



Microsoft Research  
**FacultySummit**  
**Citizen Sensing**

**Opportunities and Challenges in  
Mining Social Signals and Perceptions**

Amit P. Sheth      amit@knoesis.org

LexisNexis Ohio Eminent Scholar  
Ohio Center of Excellence in Knowledge enabled Computing (Kno.e.sis)  
Wright State University, Dayton, OH <http://knoesis.org>

Thanks: Kno.e.sis team, esp. Wenbo Wang, Chen Lu, Cory, Hemant, Pavan

# Semantics as core enabler, enhancer @ Kno.e.sis

Ohio Center of Excellence in  
Knowledge-enabled Computing

*Computing for Human Experience*

*Social Computing*

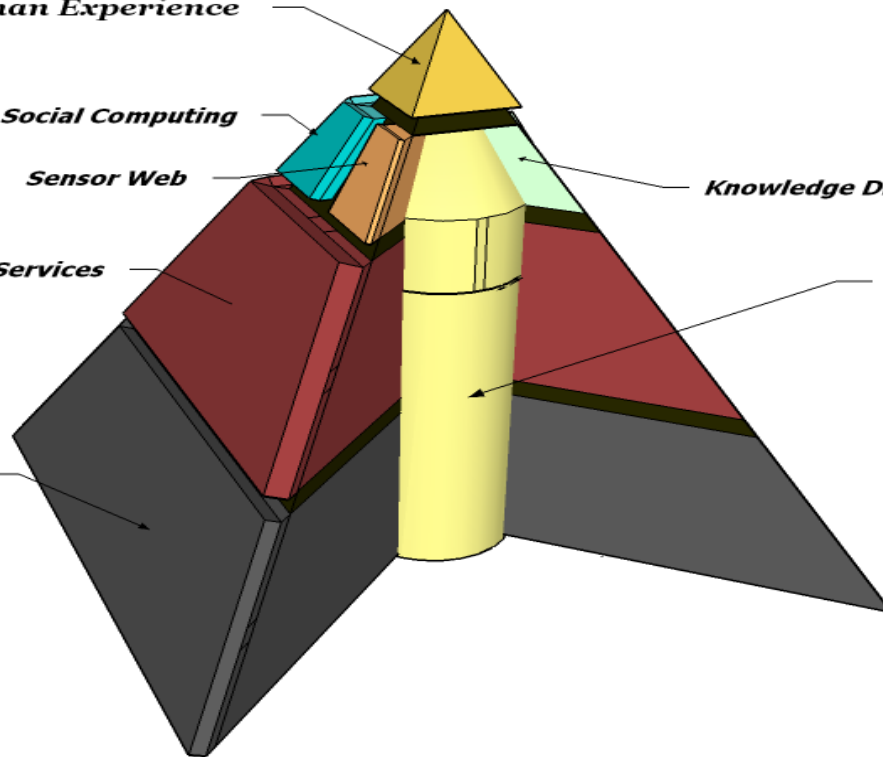
*Sensor Web*

*Web Services*

*Cloud Computing*

*Knowledge Discovery*

*Semantic  
Technologies*



one of the two largest  
academic groups in  
Semantic Web;  
multidisciplinary

# Semantics & Semantic Web in 1999-2002

## (12) **United States Patent** **Sheth et al.**

---

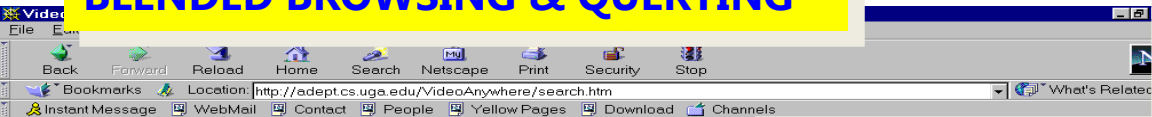
(54) **SYSTEM AND METHOD FOR CREATING A SEMANTIC WEB AND ITS APPLICATIONS IN BROWSING, SEARCHING, PROFILING, PERSONALIZATION AND ADVERTISING**

(75) Inventors: **Amit Sheth**; **David Avant**, both of Bogart; **Clemens Bertram**, Athens, all of GA (US)

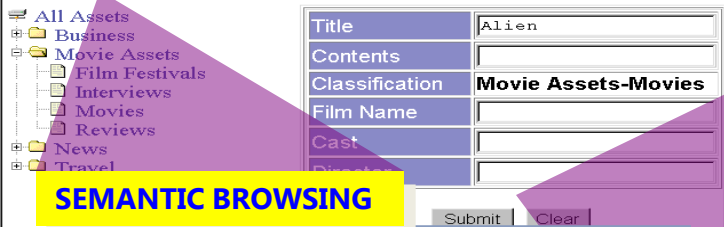


# Taalee Semantic/Faceted Search & Browsing (1999-2001)

## BLENDING BROWSING & QUERYING



## ATTRIBUTE & KEYWORD QUERYING



## SEMANTIC BROWSING

uniform view of worldwide distributed assets of similar type

### Details

**Title:** Alien-Resurrection  
**Director:**  
**Actors:** Sigourney Weaver Winona Ryder  
 The experiments begin. An unholy combination of human and alien genetics, made possible by an uneasy alliance between a renegade band of smugglers and a zealous cadre of scientists and officials. One subject is familiar -- a woman horrifically linked to the alien species that now elicits so much scrutiny. Ripley is back and all is not what it seems... Alien Resurrection stars Sigourney Weaver as Ripley, the role she originated in Ridley Scott's 1979 classic, Alien, and reprised in James Cameron's Aliens (for which she received an Academy Award nomination)

**Contents:**

### Search Results

Title	Details	View	Buy It
Alien-Resurrection			
Alien Resurrection			
Alien Resurrection			
Alien Resurrection			
Alien Resurrection			

## Targeted e-shopping/e-commerce



## assets access



Taalee Semantic Search ...

# Semantic Search/Browsing/Directory (2001-....)

**Links to news on companies that Commerce One**

**Crucial news on Commerce One's competitors (Ariba) can be accessed easily and automatically**

**Other Frequently mentioned**

1. [Ariba, Inc.](#)
2. [Prudential plc](#)
3. [Centrica plc](#)
4. [The IT Group, Inc.](#)
5. [General Motors Corp.](#)
6. [The Union Corp.](#)
7. [First Union Corp.](#)
8. [VerticalNet, Inc.](#)
9. [SAP AG](#)
10. [The Ross Co.](#)

**Links**  **hide** **Company**

Sites of Interest:  
[On24](#)  
[FT News](#)  
[Bloomberg Television](#)  
[Bloomberg](#)  
[Medialink Business](#)  
[StartDust Talk Radio](#)  
[CNET Radio](#)  
[CNNFN The Financial Network](#)  
[FT.com archive](#)  
[DallasBusiness](#)  
[Hoovers Company Data](#)  
[CNNfn The Financial Network](#)  
[CNNfn The Financial Network: Digital Jam](#)  
[dbusiness](#)  
[OhioBusinessEducationClass](#)  
[CBS MarketWatch](#)  
[Executive Radar](#)  
[CATO](#)  
[CNNFN Briefs](#)  
[CNBC Video Highlights](#)

**Ariba, Inc.** via Company

See [Relationships](#)

**Audio/Video clips for Ariba, Inc.**  
Pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [Next](#) ([Next 10 pages](#)) of 24

1. [Ariba, Inc.](#)  
ON The Move: B2B Marketplace Stocks Slump on Downgrade  
Source : ON24  
Category: Business  
Posted: 3/09/2001  
Company Name : Ariba, Inc.
2. [Analyst Cuts Estimates for B2B and E-business Software...](#)  
Thomas Weisel Partners Analyst David Gremmels downgrades Ariba, i2 and Siebel from Strong Buy to Buy and keeps his Market Perform...  
Posted: 3/07/2001  
Source : ON24  
Category: Business  
Company Name : Thomas Weisel Partners LLC, Ariba, Inc., Commerce One, Inc., VerticalNet, Inc., Oracle Corporation
3. [Analyst Cuts Estimates for B2B and E-business Software...](#)  
Thomas Weisel Partners Analyst David Gremmels downgrades Ariba, i2 and Siebel from Strong Buy to Buy and keeps his Market Perform...

