!**
- http://www3.nd.edu/~dwang5/courses/spring15/project/project.html
- However, if you do have a cool idea that does not meet the above requirement, please schedule a meeting with the instructor to discuss about it.

Collaborative Projects (**Milestones**)

- Project title, abstract, and member list is due **Jan 30th**.
- Meetings to discuss progress and problems on your project of interest.
- 2-page project proposal by **Feb 20th**.
- Mid-term project presentation, **March 17th-19th**.
- 4-page mid-term project report due: **March 20th**
- Final project presentation, starting from the **week of April 21th**
- Final project paper is due **May 4th**. The paper follows the standard technical paper format.
- Successful projects could result in a conference or journal quality research paper.

All project documents (abstract, proposal, report and final paper) are due **at Noon of the due day.**

Getting Help

- **Office Hours** – Schedule on the course website
- **Email** - Contact Prof. Wang for questions about grades, course policies, etc.
- **Piazza Discussion Page** - General announcement, lecture notes and Q&A after class
- **Grades** are available in Sakai.

Readings

- First week, 2 overview papers
- Starting next week
  - Papers are posted on the course website (about 4 papers per week)

Schedule

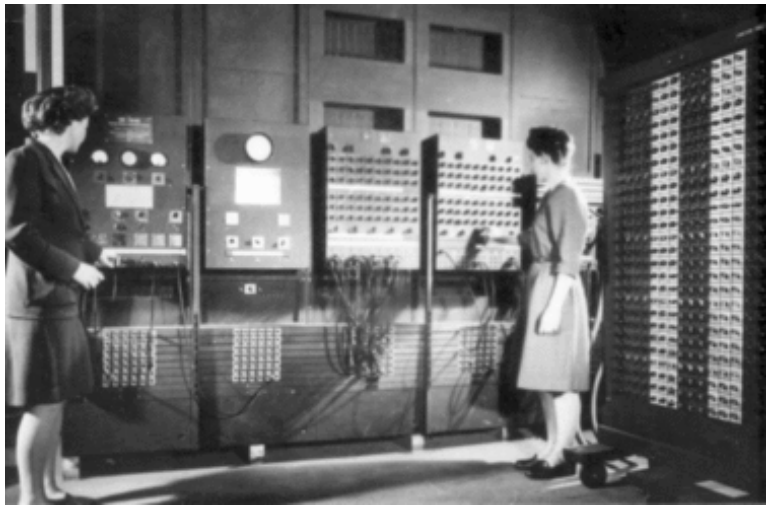
- See Course Website: http://www3.nd.edu/~dwang5/teach/spring15/
- Lectures, Project Meetings, and Presentations
- Lecture notes will be available at: https://piazza.com/nd/spring2015/cse4043760437/resources

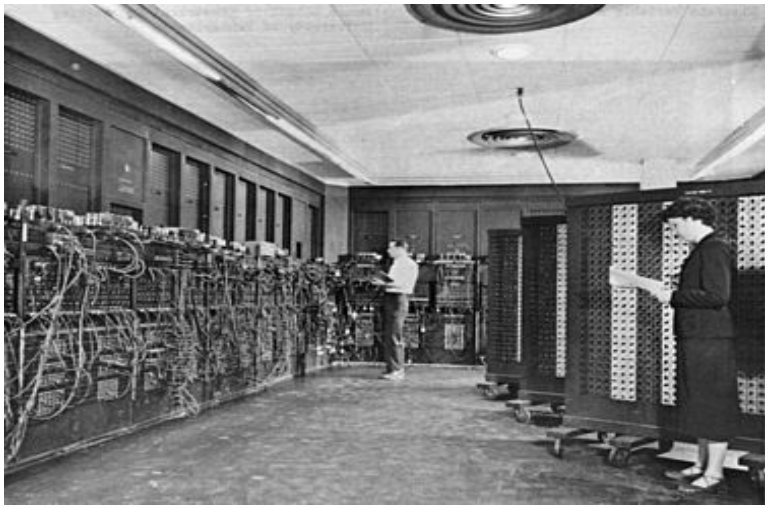
Where is Computer Science Research Going?

The first general purpose computer in the world

ENIAC (Electronic Numerical Integrator And Computer)

Born in Feb. 1946 at University of Pennsylvania

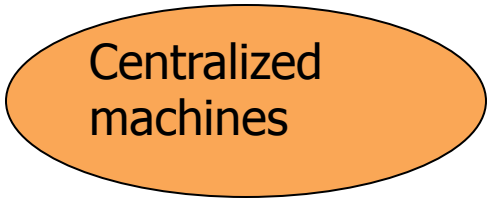


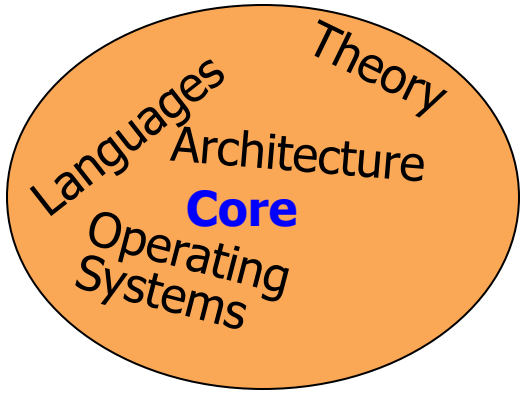


Where is Computer Science Research Going?

1946-1970s:

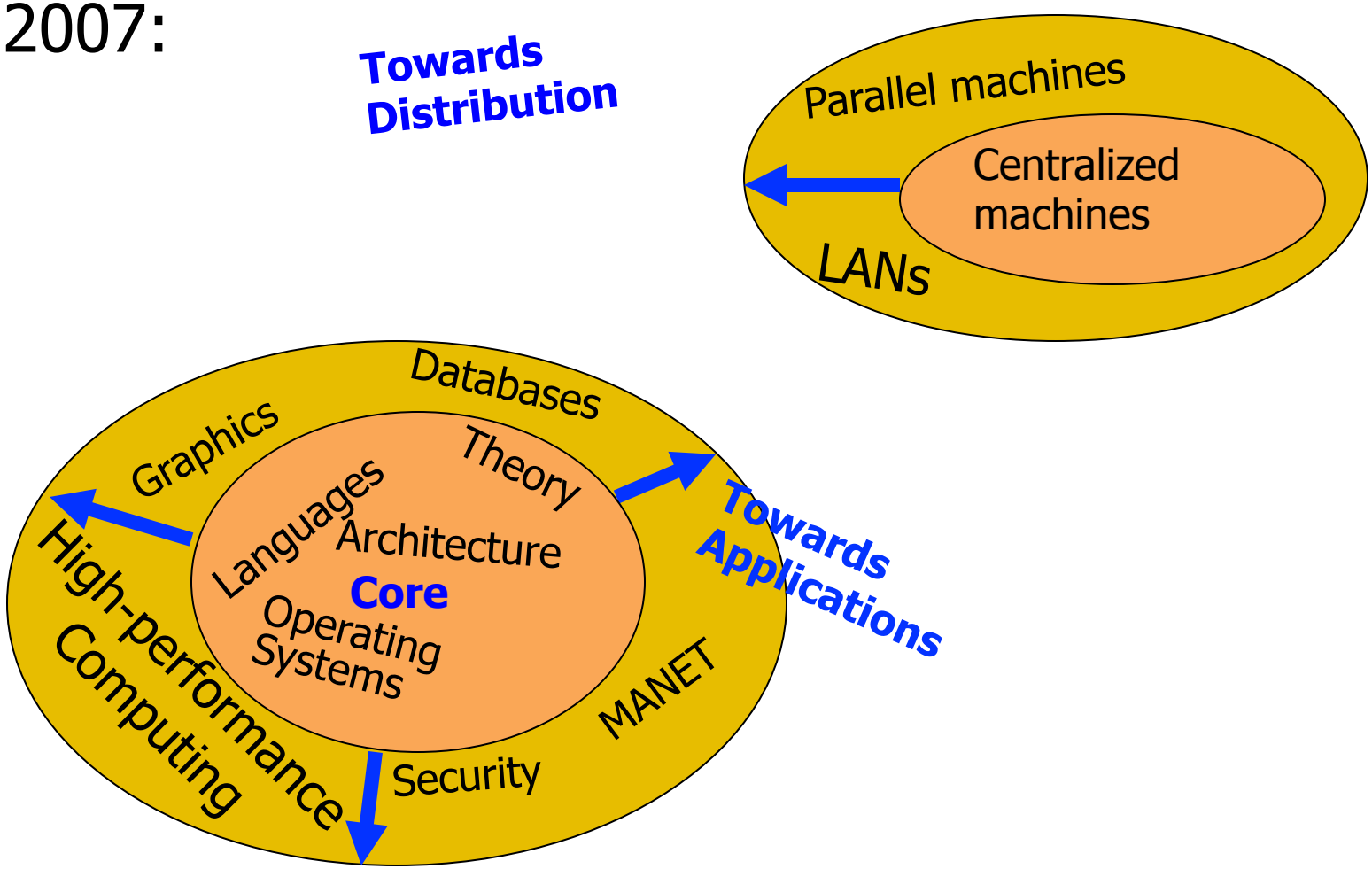
The beginning:





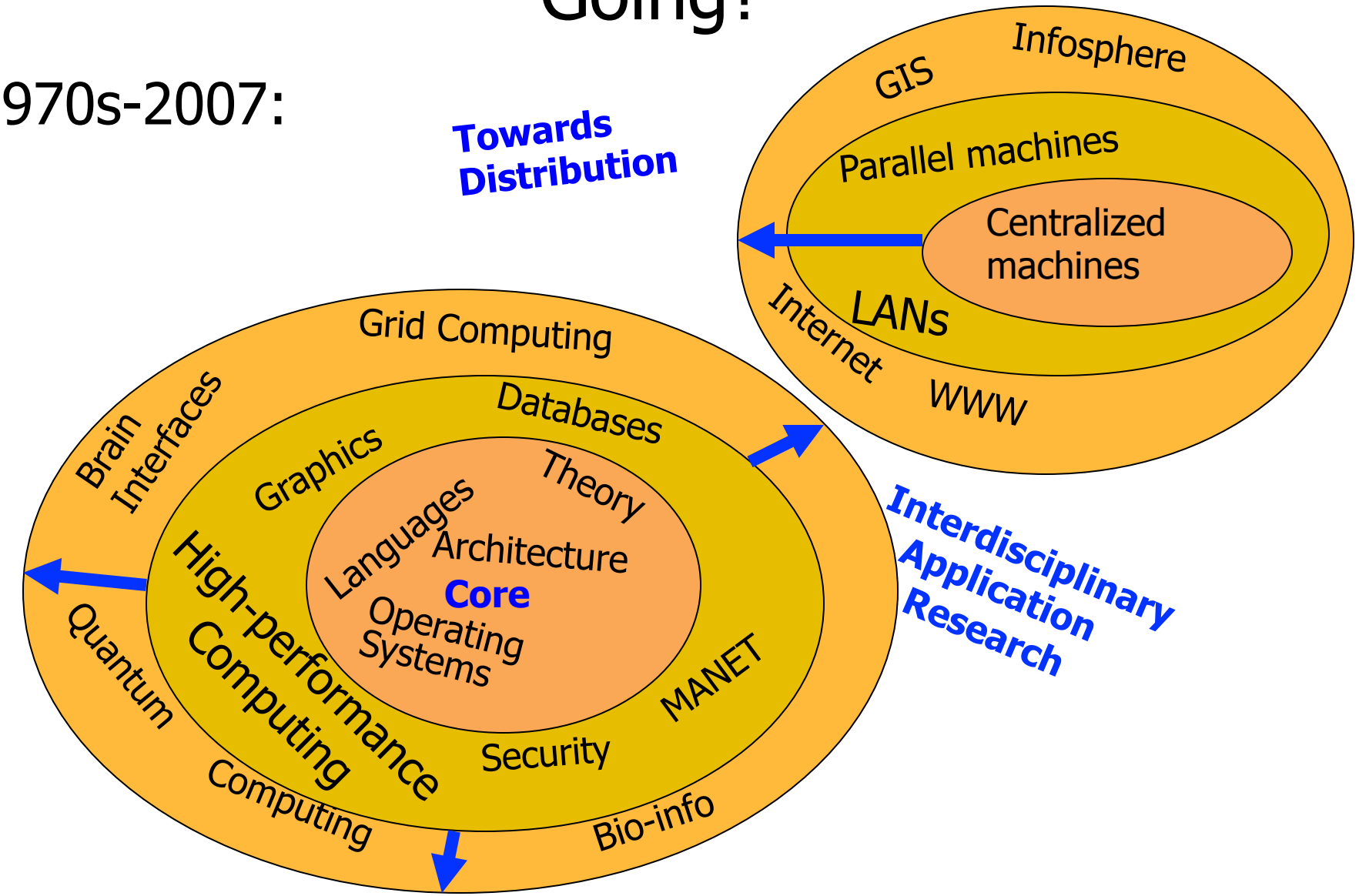
Where is Computer Science Research Going?

