

# Introduction

Jason Baldridge  
The University of Texas at Austin  
Language and Computers





# The machine is us/ing us

---

- [http://www.youtube.com/watch?v=NLI GyptXT\\_g](http://www.youtube.com/watch?v=NLI GyptXT_g)

# Word clouds: documents as bags of words (<http://www.wordle.net/>)



# New York Times Inaugural (Bag of) Words Visualization



1789  
George Washington

[Full text of the address](#)

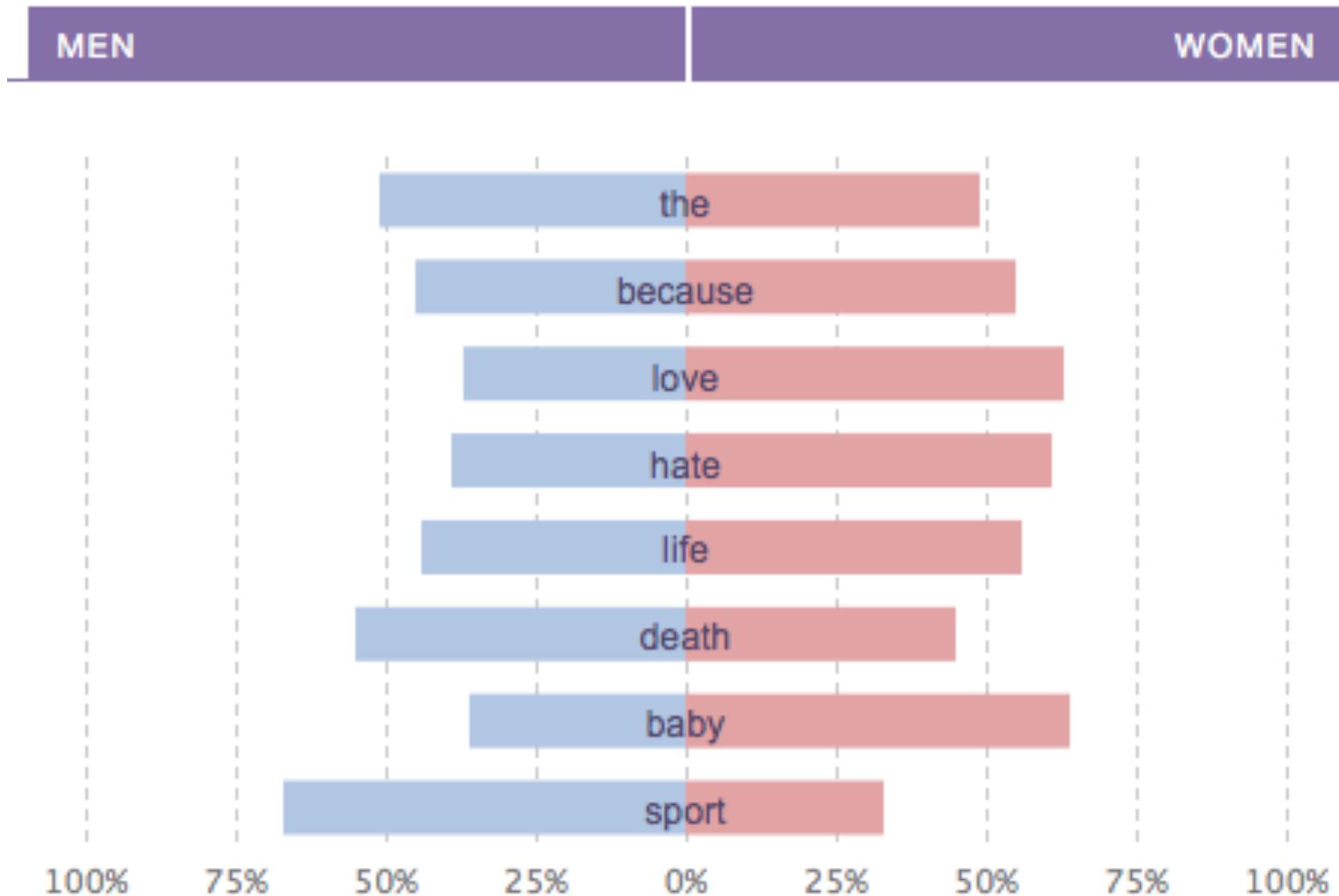
In the first inaugural speech, George Washington emphasized his duty to the newly formed nation and its citizens. He had been called to the position much as he had been called to serve in the Revolutionary War. He would obey the "public summons" to serve the "public good": "The voice of my country called me."