# Semantic Search/Browsing/Directory (2001-....)

....)

**SEMANTIC DIRECTORY**  
se-man-tic (sī-mān'tik) relating to meaning

Enter Search Term:  Search

Narrow your search by:  
Start Date:   
End Date:

## Golf

[Top Level](#)  
[Sports > Golf](#)

**Attributes of Golf**  
[Location](#), [Tournament](#), [Course Name](#), [Players]

**Top 5 Players**  
1: [tiger woods](#) 2: [david duval](#) 3: [davis love III](#) 4: [phil mickelson](#) 5: [hal sutton](#)

**Players**  
[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) #

Looking for [david duval](#) [sass] as Players in Golf  
There are 85 results

- [David Duval tests out his swing Monday](#)  
Monday, April 3- Practice begins  
Source : CNNSI Posted: 4/10/2000 Location : Augusta, Georgia
- [Augusta's winds have no friends](#)  
With winds gusting at 42mph, pine cones and limbs strewn all over the course, Vijay Singh still managed to forge ahead with two...  
Source : FoxSports.com Posted: 4/08/2000 Location : Augusta, Georgia
- [Olin Browne chips in on the 14th hole. He would shoot a...](#)

**Related Sites**  
[Fox Sports Golf](#)  
[ESPN US Open](#)  
[PGA Tour 1999](#)  
[PGA Tour 2000](#)  
[Broadcast.com Golf Highlights](#)  
[AugustaGolf](#)  
[Augusta Golf Audio](#)  
[Learn Sports/ Golf](#)  
[U.S. Open](#)  
[British Open 2000](#)

System recognizes ENTITY & CATEGORY

Relevant portion of the Directory is automatically presented.

# Semantic Search/Browsing/Directory (2001-....)

SEMANTIC DIRECTORY

Enter Search Term: david duval

Narrow your search by: Start Date: End Date:

---

SEMANTIC DIRECTORY

se-man-tic (sĭ-măn'tĭk) relating to meaning

Enter Search Term: robert duval

Search

Narrow your search by: Start Date: End Date:

**Users can explore Semantically related Information.**

## Movies

[Top Level](#)  
[Entertainment](#) > [Movies](#)

### Attributes of Movies

[Genre](#), [\[Cast\]](#), [Credits](#), [Location](#), [Film Title](#)

### Top 5 Casts

1: [jack nicholson](#) 2: [matthew broderick](#) 3: [mike myers](#) 4: [bruce willis](#) 5: [milla jovovich](#)

### Cast

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [\[R\]](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) <#>

Looking for [robert duvall](#) <sup>isassi</sup> as Cast in Movies

There are 3 results

#### 1. [The Apostle](#)

Robert Duvall writes, directs and stars in this labor-of-love project about a complex Southern preacher. An intense personality...

Source : Hollywood Posted: 3/29/2000  
 Category: Movie Clips Genre : Drama

#### 2. [A Civil Action](#)

Attorney John Travolta and his small law firm are hired to sue industrial giants accused of contaminating two water wells in a...

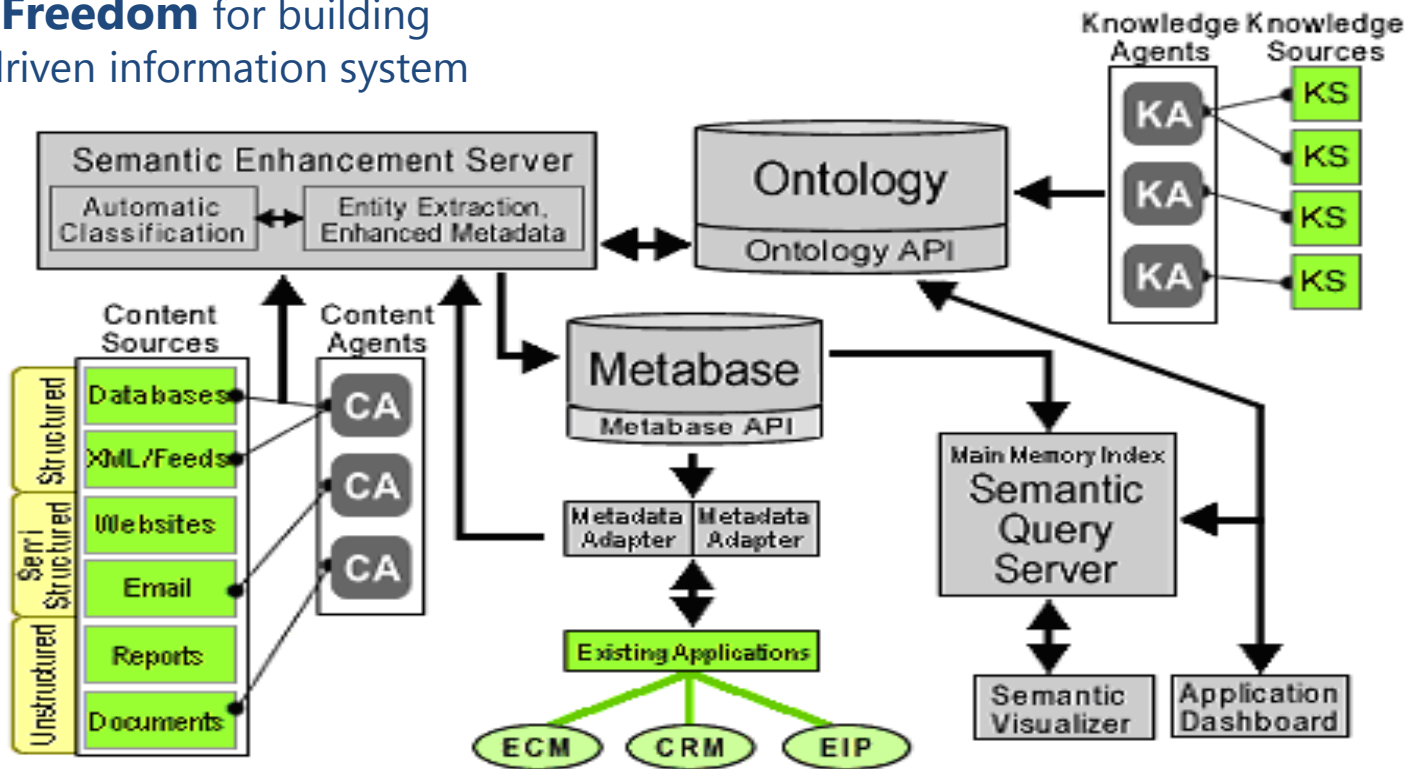
Source : Hollywood Posted: 3/29/2000

### Related Sites

- [SonyMovies](#)
- [MovieWeb](#)
- [Hollywood Movies](#)
- [IFilm\\_video](#)
- [VideoSeeker Movies](#)
- [PayPerView Movies](#)
- [Hollywood Videos](#)
- [Net Movie Mania](#)
- [ComingSoon](#)
- [Sookie Todd&apos;s](#)
- [Instuclional Videos](#)
- [Pulp Kitchen](#)
- [Drivein- Sony](#)
- [Oscar](#)
- [indiafmscene](#)
- [moviesounds12monkeys](#)
- [MovieSounds](#)
- [movieclips](#)
- [oxygenclips](#)
- [film.com-Trailer-fishing](#)

# Extracting Semantic Metadata from Semistructured and Structured Sources (1999 – 2002)

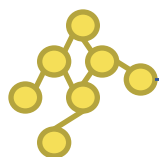
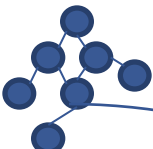
**Semagix Freedom** for building ontology-driven information system