1970s-2007:



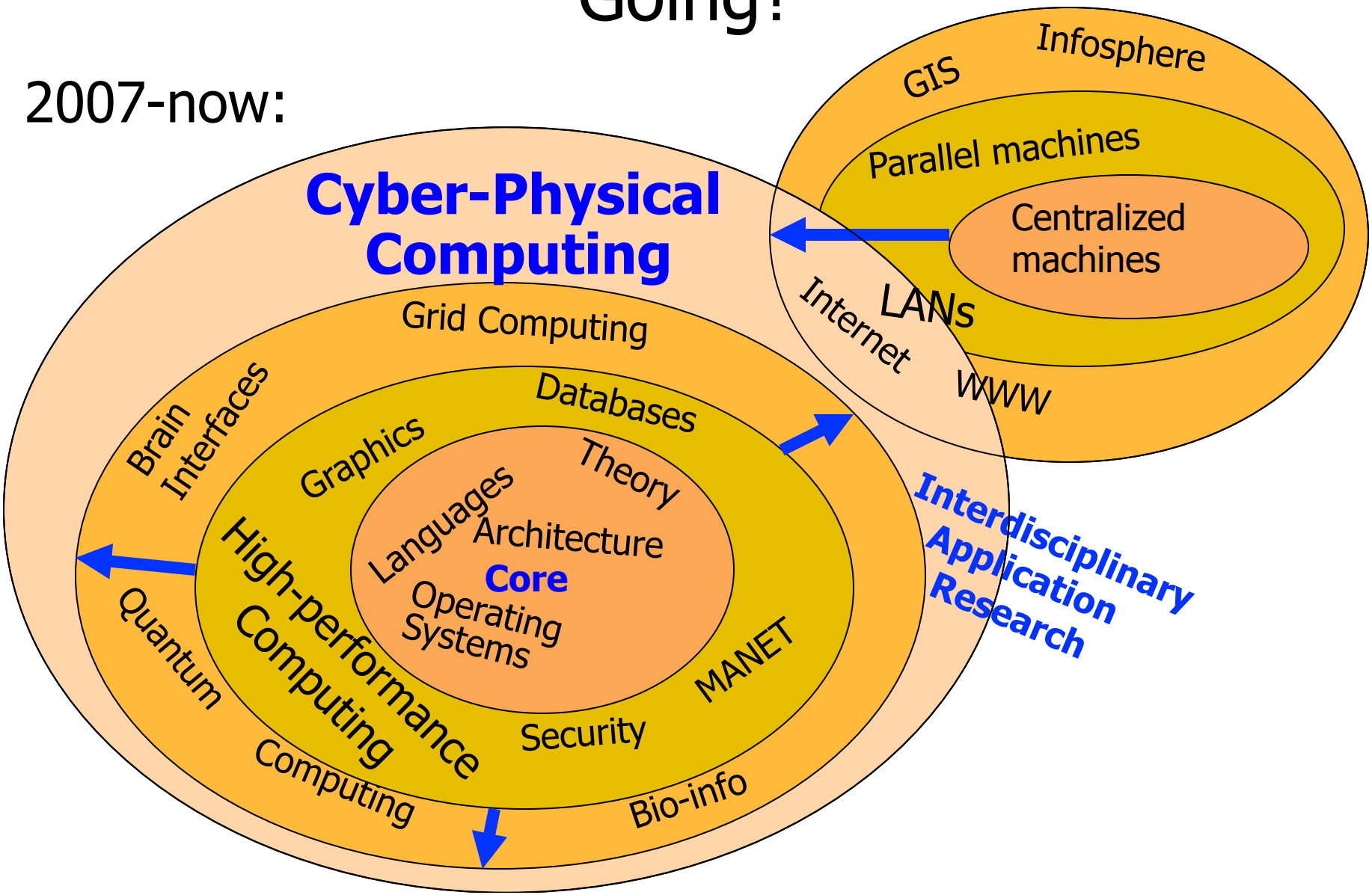
Where is Computer Science Research Going?

1970s-2007:



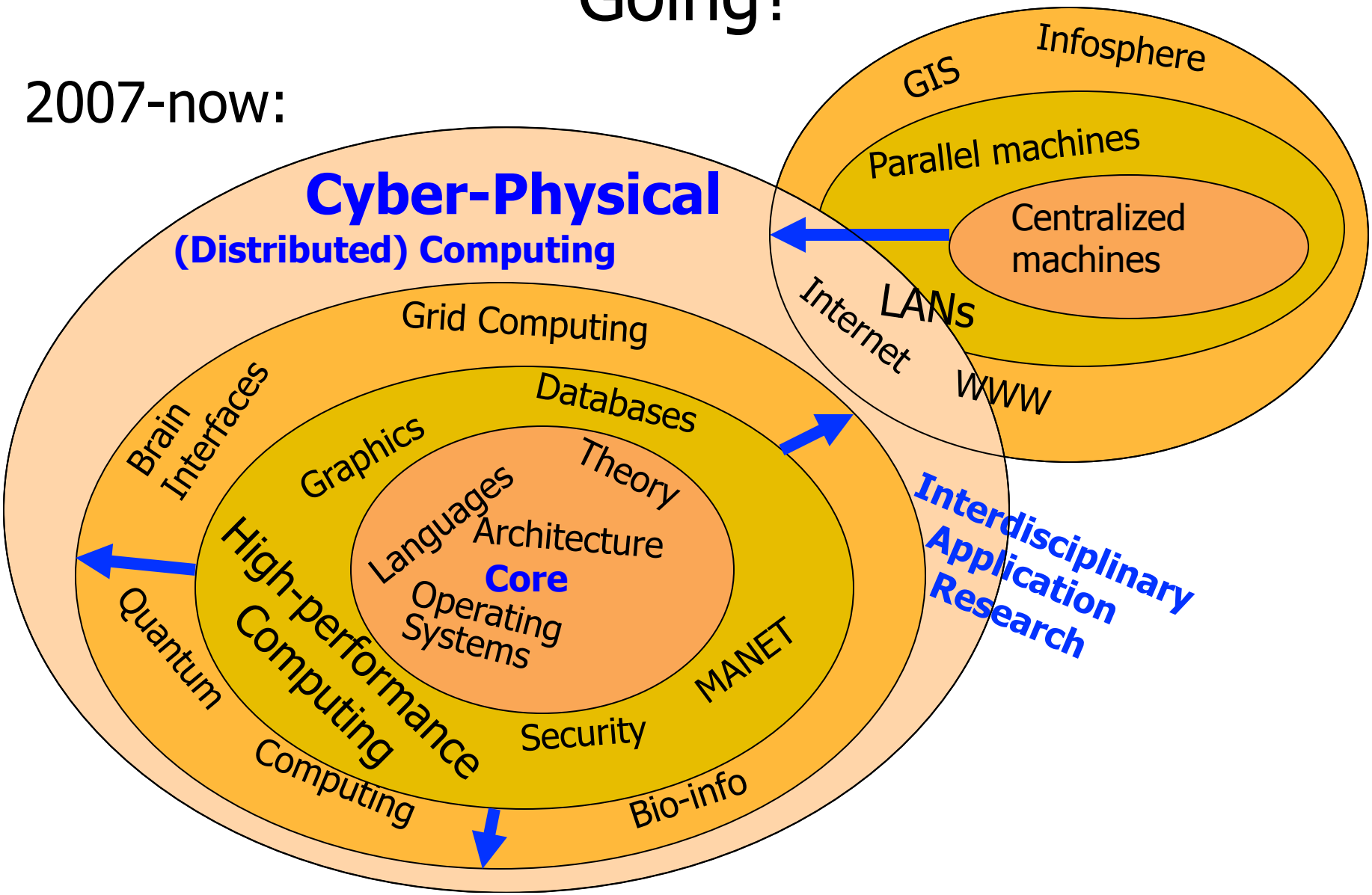
Where is Computer Science Research Going?

2007-now:



Where is Computer Science Research Going?

2007-now:



Where is Computer Science Research Going?

2007-now:

Cyber-Physical Systems

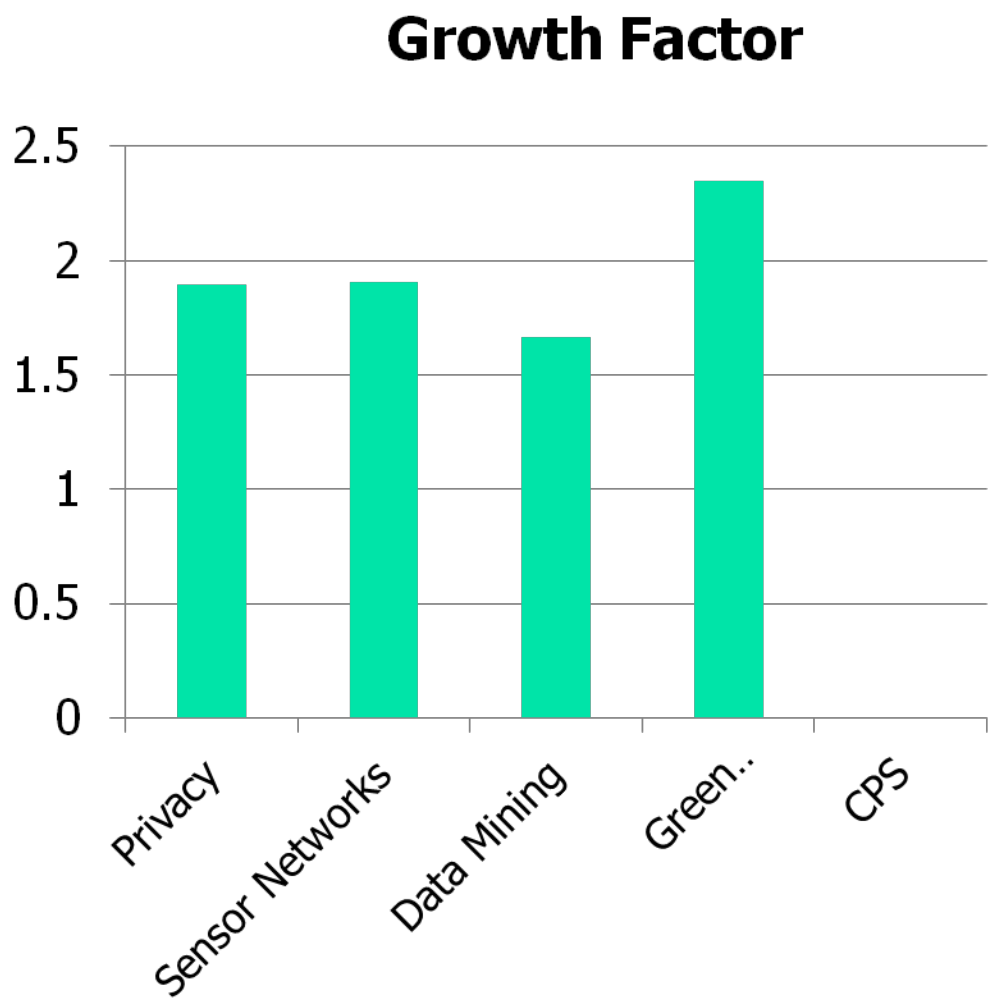
For example:

In the US, in 2007, the **Presidential Counsel of Advisors in Science and Technology** named systems that interact with the physical world the