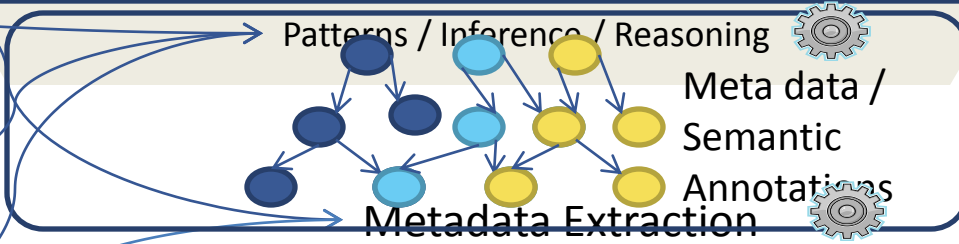


Fast forward to 2010-2011

Domain Models



Search  
 Integration  
 Analysis  
 Discovery  
 Question  
 Answering  
 Situational  
 Awareness



# Let Us Start with Social Data

# Jan. 2011 Egypt Protest

Image:<http://bit.ly/qHI7wl>



Image:<http://bit.ly/g4yPXS>



Image:<http://bit.ly/qmDocA>



monkey: With @Gemhood @aGharbeia @oshaokhtmeligi for sharing food and water. Awesome. #jan25

ben since 1991. Because you can't firewall a rumor. #Jan25

Tweet By: @AnonOnABike

RT @GabyVerdier: RT @justicentric: They r preping to attack us now. Tightening cordons and working out antiriot cars. Pray for us #jan25

Tweet By: @egyptenguiden

Tear gas bombs are being fired at the protesters in tahrir sq. #jan25

Tweet By: @Bounz

RT @fustat: R @bencnn Hearing bangs and pops from Tahrir....sounds like trouble #Jan25

Tweet By: @bernhardhopfner

News

<http://bit.ly/nP1E4q>

# Twitter, Facebook become lifelines after Japan quake

Image:<http://bit.ly/fl4gEJ>

<http://cnet.co/jdQgME>

MAY 6, 2011 5:51 PM PDT

## Japan radiat

by Eric Mack

Recommend 91



Safecast.org seek  
(Credit: Screen ca

011

<http://bit.ly/gWboib>

Recently funded NSF proposal: Social Media Enhanced Organizational Sensemaking in Emergency Response



Image:<http://bit.ly/nqq6Wj>



**Heater34:** [@VinniePolitan](#) hey Vinnie, waiting on the word from you, I'm sitting in holiday **traffic** on **I75** in KY!!!  
about 24 hours ago via *Twitter for Android* · [Reply](#) · [View Tweet](#)



**nickmomrik:** Just used an on ramp to get off **I75**. **#traffic**  
2 days ago via *Twitter for iPhone* from *Kawkawlin, MI* · [Reply](#) · [View Tweet](#)



**rickirose19:** Stuck in **traffic** on **#I75**, sucks.  
2 days ago via *Twitter for iPhone* · [Reply](#) · [View Tweet](#)



**MrsWright8579:** **I75** heading up North (MI) **traffic** jam..anyone know what's up?  
Didn't expect to be down to one lane :(  
2 days ago via *Twitter for BlackBerry®* · [Reply](#) · [View Tweet](#)



**Tornadogeneral:** N a **traffic** jam on **I75** siting still. That's all I do is sit in cars behind the wheel **#trafficwreck**  
3 days ago via *Twitter for iPhone* from *Alabama, US* · [Reply](#) · [View Tweet](#)

Jul. 2011

# I-75 Traffic Jam in US

# Citizen Sensing

- Who?
  - An interconnected network of people
- What?
  - Observe, report, collect, analyze, and disseminate information
- How?
  - Via text, audio, video and built in device sensor (and smart devices)



Image: <http://bit.ly/nvm2iP>

# Enablers: Mobile Devices & Ubiquitous Connectivity

- Mobile Platforms Hit Critical Mass, Over 5 billions users
  - 1+B with internet connected mobile devices (2010)
  - Smartphones > Notebooks + Netbooks (2010E)
  - 500K+ mobile phone applications
  - 74% of mobile phone users (2.4B) worldwide used SMS (2007)
- Mobile is Global
- Ubiquity, 24x7
- Built in sensors
  - environmental, biometric/biomedical,...



Image: <http://bit.ly/mYqcPF>



# Enablers: Web 2.0 & Social Media

- A huge number of users
  - 750M+ active Facebook Users
  - 1+B tweets/wk; 175M+ Twitter users
  - Internet Users: 2 Bln



Image: <http://bit.ly/euLETT>

# Role of Semantics in Citizen Sensing

- Key of citizen sensing: extract metadata/annotate
  - different types of metadata (depend on application need)
    - Spatial, temporal, thematic: key phrase, named entity, relationship, topic/category, event descriptors, sentiment ...
    - People, network, content
- Semantics: provide the meaning of data
  - various forms of semantic models: core vocabularies/nomenclatures, community created dictionaries/folksonomies/reference databases, automatically extracted domain models, manually created taxonomies, formal ontologies
  - deal with complexities of user generated data; supplement well-known statistical and natural language processing (NLP) techniques

# Research Application: Twitris

# Twitris - Motivation

- What were people in U.S.A. saying about Bin Laden's death?
- How about people in Egypt?
- How about people in India?
- TOO MANY tweets to be read each day!!!
- Twitris
  - Now: WHEN, WHERE, People are talking WHAT
  - Future: socio, cultural, behavioral studies

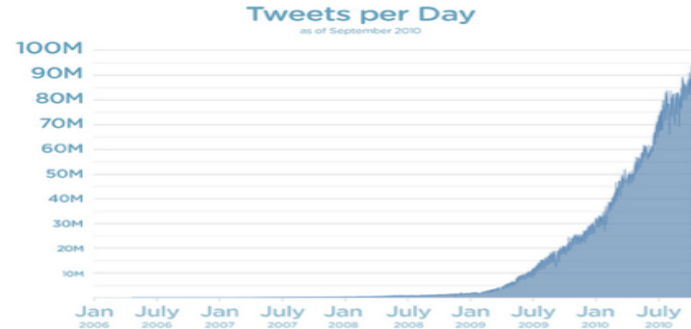


Image: <http://bit.ly/etFezl>

# Twitris: Semantic Social Web Mash-up

N-gram summaries

Select date

Topic tree

Select topic

How to Twitris

Date: 06/22/2009

Twitris Map : Iran Election 2009

Change the Topic of Interest

June 2009

You are looking at data from United States

Images & Videos

Tweet traffic

Sentiment Analysis

Spatial Marker

Related tweets

Reference news

Wikipedia articles

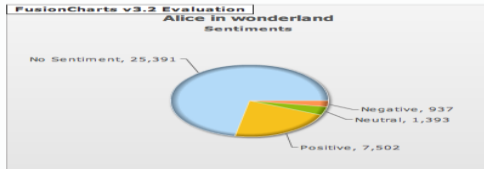
Related Tweets and Articles

Entity Relationships

Multimedia

Tweet Traffic

Sentiment Analysis



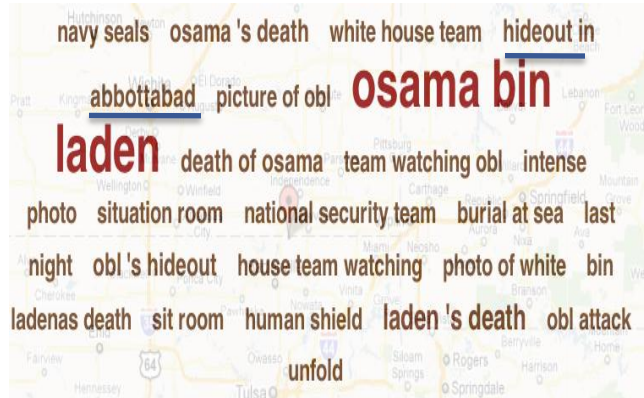
## Positive Sentiments

- just watched alice in wonderland with kanthes. so--so only. not 'WOW!' and the 3D effects was.. lame.
- Alice in Wonderland in 3D ... So cool!!!! I really like Tim Burton's movies!!! effects was.. lame. --- Thinking twice..
- Saw Alice in Wonderland was a bit slow but loved it anyways
- It's not so much that the critics were wrong... But Alice In Wonderland was still visually stunning and fun to watch
- OMG! I found the New Alice in Wonderland Online! It's Great! Use your email to watch and enjoy! Hurry! promo <http://tinyurl.com/ybgqzn4>
- A little tough morning prayer only had a couple hours asleep watched a late showing Alice in wonderland love the mad hatter
- saw Alice in Wonderland this weekend! lovely movie! listening to soundtrack now, it's pretty cool too.
- Alice in Wonderland Is Another of sweet Movie from Disney<3 ReCommended :)
- Alice in wonderland in 3D is amazing. Takes me back to my younger days. Sigh.

# Analyzing Events from Temporal Perspective

- How did tweets in **United States** on the **death of Bin Laden** evolve over time?

May  
2nd



RT @ReallyVirtual: Here's a picture of OBL's **hideout in Abbottabad**, as shared by a friend @Rahat <http://yfrog.com/h7w4izmj>

May  
4th



Img:<http://www.twitpic.com/4t1mt0>

# Analyzing Events from Spatial Perspective

- Tweets (**Death of Bin Laden**) in Egypt VS tweets in India

May  
2<sup>nd</sup>  
Egypt



"#Egypt foreign minister: Egy gov't has no official comment but we condemn all **forms of violence** in international relations. #osama #obl"

May  
2<sup>nd</sup>  
India



"RT @mvatlarge: U.S. has given **Pakistani military** nearly \$20 billion since 9/11 for the privilege of housing bin Laden: <http://is.gd/xegnFm>"

# A sample of current research @ Kno.e.sis demonstrating role of semantics in Citizen Sensing & Social Media Analysis



# User-community Engagement Analysis

# User-community Engagement

- How do we understand the phenomenon of user participation (engagement) in topic discussions?
  - How communities form during the product launch?
  - What factors can attract users to engage in these communities, therefore further spreading the message?
  - How quickly we can disseminate information between resource providers and people in need of resources in case of emergency?

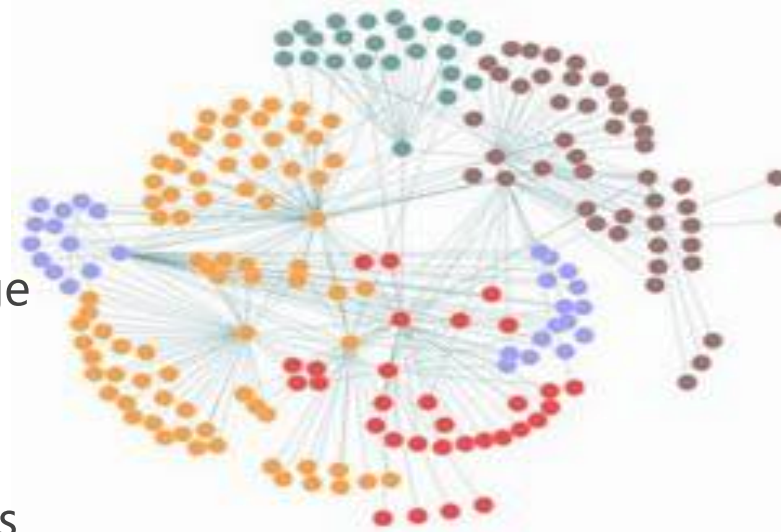
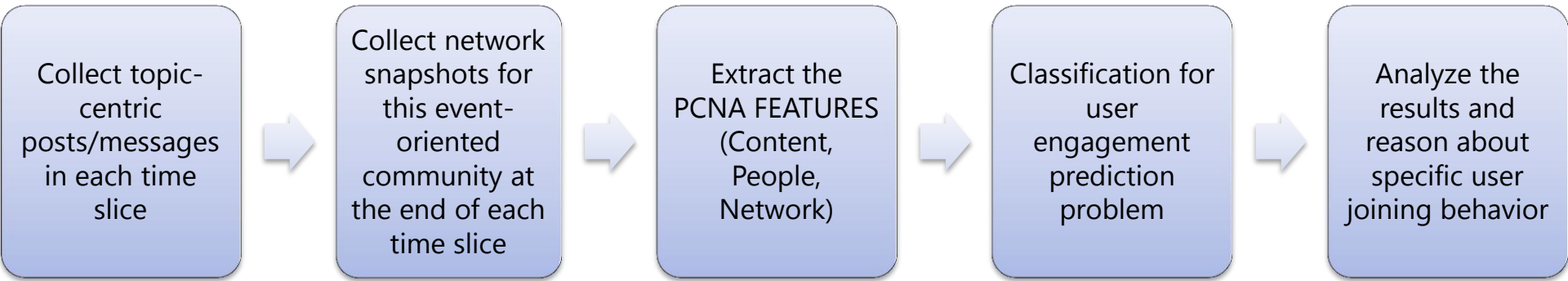


Image: <http://itcilo.wordpress.com>

# Analysis Framework:

## People-Content-network Analysis (PCNA)



# Three Sets of Features

Which set is more important?

- **Content features** [Characteristics of tweets posted by active friends of  $U$ ]:
  - **keywords**: number of event-relevant keywords
  - **hashtags**: number of event-relevant hashtags
  - **retweet**: number of retweets
  - **mention**: number of mentions
  - **url**: number of relevancy-adjust hyperlinks
    - Irrelevant hyperlink is given number -1
  - **subjectivity**: Subjectivity scores for words and emoticons
  - **Linguistic Cues** (LIWC<sup>1</sup> analysis): Features for the language usage. Top-3 transformed features using Principle Component Analysis (PCA) extracted
- **Community features**: [Characteristics of the active community/network under consideration]
  - **wccSize**: size of the weakly-connected component (WCC) which  $U$ 's friends belongs to in the active network.
  - **wccPercent**: ratio of **wccSize** to the size of the active network.
  - **connectivity**: number of active friends (i.e. followees) in the community.
  - **communitySize**: size of the active community.
- **Author features** [Characteristics of friends that  $U$  is following]:
  - Only friends in the active community are considered.
  - **logFollower**: logarithm of follower count
  - **logFollowee**: logarithm of followee count
  - **Klout<sup>[1]</sup>**: a integrated measure of user influence and popularity
  - Other profile information and activity history<sup>[2]</sup>.