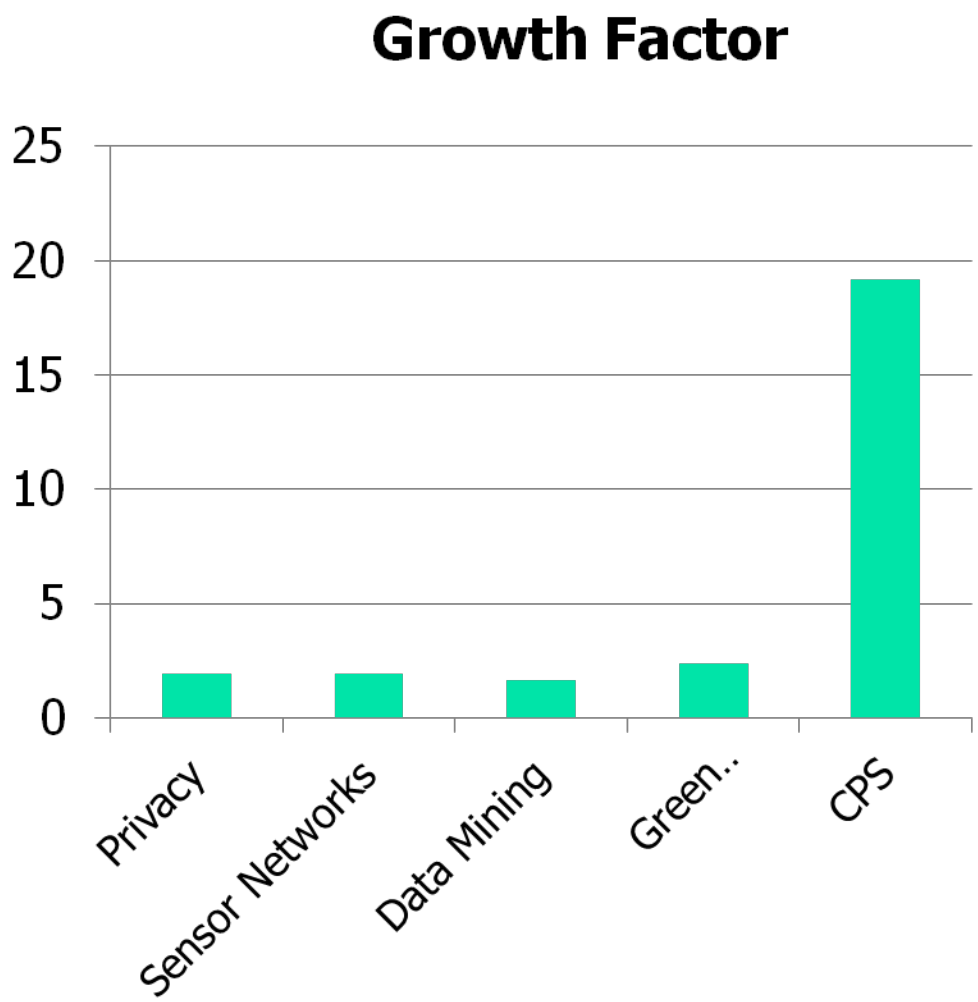
#1 Research Priority in the US

Keyword Trends (on INSPEC) in Computers/Control Eng.

Keyword Trends (Continued): 2011/2006 Multiplicative Factor

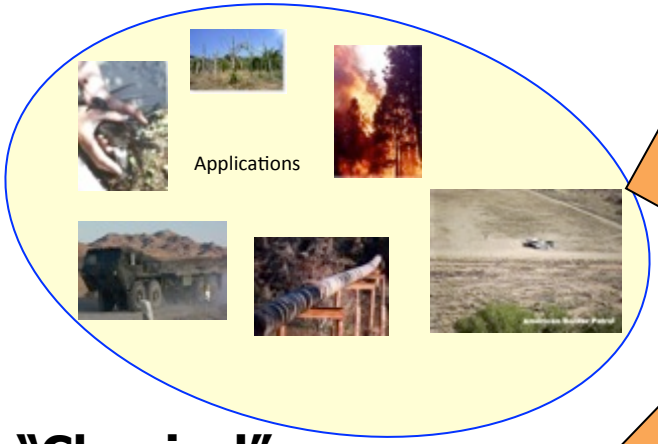


Keyword Trends (Continued): 2011/2006 Multiplicative Factor



Force #1: Moore's Law
(Device proliferation)

Cyber Physical Networks

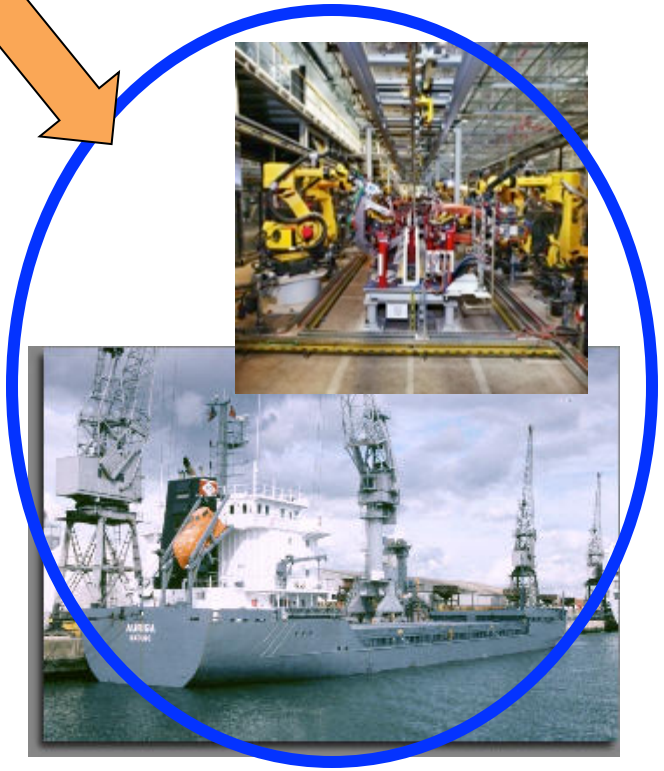


"Classical"
- Unattended multihop ad hoc wireless

Industrial
- Single-hop: monitor cargo, machinery factory floor, …
- Send to base.

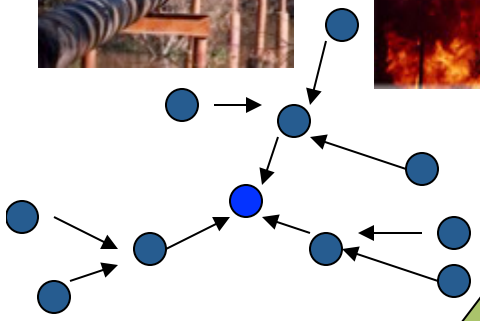
Medical

Ubiquitous Computing



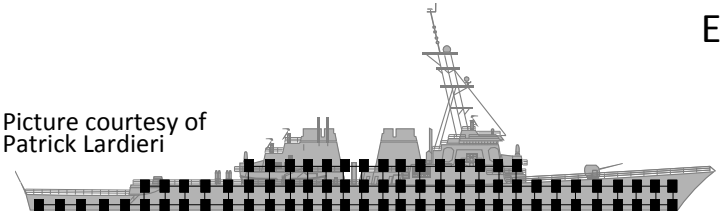
Force #2: Integration at Scale
(Isolation has cost!)

- Low end: ubiquitous embedded devices
  - Large-scale networked embedded systems
  - Seamless integration with a physical environment



World Wide Sensor Web
(Feng Zhao)

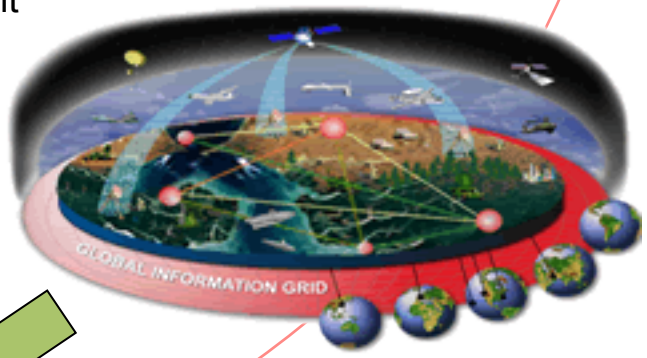
- High end: complex systems with global integration
  - Examples: Global Information Grid, Total Ship Computing Environment



Picture courtesy of Patrick Lardieri

Total Ship Computing Environment (TSCE)

Global Information Grid



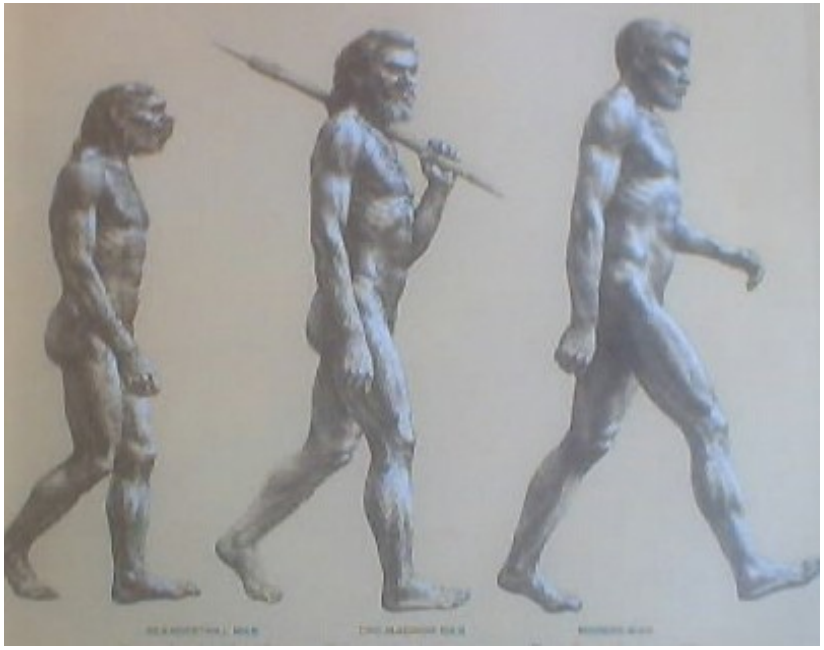
Future Combat System
(Rob Gold)