# Experiments: *Results*

Content is the key to understand user community engagement

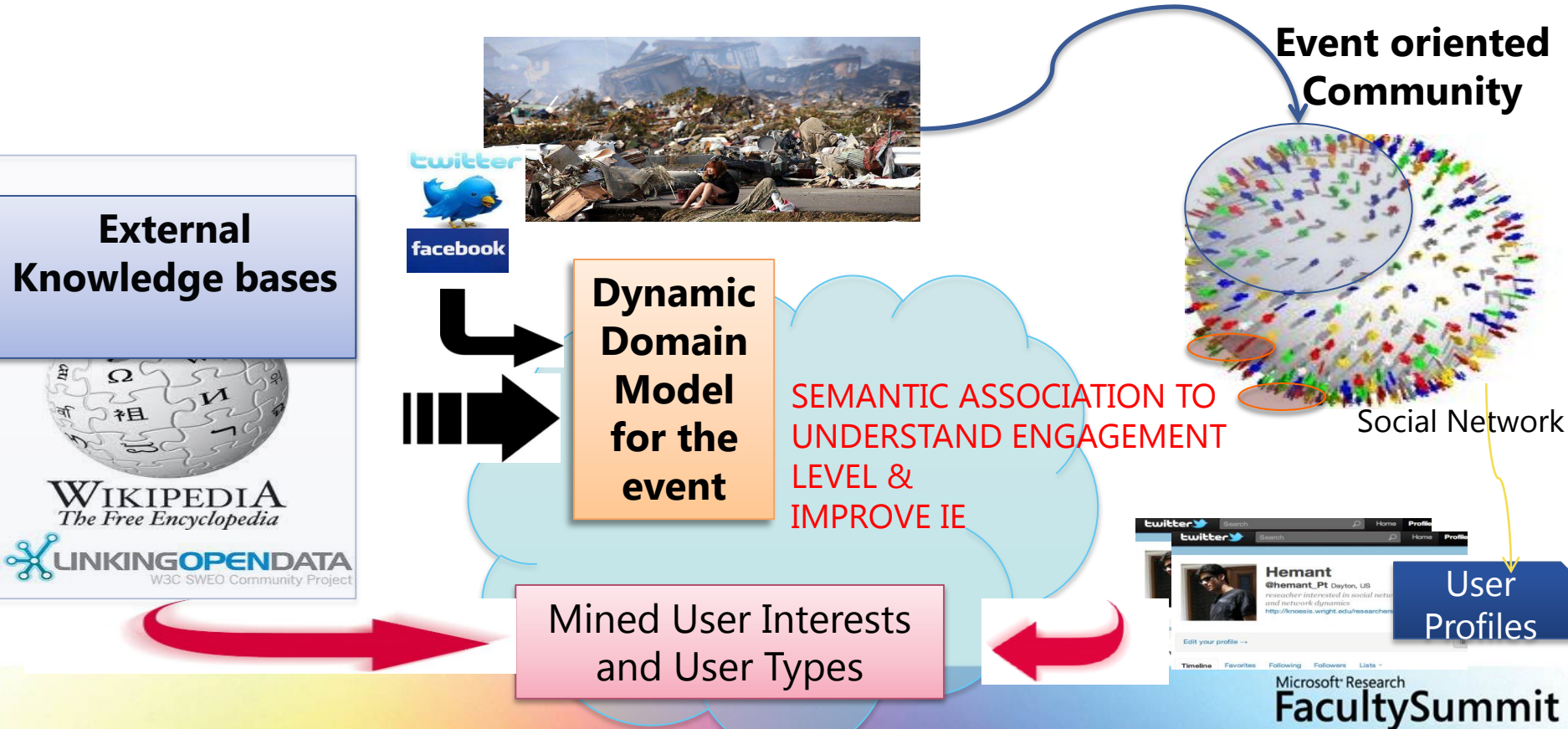
Performance High All Content People Network Low



Events	All	Con.	Aut.	Com.	Event-Type	
					U	L
DiscoveryBuildingCrisis	<b>77.86</b>	75.95	71.31	69.65	U	L
GoogleInstantSearch	<b>76.25</b>	74.92	72.23	52.60	U	L
RichCroninDeath	<b>90.68</b>	<b>90.96</b>	<b>90.36</b>	68.47	U	L
StuxnetWorm	<b>76.05</b>	<b>76.46</b>	72.05	57.51	U	L
HeismanTrophy	<b>76.88</b>	75.28	69.94	61.85	U	C
ClevelandShowPremiere	<b>86.11</b>	<b>85.77</b>	<b>85.65</b>	67.36	D	L
EmmyAwards	<b>77.00</b>	<b>77.39</b>	70.93	56.23	D	L
IowaStateFair	83.34	<b>84.25</b>	81.62	70.09	D	L
LindsayLohanHearing	<b>80.09</b>	79.30	77.22	52.57	D	L
LondonTubeStrike	<b>82.40</b>	<b>82.96</b>	80.07	56.22	D	L
ScottPilgrimRelease	<b>78.16</b>	<b>77.86</b>	75.32	59.81	D	L
JewishNewYear	<b>75.15</b>	74.14	69.16	55.63	D	C
LinuxCon	80.77	<b>82.17</b>	76.97	71.97	D	C
SESSanFrancisco	<b>75.50</b>	<b>76.40</b>	71.69	58.34	D	C

Summary of Prediction Accuracy (%)  
 Statistical significant results are in bold

# Background Knowledge to improve Social Data Analysis



# Analysing the Content can be Hard...

This **new** Merry Christmas tune.. **SO GOOD!**

Is Merry Christmas a song?

If it is, which 'Merry Christmas' since there are 60 songs of the same name.

'So Good' is also a song!

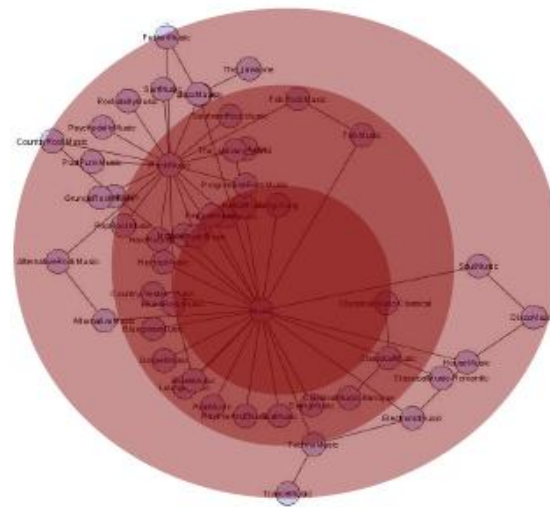
Using a domain model (E.g., MusicBrainz)

Using context cues from the content

- e.g. **new** Merry Christmas tune

Reduce potential entity spot size (with restrictions)

- e.g. **new** albums/songs



# Real Time Social Media Data Analysis



# Motivation

- People can't wait for Information
  - Disaster Management
    - Ushahidi ([www.usahidi.org](http://www.usahidi.org))
  - Real-Time Markets
    - RealTimeMarkets (<http://www.realtimemarkets.com/>)
  - Brand Tracking
    - Twarql (<http://wiki.knoesis.org/index.php/Twarql>)
  - Movie reviews
    - Flicktweets ([www.flicktweets.com](http://www.flicktweets.com))
  - Journalism

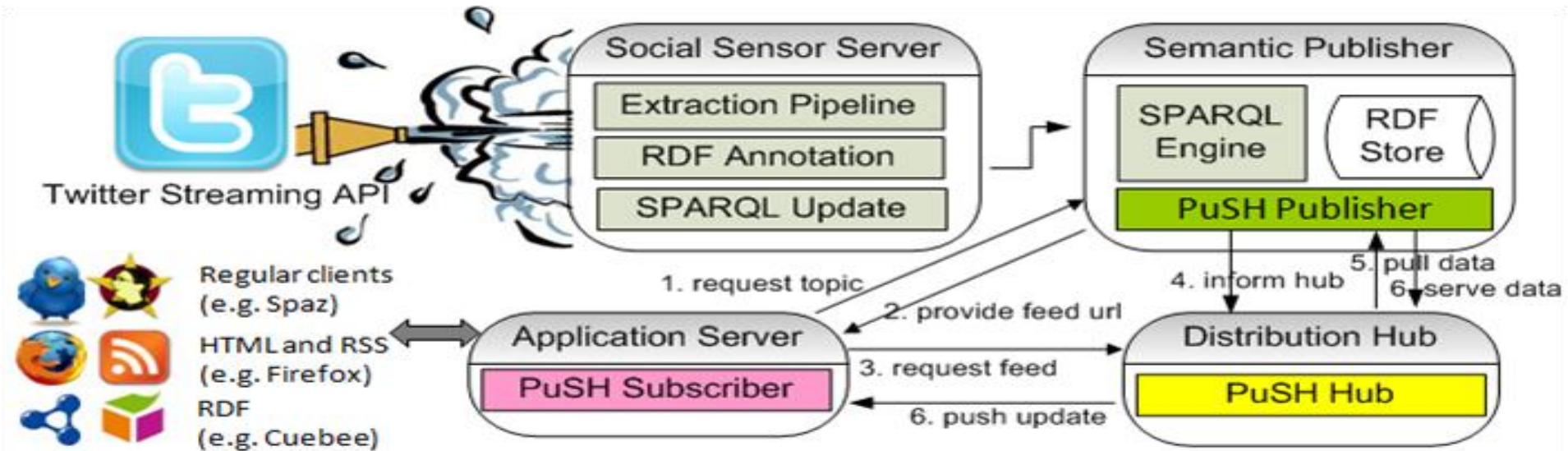
# Scenarios

- Brand Tracking
  - Give me a stream of locations where Kinect is being mentioned right now
  - Give me all people that have said negative things about Kinect
  
- How can we do this?