Integration and Scaling Challenges

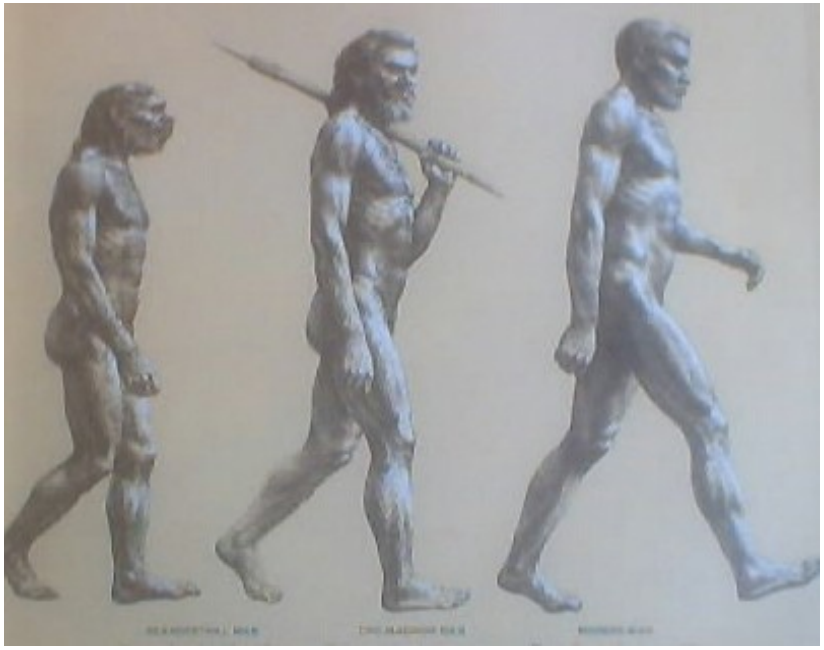
Low End

High End

Force #3: Biological Evolution



Force #3: Biological Evolution

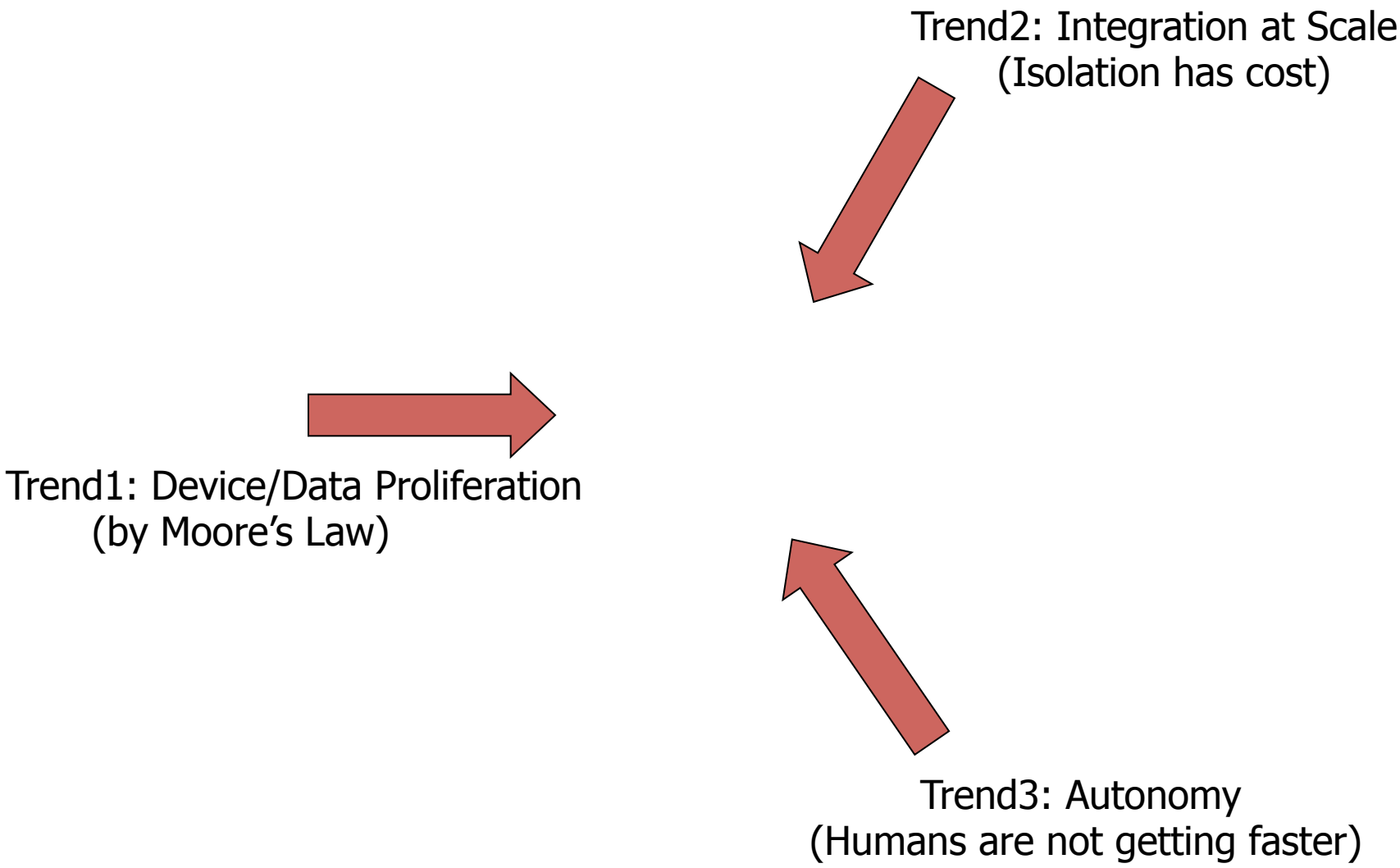


- **It's too slow!**
  - The exponential proliferation of data sources (afforded by Moore's Law) is ***not*** matched by a corresponding increase in human ability to consume information!

  → Increasing focus on information distillation to support decision making

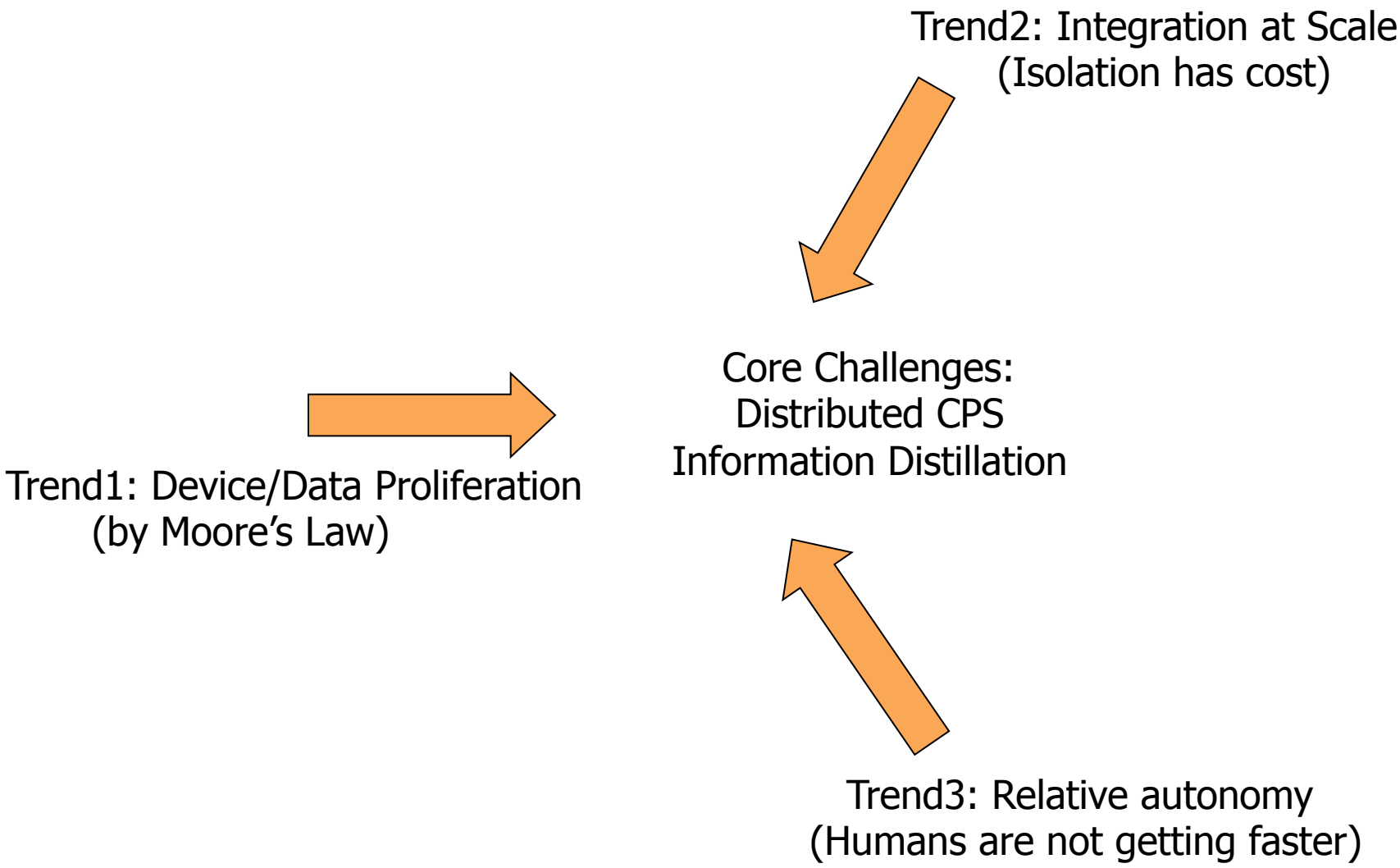
Confluence of Trends

The Overarching Challenge

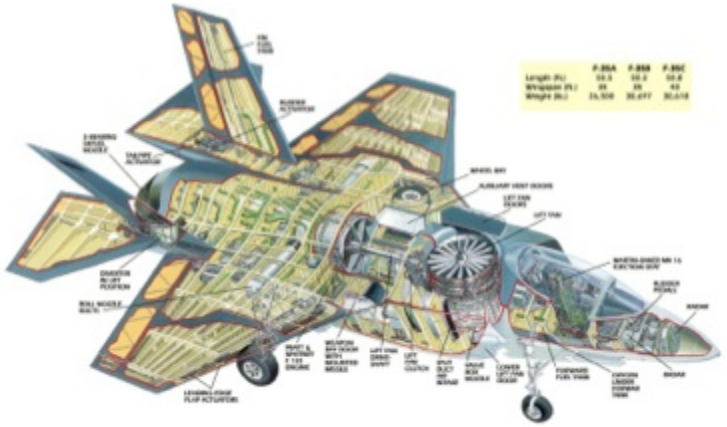


Confluence of Trends

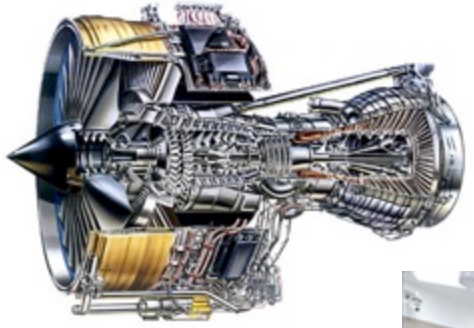
The Overarching Challenge



Traditional Embedded Computing (Cyber+Physical)



Embedded Computing Systems



Emerging Directions






Distribution,
Humans in the Loop,
"Big" Data from the Physical World

Emergence of Social Sensing

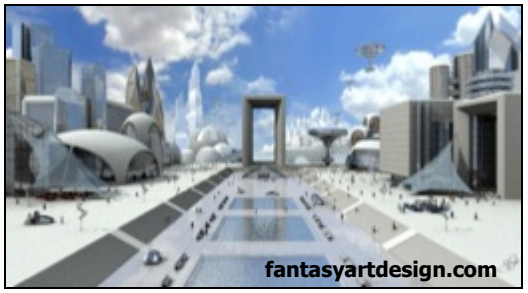
Information Services for a Smarter World

People

Sensors

Analytics

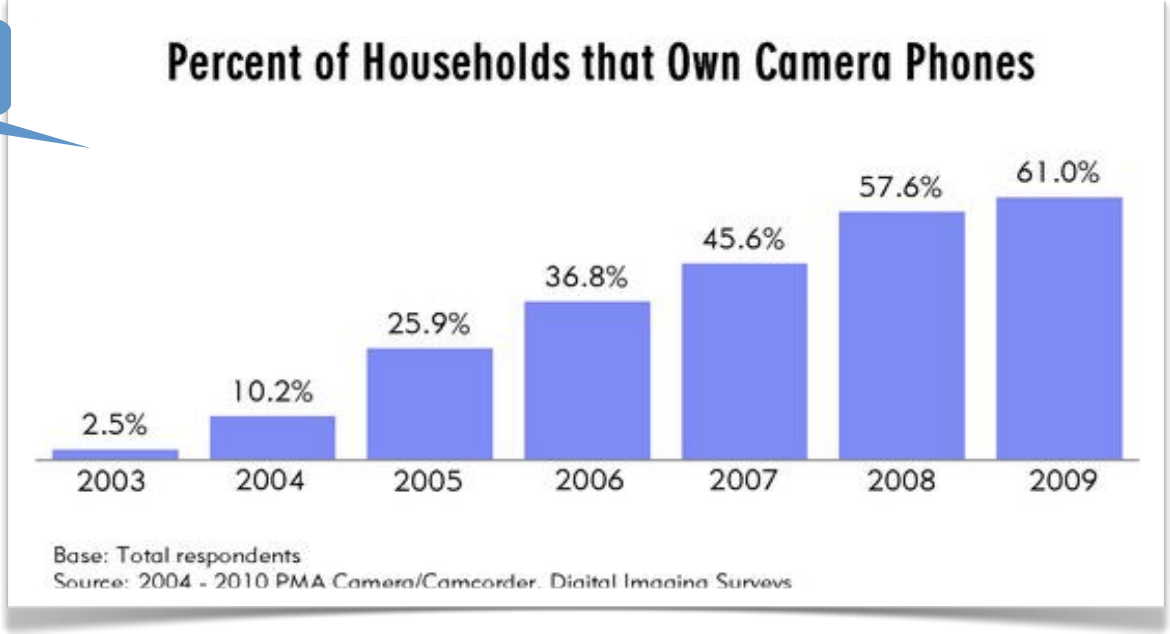
Data



Future Applications

Growth Statistics of Mobile "Sensors"

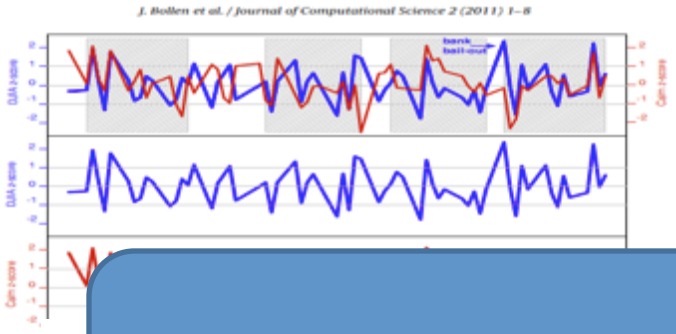
The number perhaps already is much higher now



Household penetration, U.S.	2009	2010	2013
U.S. Households	117.1M	118.6M	122.9Mh
In-vehicle navigation systems	8.1	9.1	13.6
Dedicated navigation devices	34.9	37.4	32.6
Smartphone/PDAs w/navigation	6.6	9.7	27
Handsets w/navigation	6.8	9.7	36.9
Total portable navigation systems	48.3	56.8	96.5

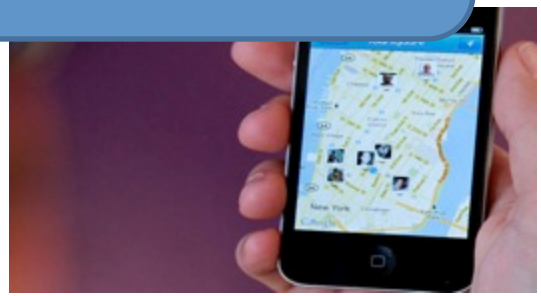
Social Sensing

A set of applications where data are collected from ***human sources*** or devices on their behalf.