# Twarql (Twitter Feeds through SPARQL)

- Semantically annotate tweets with entities, hashtags, URLs, sentiments, etc.
- Encode content in a structured format (RDF) using shared vocabularies (FOAT, SIOC, MOAT, etc.)
- Structured querying of tweets
- Subscribe to a stream of tweets that match a given query
- Real-time delivery of streaming data.

# Twarql Architecture



# Back to the Scenario

- Give me a stream of locations where Kinect is being mentioned now

```
SELECT ?location
WHERE {
?tweet moat:taggedWith dbpedia:Kinect .
?presence opo:currentLocation ?location .
?presence opo:customMessage ?tweet .
}
```

- Give me all people that have said negative things about Kinect

```
SELECT ?user
WHERE {
?tweet sioc:has_creator ?user .
?tweet moat:taggedWith dbpedia:Kinect .
?tweet twarql:sentiment twarql:Negative .
}
```

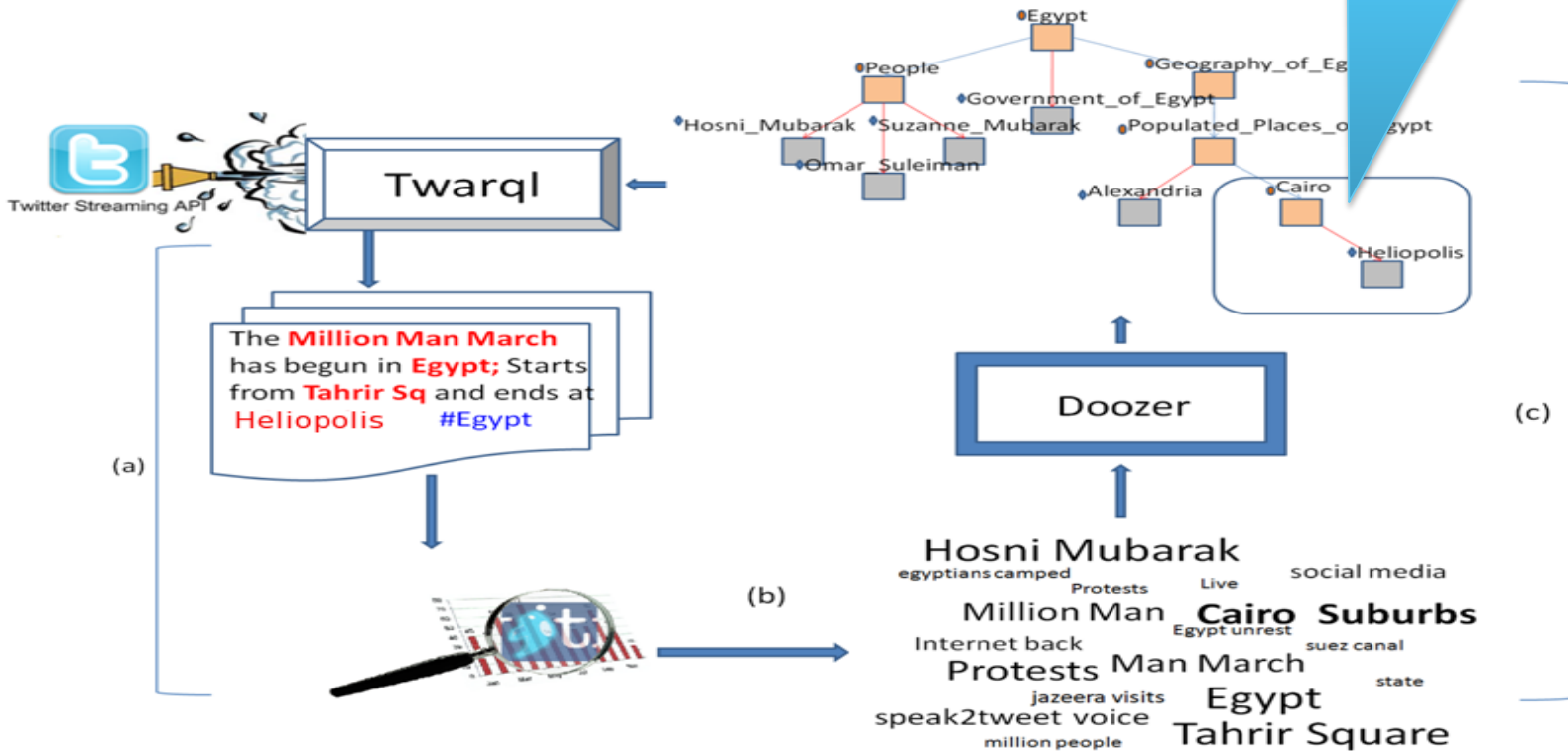
Dynamic Domain Models for Semantic  
Analysis of Real-Time Data  
aka  
Continuous Semantics

# Motivation

- Semantic processing using a model of the domain
- But it is difficult to model **dynamic domains** on social web
  - spontaneous (arising suddenly)
  - real-time data requiring continuous searching and analysis
  - distributed participants with fragmented and opinionated information
  - diverse viewpoints involving topical or contentious subjects
  - feature context colored by local knowledge as well as perceptions based on different observations and their sociocultural analysis.

# Dynamic Model Creation

Heliopolis is a suburb of Cairo.



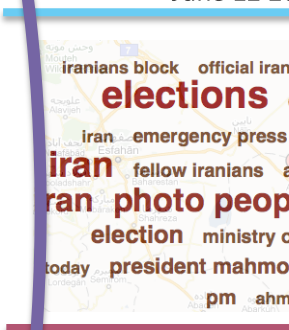


# Dynamic Evolving Models to underpin Semantics

Events  
↓  
Key phrases  
↓  
Models

"Both **Ahmadinejad & Mousavi** declare victory in **Iranian Elections**."

June 12 2009



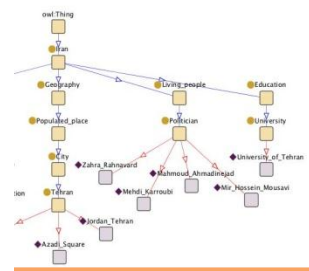
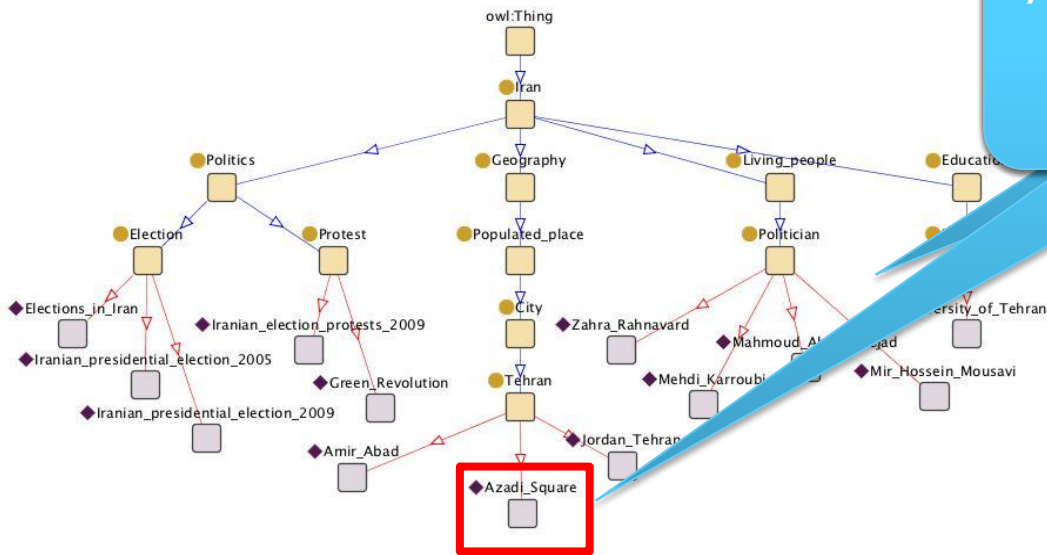
"situation in **tehran University** is so worrisome. police have attacked to girls dormitory #tehran #iranelection"