Human + Cyber + Physical

Japan Tsunami and Nuclear Event, 2011



FourSquare helps blind people navigate , 2012

Why Social Sensing?
A Confluence of Three Trends

Mass Dissemination Media

twitter

flickr

facebook

YouTube

twitpic
share photos on twitter

Connectivity

Cars on Internet

wimax

4G

Cell-phones

Sensors

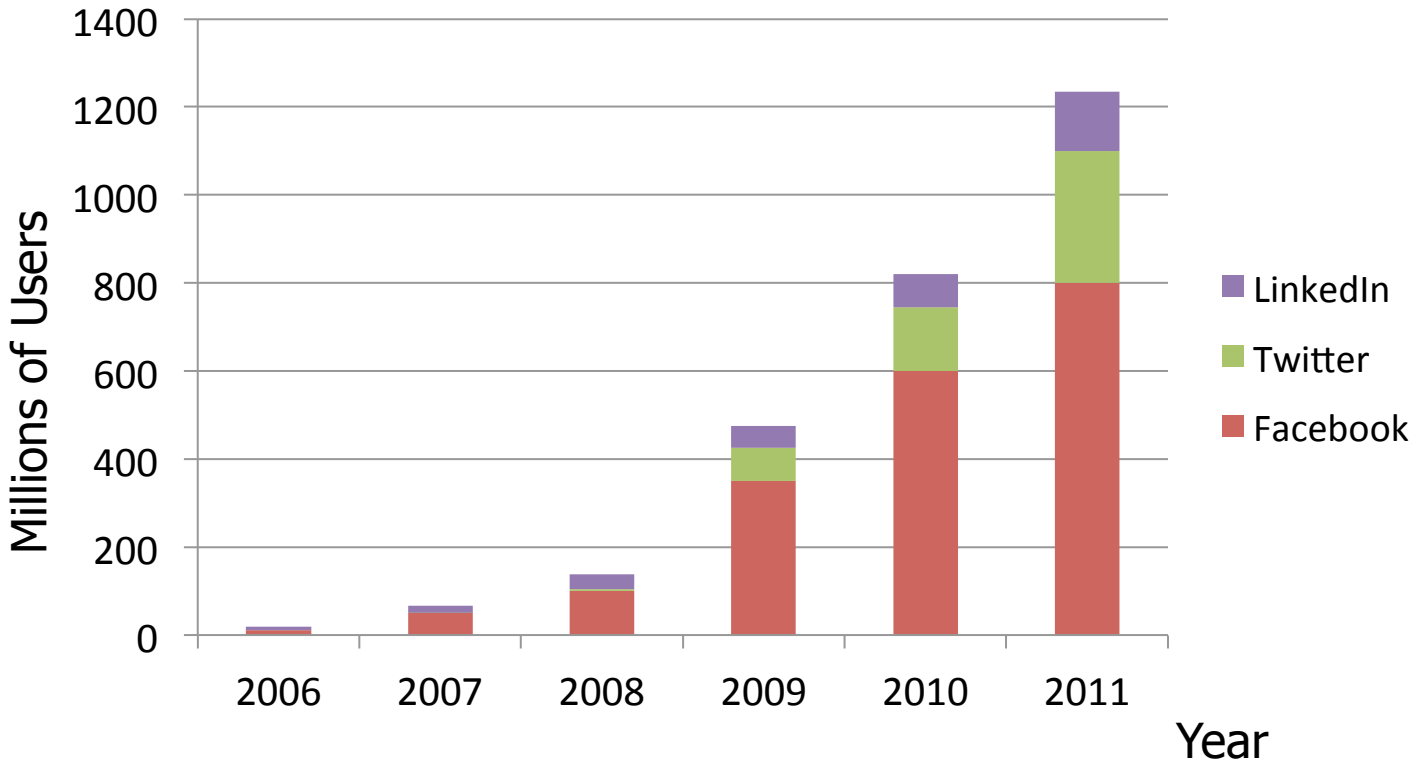
Smart Phone

GPS

Smart Meter

The Rise of Social Media
Platforms for Information Dissemination

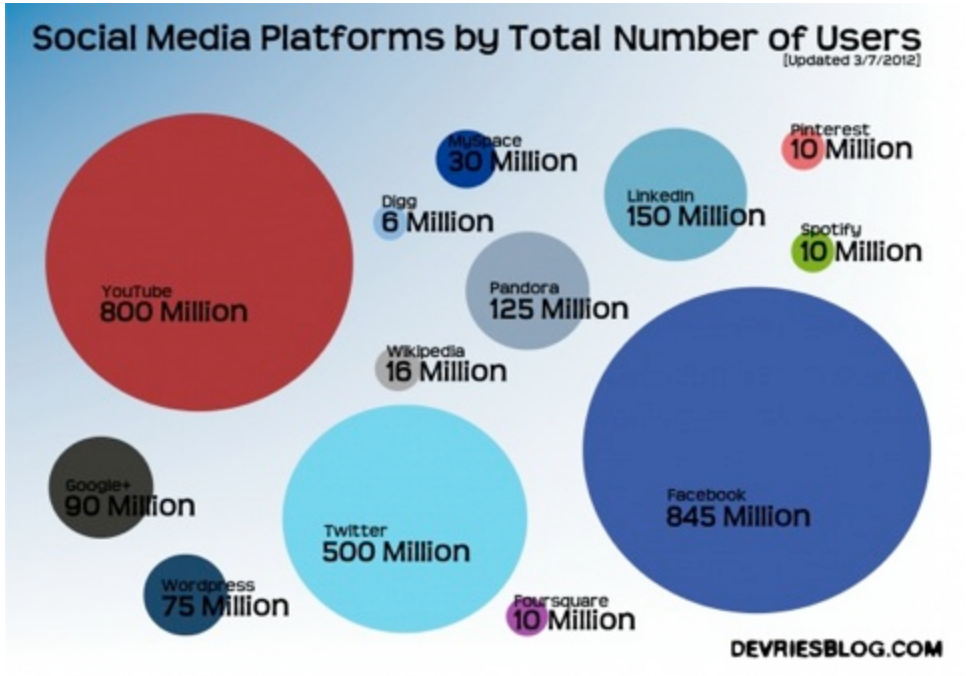
- Significant growth in social networks over the last decade



http://dstevenwhite.com

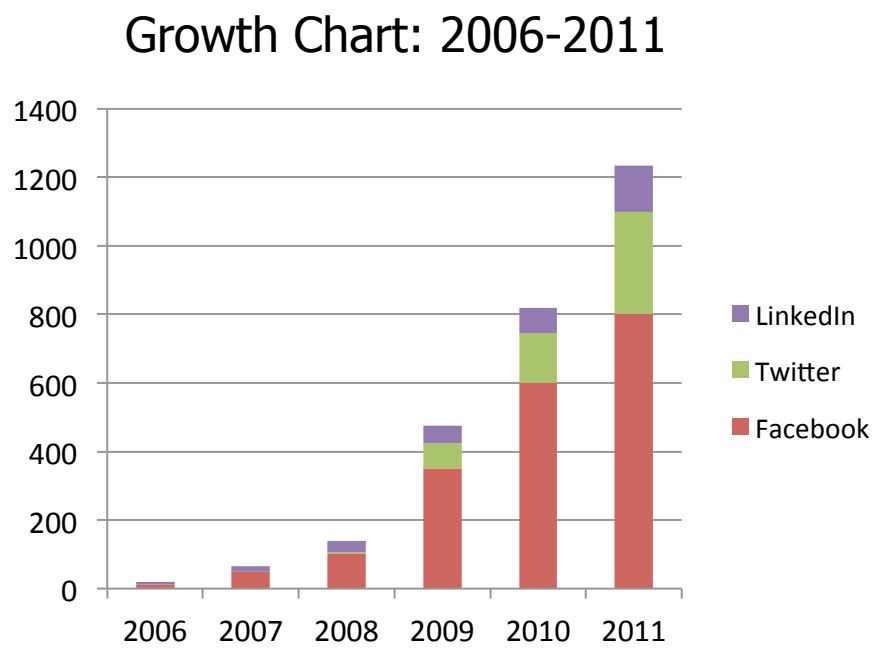
The Rise of Social Media
Platforms for Information Dissemination

- 100s of millions of users on Facebook, YouTube, Twitter, LinkedIn, and Pandora



Total Users in 2011

http://devriesblog.com/category/social-networking/



http://dstevenwhite.com

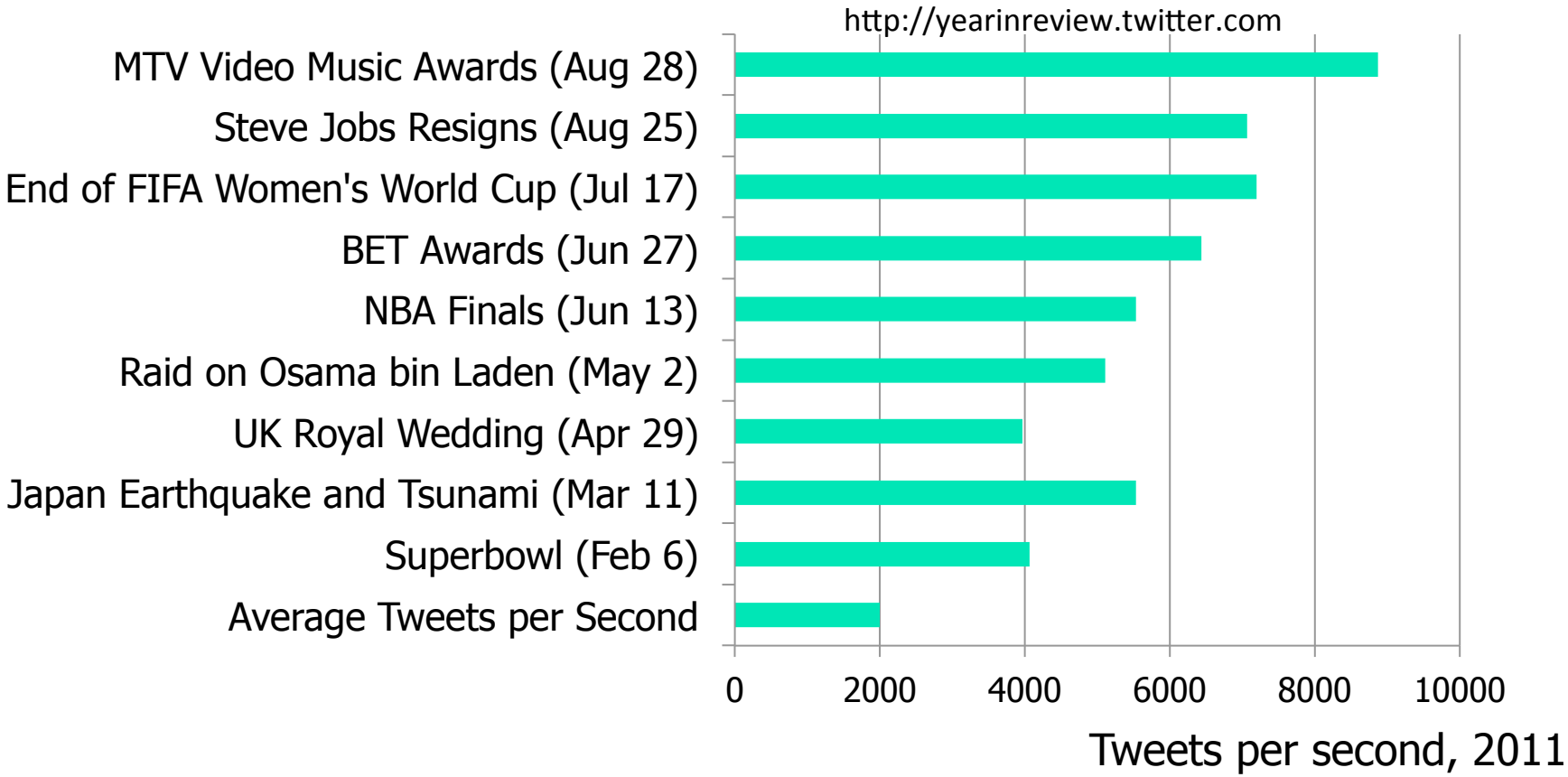
A Wealth of Social Sensing Content

Every Minute:

- More than **350,000 tweets** made on Twitter
  http://www.internetlivestats.com/twitter-statistics/
- Almost **700,000 status updates** made on Facebook
  http://www.themarketingbit.com/infographics/online-for-one-minute/
- More than **3,500 images** uploaded to Flickr
  http://www.pcmag.com
- More than **2000 check-ins** on FourSquare
  http://articles.businessinsider.com
- More than **100 hours of video** uploaded to YouTube
  http://www.youtube.com/t/press_statistics

Significant Information on Ongoing Events

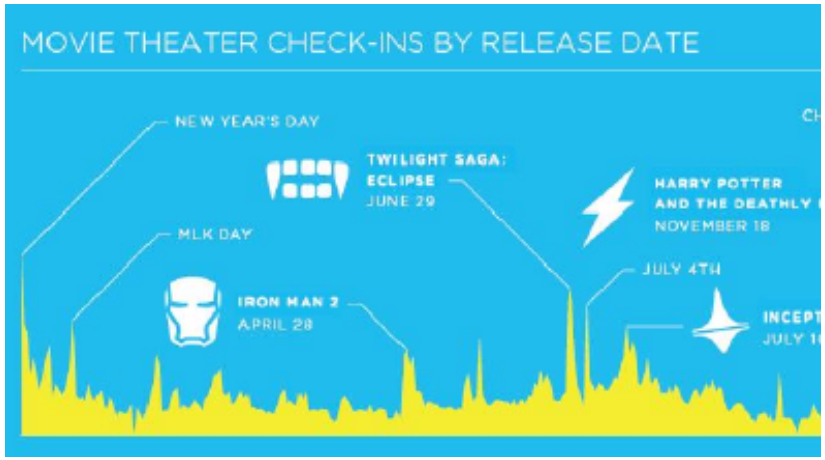
- Thousands of tweets per second during key events



Examples of Real-time Content Analytics

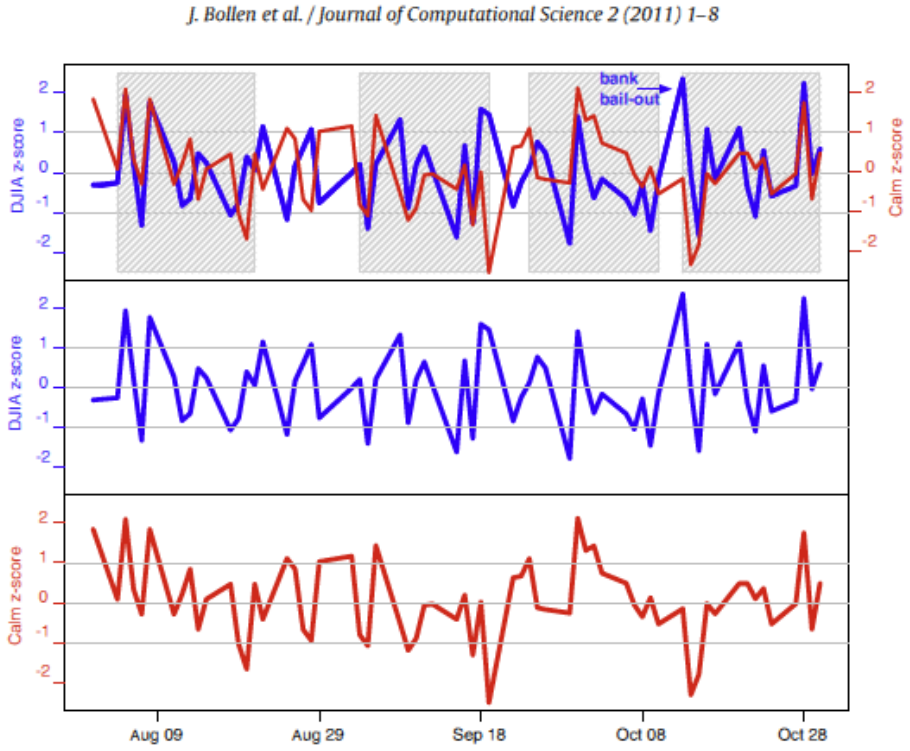
Dataminr

Uses Twitter to analyze and predict global events



FourSquare: Keeps track of human spatio-temporal data

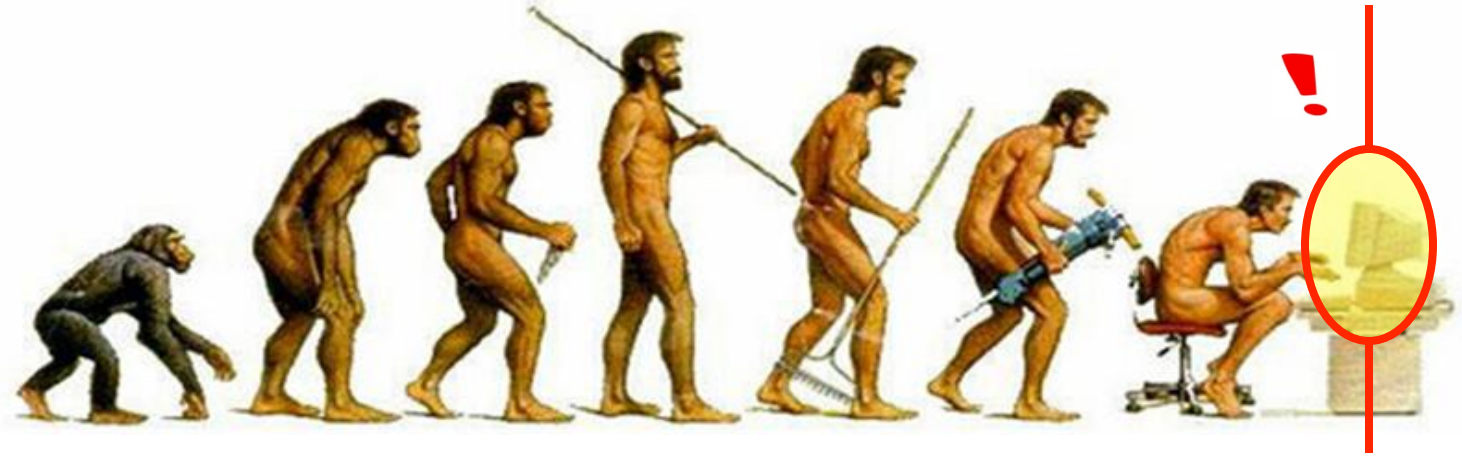
Twitter mood predicts stock market trends



A Deluge of Information!!!

Huge amount of data grows much faster than our cognitive ability to consume it

Fast data growth!



Social
Sensing
Information
Distillation
Services

Slow Evolution!

Towards Information Distillation Services

- Much like Google organizes (relatively static) world content, we need an engine for organizing real-time/streaming data feeds and:

Reconstructing the "State of the World", Physical and Social!

Information distillation

Clean structured representation, high quality of information

A firehose of text, images, video, sound, and time-series data

Towards Information Distillation

- An information refining pipeline

First stage - rough cut: reduce vast amounts of crude input to a much smaller amount of higher value material

Subsequent stages - refinement: Employ specialized techniques to extract final value for end user

Towards Information Distillation

- An information refining pipeline

Our focus:

First stage - rough cut: reduce vast amounts of crude input to a much smaller amount of higher value material

Add-ons:

Subsequent stages - refinement: Employ specialized techniques to extract final value for end user

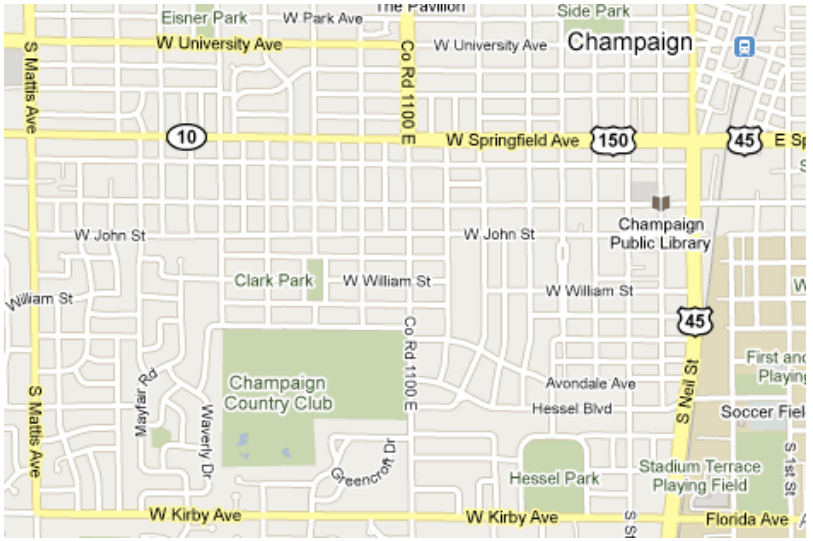
Transportation
EPA Statistics

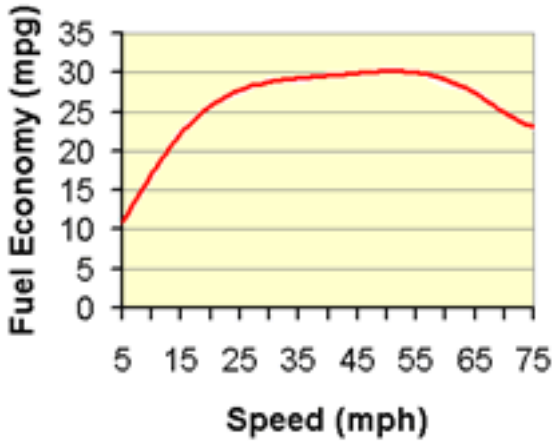
- 200 million light vehicles on the streets
- Each driven 12000 miles annually on average
- Average MPG is 20.3 miles/gallon
- 118 Billion Gallons of Fuel per year!
- Savings of 1% = **One Billion Gallons**

Fuel Efficient Routing

Fuel efficient routes may differ from shortest or fastest routes

- Congestion => shortest may not be fuel efficient
- MPG vs. speed is non-linear => fastest may not be fuel efficient





Source: US EPA

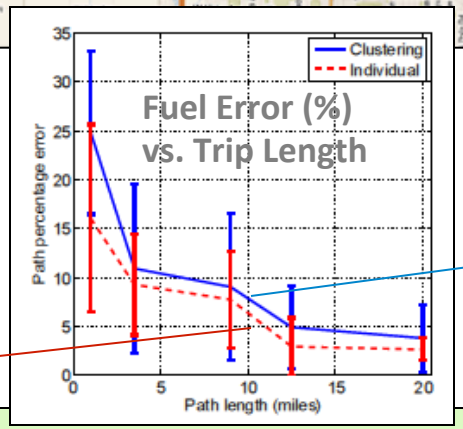
Green GPS

Saves 6% over shortest path
and 13% over fastest path

Shortest and fastest

Most fuel-efficient

Green GPS



Subscribers:
Premium service
High savings

Open access:
Standard service
Average savings

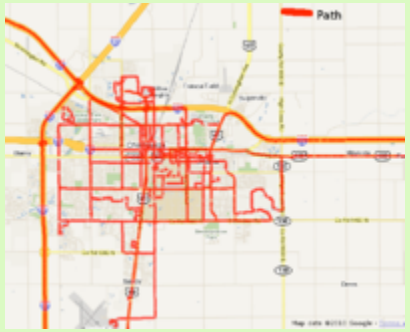
Subscribers

OBDII-WiFi Adaptor ($50)

+

GPS Phone

Fuel Data + Physical Models



$$F_{engine} = \frac{\Gamma(\omega) G g_k}{r}$$

$$F_{air} = \frac{1}{2} c_d A \rho v^2$$

$$F_{friction} = c_{rr} mg\cos(\theta)$$

$$F_g^g = mg\sin(\theta)$$

$$F_{car} = F_{engine} - F_{friction} - F_{air} - F_g$$

Server

Japan's Tsunami and Nuclear Event (2011)

Japan's Tsunami and Nuclear Event (2011)

Zero Energy Buildings

- How can sensing and computing help?

Environmental Technology Center at Sonoma State University

Science House at the Science Museum of Minnesota

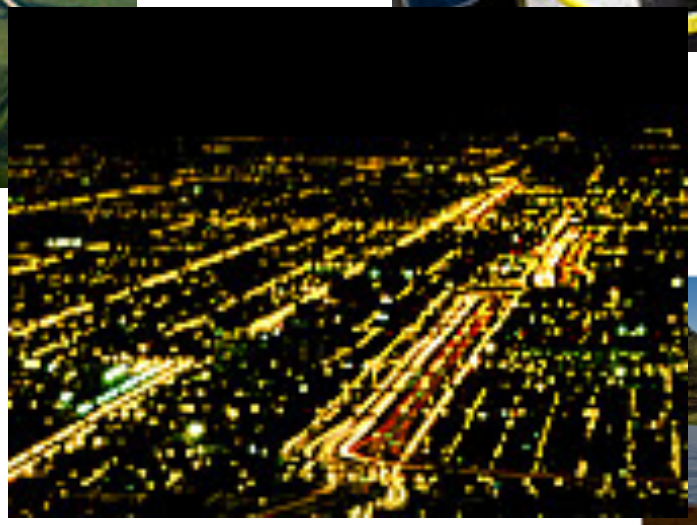
Aldo Leopold Legacy Center

Hawaii Gateway Energy Center

Oberlin College Lewis Center

Smart Grid

- Connecting millions of intermittent sources?



Project Ideas

- Trust and Credibility Analysis
- Disaster Tracking and Report System
- Social Media Command Center for Business
- Personalize Information Subscription Service
- Real-Time Data Analytics
- Multi-genre Network Analysis
- Big Data Processing
- Detect and Reduce Redundant Information
- Geo-location and Spatial Distribution Problem
- Assembling information for heterogeneous data types
- Your own idea ☺

Project Ideas

Trust and Credibility Analysis

- Dow Jones lost 150 points on a rumor of two explosions in the White House on April 23rd, 2013

200 Billion Lost!

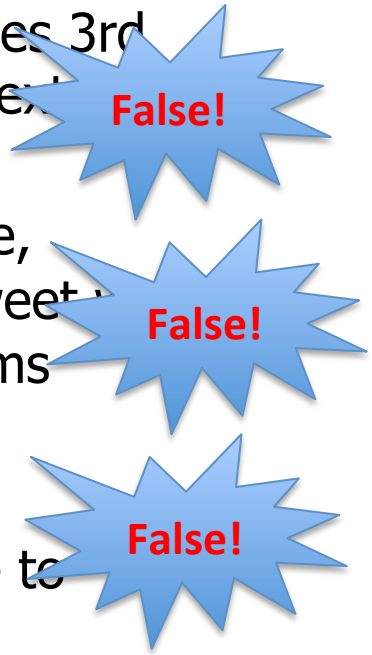
BREAKING: Two Explosions in the White House and Barack Obama is Injured

Project Ideas

Trust and Credibility Analysis

Example tweets in the aftermath of Boston Bombing, April 2013

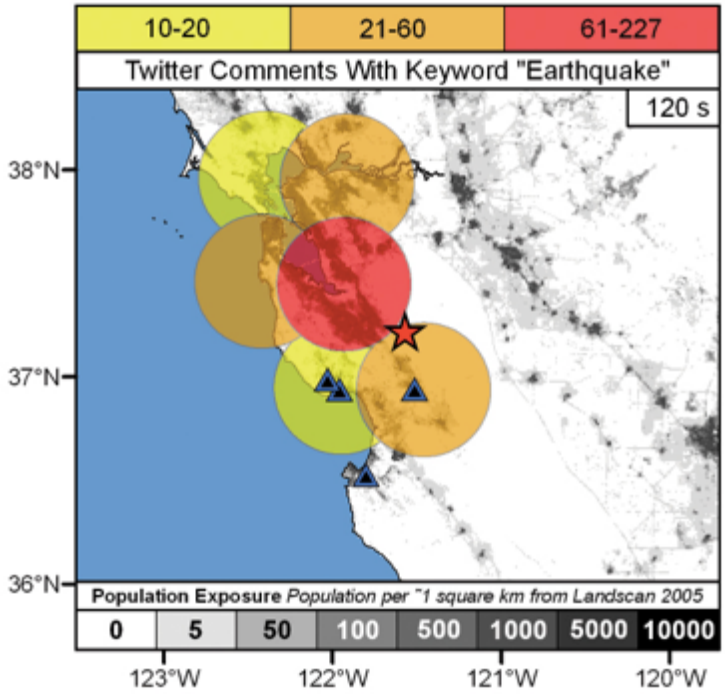
- @USATODAY Boston police commissioner says he believes 3rd blast at JFK Library is related to marathon explosions. Next briefing at 7 p.m.
- @_BostonMarathon. Posing as the organizers of the race, whoever is behind the account tweeted: "For every retweet we receive we will donate $1 to the #BostonMarathon victims #PrayForBoston."
- on Monday that police in Boston had shut down cellular networks to prevent an attacker from using a cell phone to detonate another explosive.

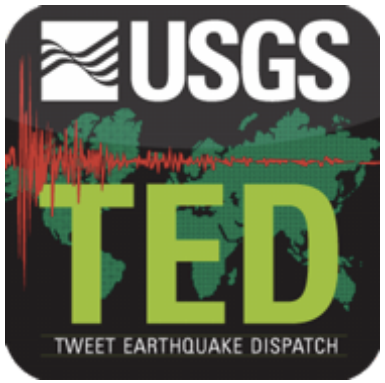


Project Ideas

Disaster Report and Event Tracking

- U.S. Geological Survey is using Twitter to detect earthquake and report sever weather conditions





Project Ideas

Disaster Report and Event Tracking

- In many unrest/riot events, both authority and common citizens use social media to obtain real-time information

Egypt Unrest Feb. 2011

Occupy Wall Street Sept. 2011

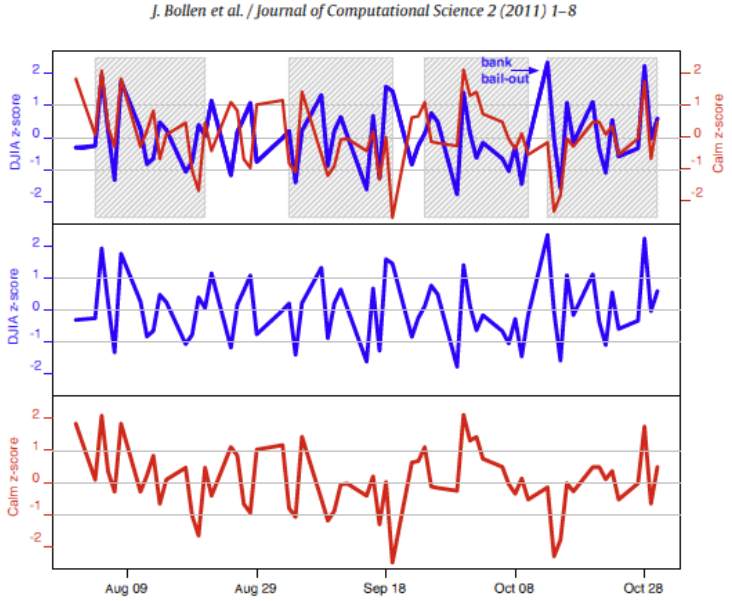
Stockholm Riots, May. 2013

Crimea Unrest, Feb. 2014

Project Ideas

Social Media Command Center

- Large companies start to build their Social Media Command Center (SMCC) to create novel ways of marketing and interacting with customers

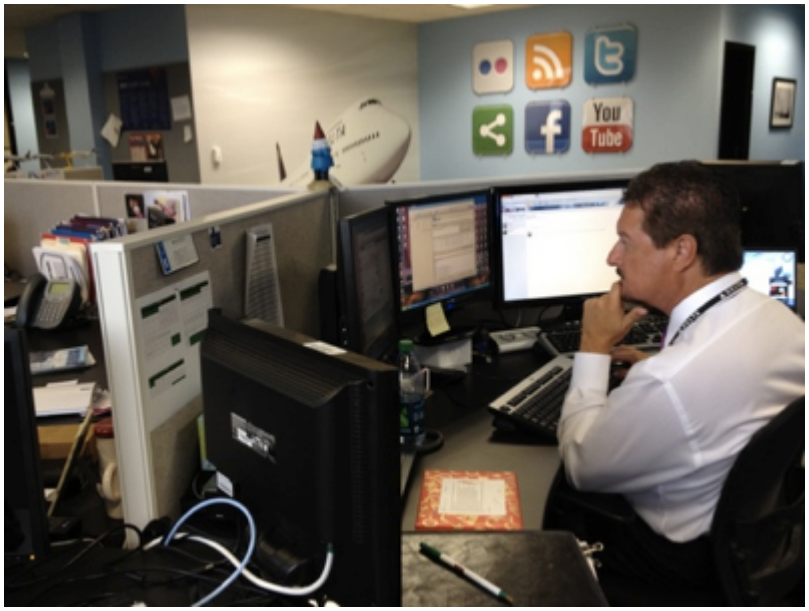


Twitter mood predicts stock market trends

Project Ideas

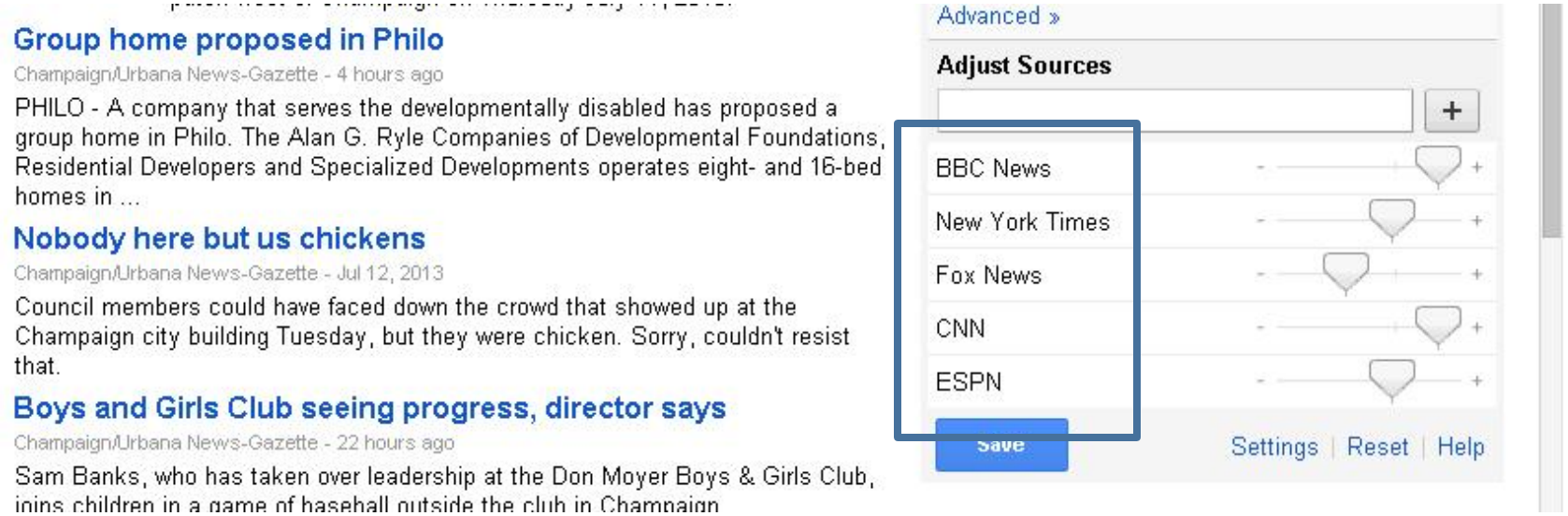
Social Media Command Center

- Delta built its own social media lab



Project Idea

New Information Subscription Service



Project Idea

New Information Subscription Service

Heterogeneous, real-time and potentially unreliable sources

Group home proposed in Philo

Champaign/Urbana News-Gazette - 4 hours ago

PHILO - A company that serves the developmentally disabled … group home in Philo. The Alan G. Ryle Companies of Developmental Foundations, Residential Developers and Specialized Developments operates eight- and 16-bed homes in …

Nobody here but us chickens

Champaign/Urbana News-Gazette - Jul 12, 2013

Council members could have faced down the crowd that showed up at the Champaign city building Tuesday, but they were chicken. Sorry, couldn't resist that.

Boys and Girls Club seeing progress, director says

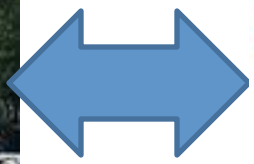
Champaign/Urbana News-Gazette - 22 hours ago

Sam Banks, who has taken over leadership at the Don Moyer Boys & Girls Club, joins children in a game of baseball outside the club in Champaign

Explore the Collective Wisdoms of Common Individuals !