June 13 2009

"Reports from **Azadi Square** - 4 people killed by police, people killed police who shot. More shots being fired #iranelections"

June 15 2009

Azadi Square is a city square in Tehran



# Sentiment/Opinion Extraction

# Challenges

- Domain/Topic-dependency: spotting the target of the sentiment is as important as finding sentiment itself
  - E.g., “**long** river” (no sentiment), “**long** battery life” (positive), or “**long** time for downloading” (negative).
- Context-awareness : encoding the context information into the extracted sentiment
  - E.g., “**must** watch a movie today” (no sentiment) and “this movie is a **must** see” (positive).
- Informal language (abbreviations, misspelling, slang...): using Urban Dictionary

# The Usage of Background Knowledge

## Background Knowledge (Identify the Target Entity)

### IMDB (movie name, year, genre...)

The King's Speech 2010 Drama  
Iron Man 2 2009 Action  
Edge of Darkness 2010 Crime...

### Freebase (person name, birth day, country...)

Colin Firth 09/10/1960 England Actor  
Helena Bonham Carter 05/26/1966  
England Actress...

## Lexical resources (Identify the Sentiment Clue)

### Urban Dictionary

**wicked**: awesome cool sweet sick amazing rad great good tight....  
(US ---> really, very, extremely UK ---> cool, good, great, funky  
France ---> (as 'terrible') great)

Loved the King's Speech, Colin Firth is so amazing.



Seen **Iron Man 2** yesterday! Really **wicked** movie! **Loved** it. (From UK?)  
Saw **Edge of Darkness**. **Pretty cool** but **wicked\_boring** in some parts. (From US?)

# Real Time Feature Stream

# Web DATA evolved over time



**Real-Time Sensor,  
Social, Multi-media  
data**

**Dynamic User  
Generated Content**

**Static Document  
and files**

# Semantic Abstraction

- Overwhelming amount of raw sensor data does not make much sense to decision makers

<b>Time</b>	<b>6pm</b>	<b>7pm</b>	<b>8pm</b>	<b>9pm</b>
Temperature 1(C)	-1	-2	-4	-4
Rainfall (mm/h)	0.5	1	1	0

Does -1,-2,-4,-4 make any sense to you?

**Freezing temperature**

**Rain**

- What if the data is from sensors on a highway?
  - Freezing temperature + rain => icy road
  - Close the highway? OR Spread salt on the road to melt ice?

So what?

A cross-country flight from New York to Los Angeles on a Boeing 737 plane generates a massive **240 terabytes** of data

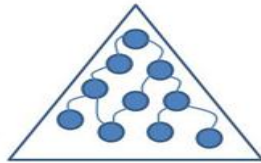
- *GigaOmni Media*

But a pilot or a ground engineer at the destination is interested in very small number of events and associated observational data that are relevant to their work.

Higginbotham, S. (2010, September). Sensor Networks Top Social Networks for Big Data. Gigaom.com. <http://gigaom.com/cloud/sensor-networks-top-social-networks-for-big-data-2/>.



# Abstraction



Background Knowledge



ABSTRACTION



**Huge amount of  
Raw Sensor Data**



**Features representing  
Real-World events**

# Weather Alert Application

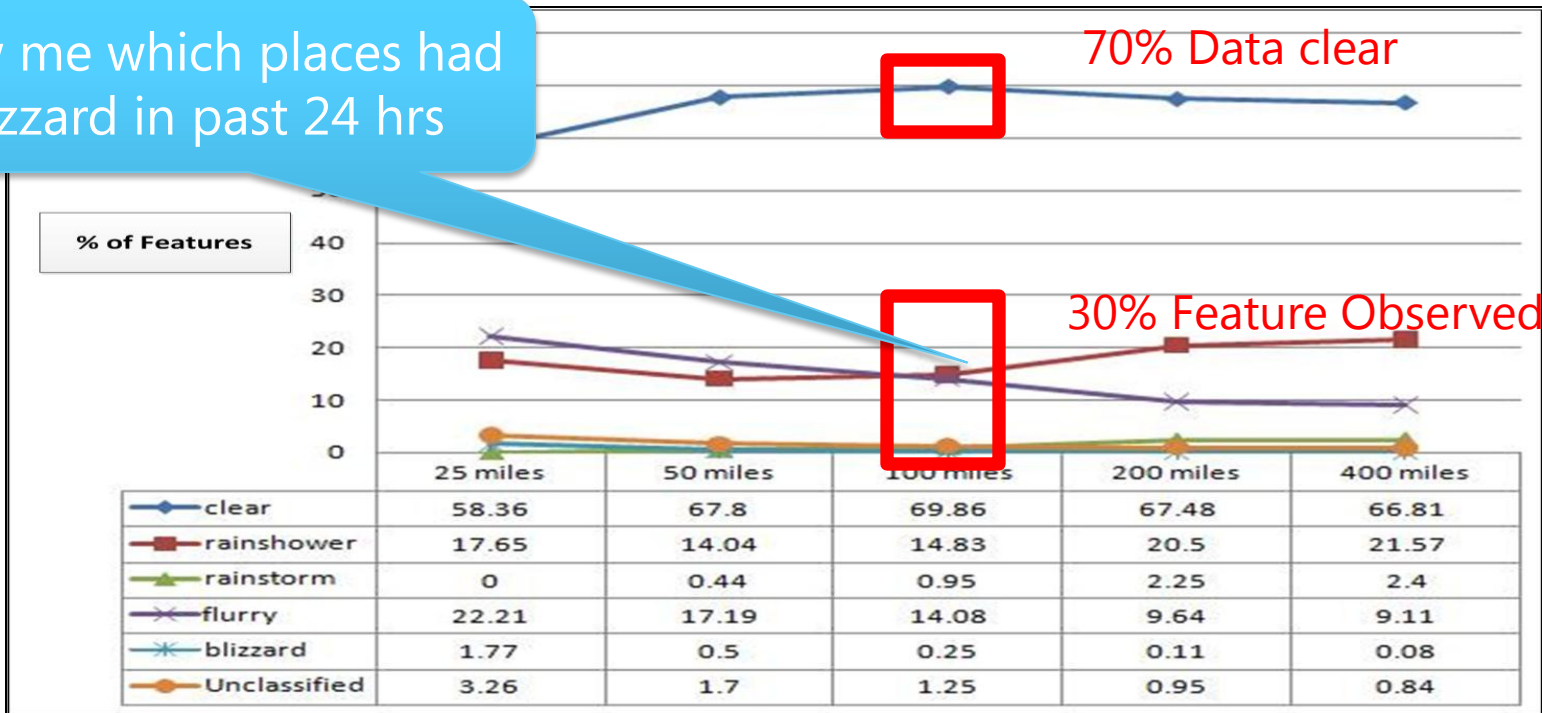
Detection of events, such as blizzards, from weather station observations



# Evaluation

- **Data Used:** Nevada Blizzard (April 1<sup>st</sup> – April 6<sup>th</sup>)