Project Idea

Multi-genre Network Analysis



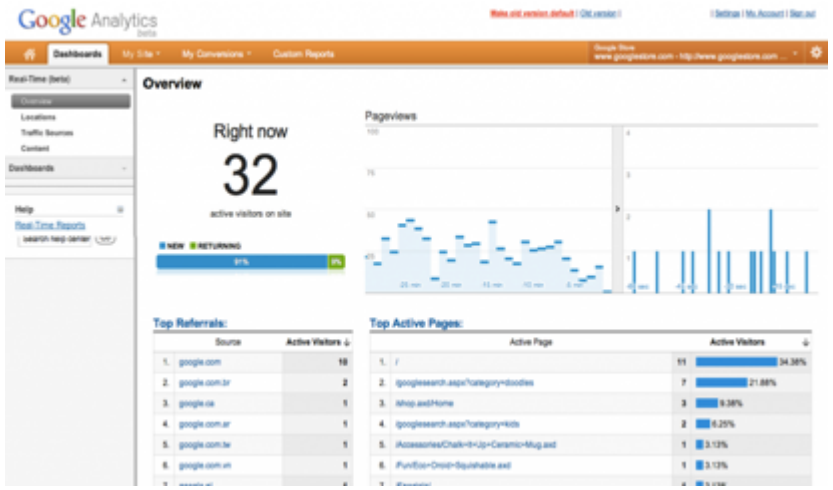
Physical World Anomaly **Social Network Anomaly**

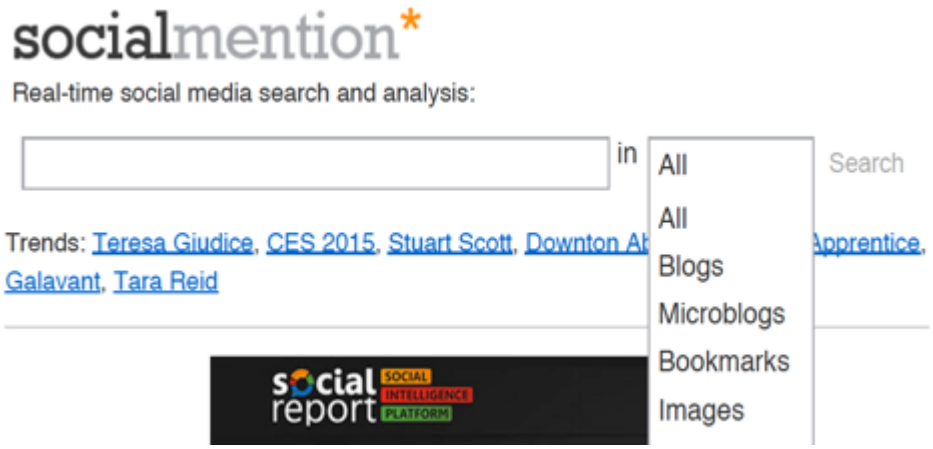
Cross-genre network analysis is essential !

Project Ideas

Real-time Data Analytics

- Google Analytics has started to develop a built-in Real-time Insights function to analyze social media

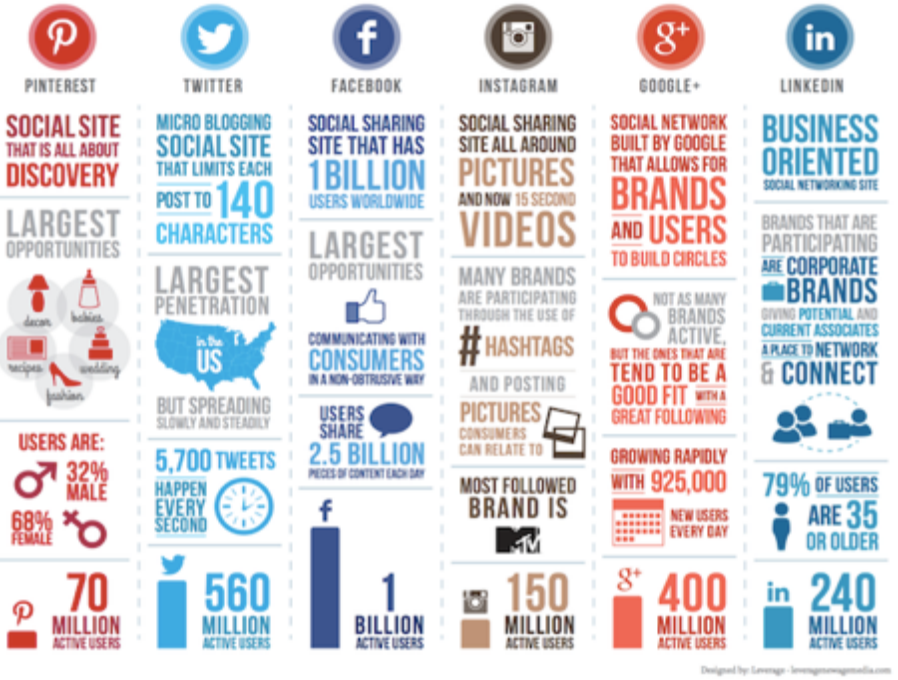




Project Ideas

Big Data Processing

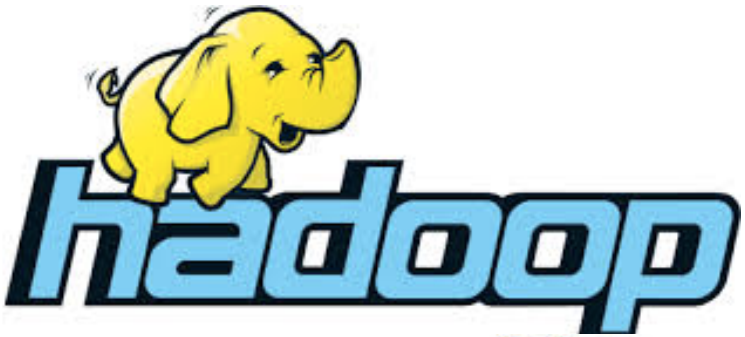
- The data generated by online social media is increasing exponentially in last five years



Project Ideas

Big Data Processing

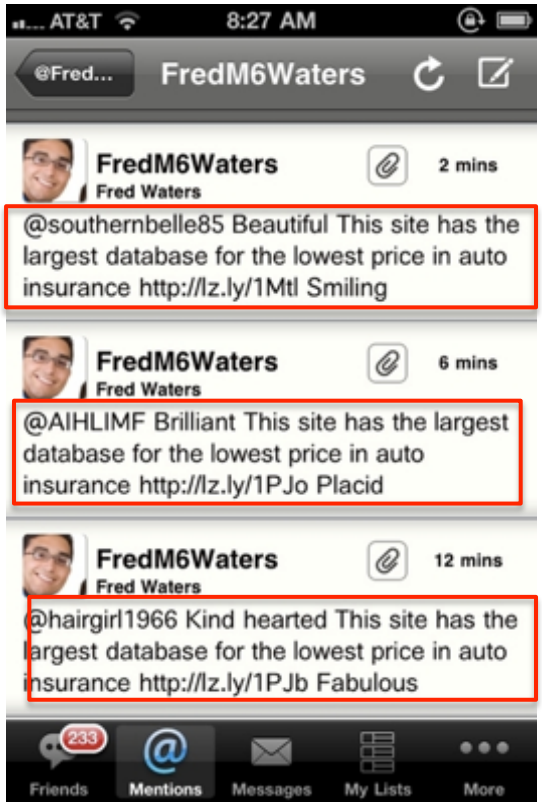
- The data generated by online social media is increasing exponentially in last five years



Project Ideas

Detect and Reduce Redundant Information

- Many messages from on-line social media are forwarded, duplicate, and redundant



t1: Huge New Toyota Recall Includes 245,000 Lexus GS, IS Sedans - http://newzfor.me/?cuye

t3: Huge New Toyota Recall Includes 245,000 Lexus GS, IS Sedans - http://bit.ly/ibUoJs

Nearly Exact Copy

t8: Federal Judge rules Obamacare is unconsitutional...

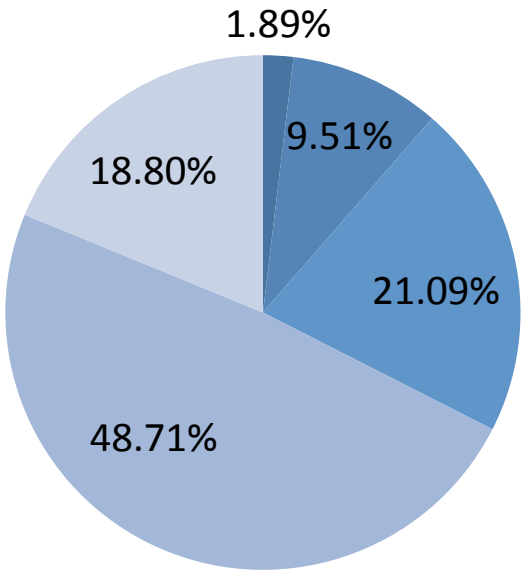
t9: Our man of the hour: Judge Vinson gave Obamacare its second unconstitutional ruling. http://fb.me/zQsChak9

Low Overlap

Project Ideas

Detect and Reduce Redundant Information

- Many messages from on-line social media are forwarded, duplicate, and redundant



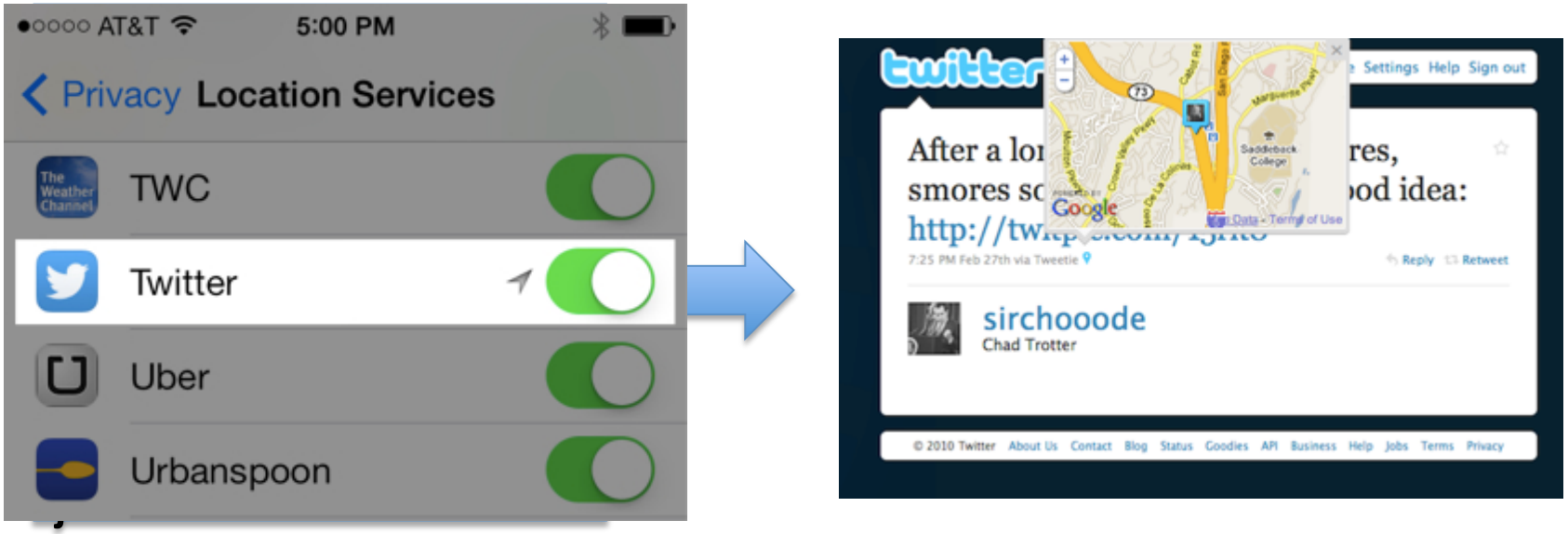
- Exact copy
- Nearly exact copy
- Strong near-duplicate
- Weak near-duplicate
- Low overlapping

1. Detect and reduce redundancy
2. Increase the diversity of search results.

Project Ideas

Geo-location and Spatial Distribution Problem

- Due to the privacy concern, a very small fraction of social media sensing data contains accurate geo-location data



Project Ideas

Geo-location and Spatial Distribution Problem

- Due to the privacy concern, a very small fraction of social media sensing data contains accurate geo-location data

Project Ideas

Structured vs Unstructured Data

- Exploring correlation between different data modalities could help enhance information fusion in social sensing

Project Ideas

Come Up with Your Own Idea!

- It is always fun to come up with your own idea and make it work!

Project Ideas

- If you have any questions, please feel free to come to my office hour or schedule a meeting.