Show me which places had blizzard in past 24 hrs

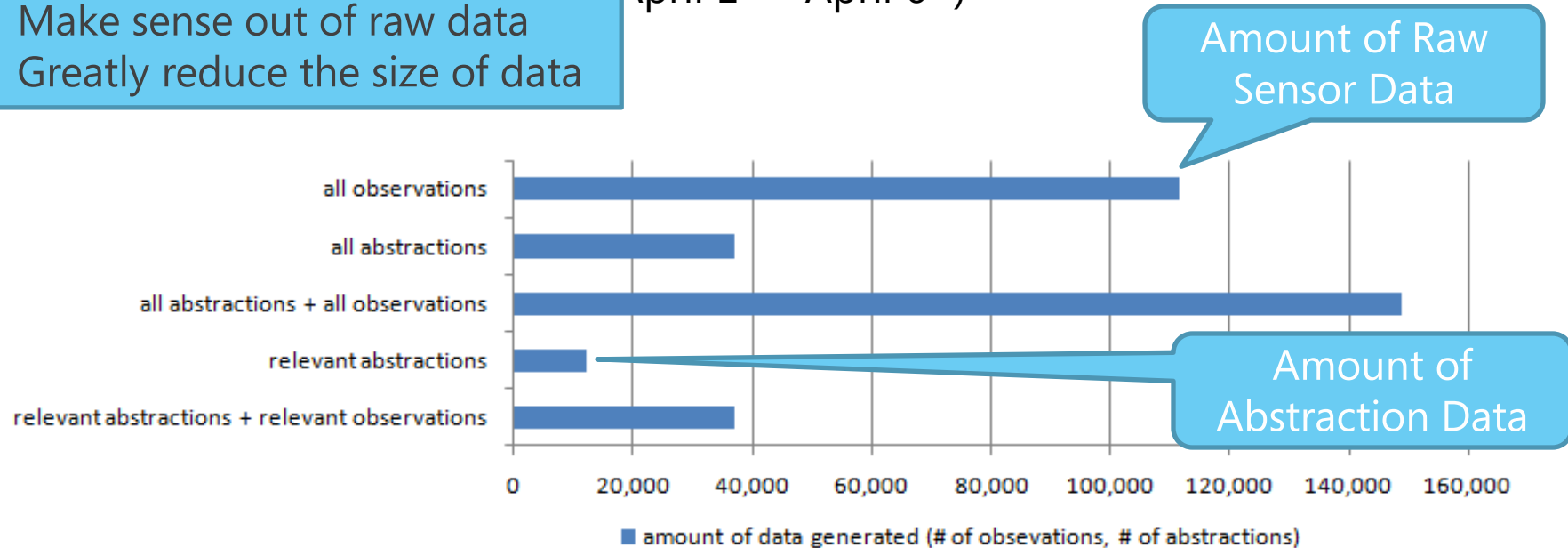


# Evaluation (cont.)

Abstraction can:

1. Make sense out of raw data
2. Greatly reduce the size of data

April 1<sup>st</sup> – April 6<sup>th</sup>



# Traffic Application

Reasons can be found via other types data: tweets, news papers, etc.

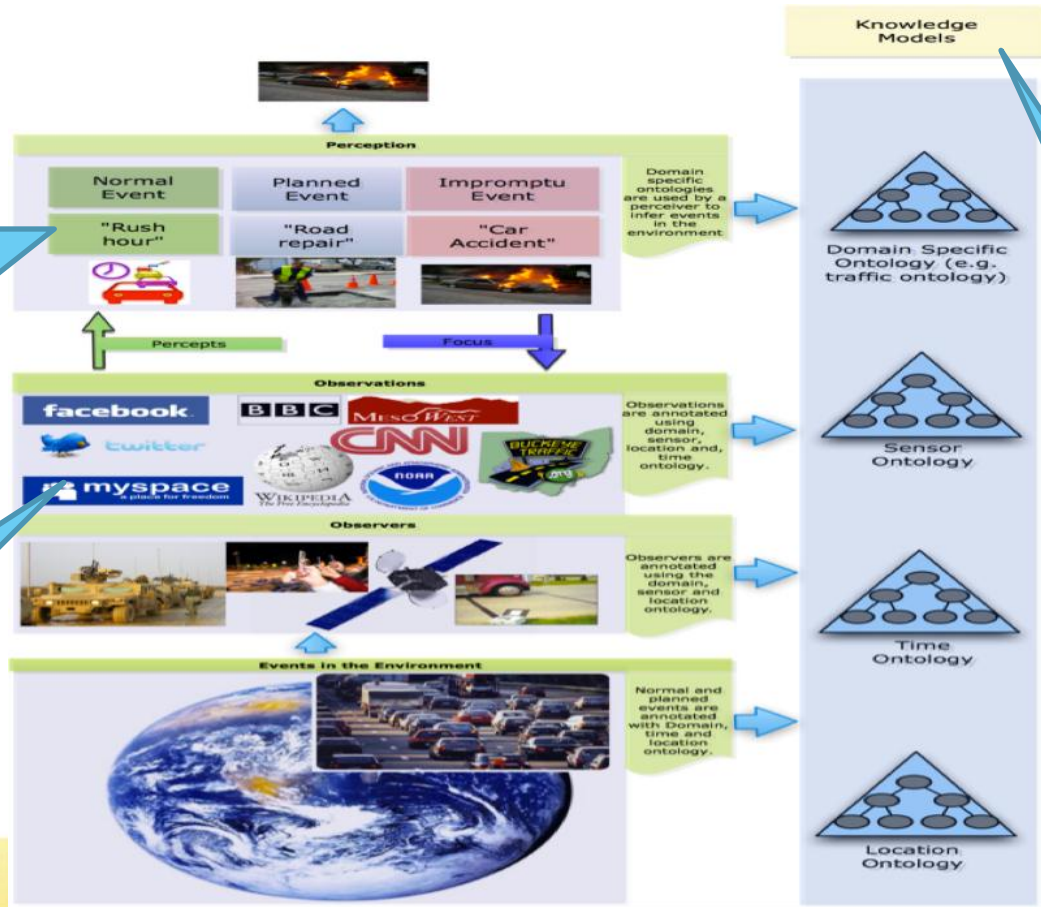


Sensor data: 10 passing cars per minute  
What might be the reason?



# Integration and Abstraction of Traffic Data

Knowledge Model Empowered Abstraction



Knowledge Models

Different types of Data

# Take Home Message

Amount of citizen sensing (and machine sensing) data is ***huge, varied,*** and ***growing rapidly.*** Search and Sift won't work.

# Take Home Message (Cont.)

Semantics play a **key** role in referring "***meaning***" behind the data. Requires progress from ***keywords*** -> ***entities*** -> ***relationships*** -> ***events***, from raw data to human-centric ***abstractions***.



# Take Home Message (Cont.)

Wide variety of semantic models and KBs (vocabularies, social dictionaries, community created semi-structured knowledge, domain-specific datasets, ontologies) empower semantic solutions. This can lead to ***Semantic Scalability*** – scalability that is meaningful to human activities and decision making.

# Interested in more?

[Kno.e.sis Wiki for the following and more:](#)

- [Computing for Human Experience](#)
- [Continuous Semantics to Analyze Real-Time Data](#)
- [Semantic Modeling for Cloud Computing](#)
- [Citizen Sensing, Social Signals, and Enriching Human Experience](#)
- [Semantics-Empowered Social Computing](#)
- [Semantic Sensor Web](#)
- [Traveling the Semantic Web through Space, Theme and Time](#)
- [Relationship Web: Blazing Semantic Trails between Web Resources](#)
- [SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups](#)
- [Semantically Annotating a Web Service](#)

**Tutorial:** [Citizen Sensor Data Mining, Social Media Analytics and Development Centric Web Applications \(WWW2011\)](#)

**Partial Funding:** NSF ([Semantic Discovery: IIS: 071441](#), [Spatio Temporal Thematic: IIS-0842129](#)), AFRL and DAGSI ([Semantic Sensor Web](#)), Microsoft Research ([Semantic Search](#)) and IBM Research ([Analysis of Social Media Content](#)), and HP Research ([Knowledge Extraction from Community-Generated Content](#)).

Microsoft Research

# Faculty Summit



FUTURE WORLD

2011 ← → 2031