government public country  
present citizen duty unite hand people  
nature circumstance great liberty nation happiness fellow  
measure establish power American hope right states union free pecuniary  
peculiarly humble recommendation call policy day particular decide sentiment executive  
administration article render human system decline private qualification future confidence  
success subject expedient trust require dare heaven share blessing object judge influence  
view character former house station event experience voice light love impression

[http://www.nytimes.com/interactive/2009/01/17/washington/20090117\\_ADDRESSES.html](http://www.nytimes.com/interactive/2009/01/17/washington/20090117_ADDRESSES.html)

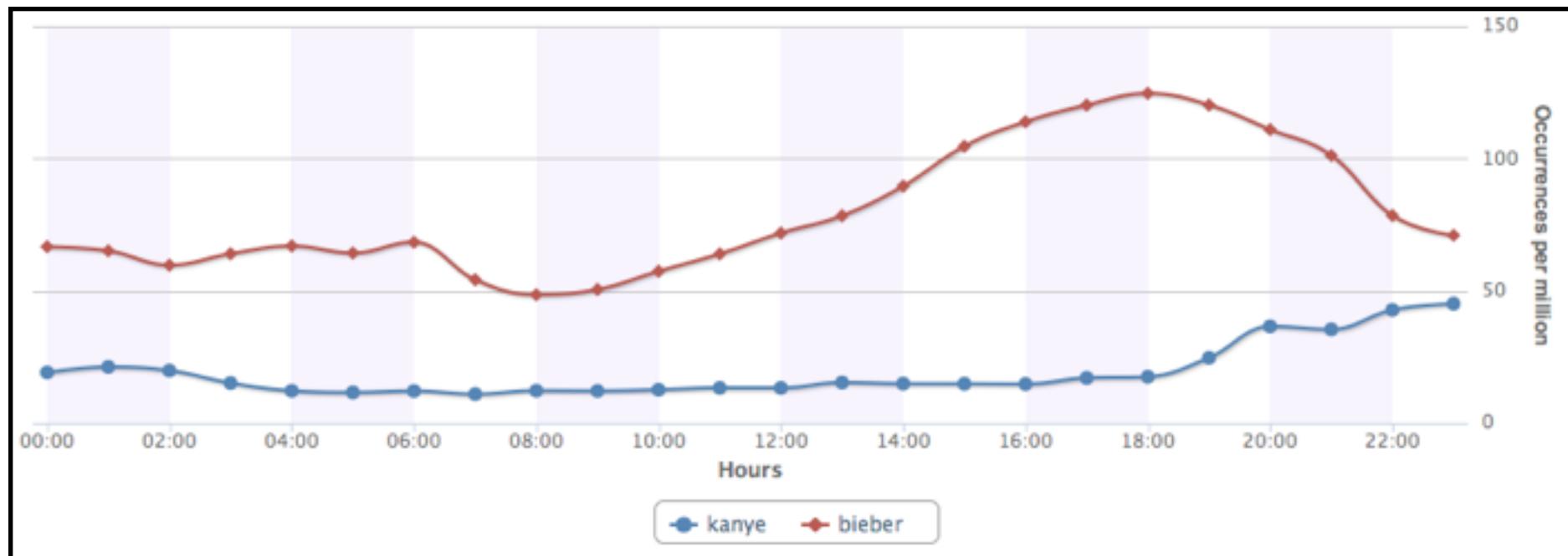
# He says, she says

<http://www.tweetolife.com/gender/>



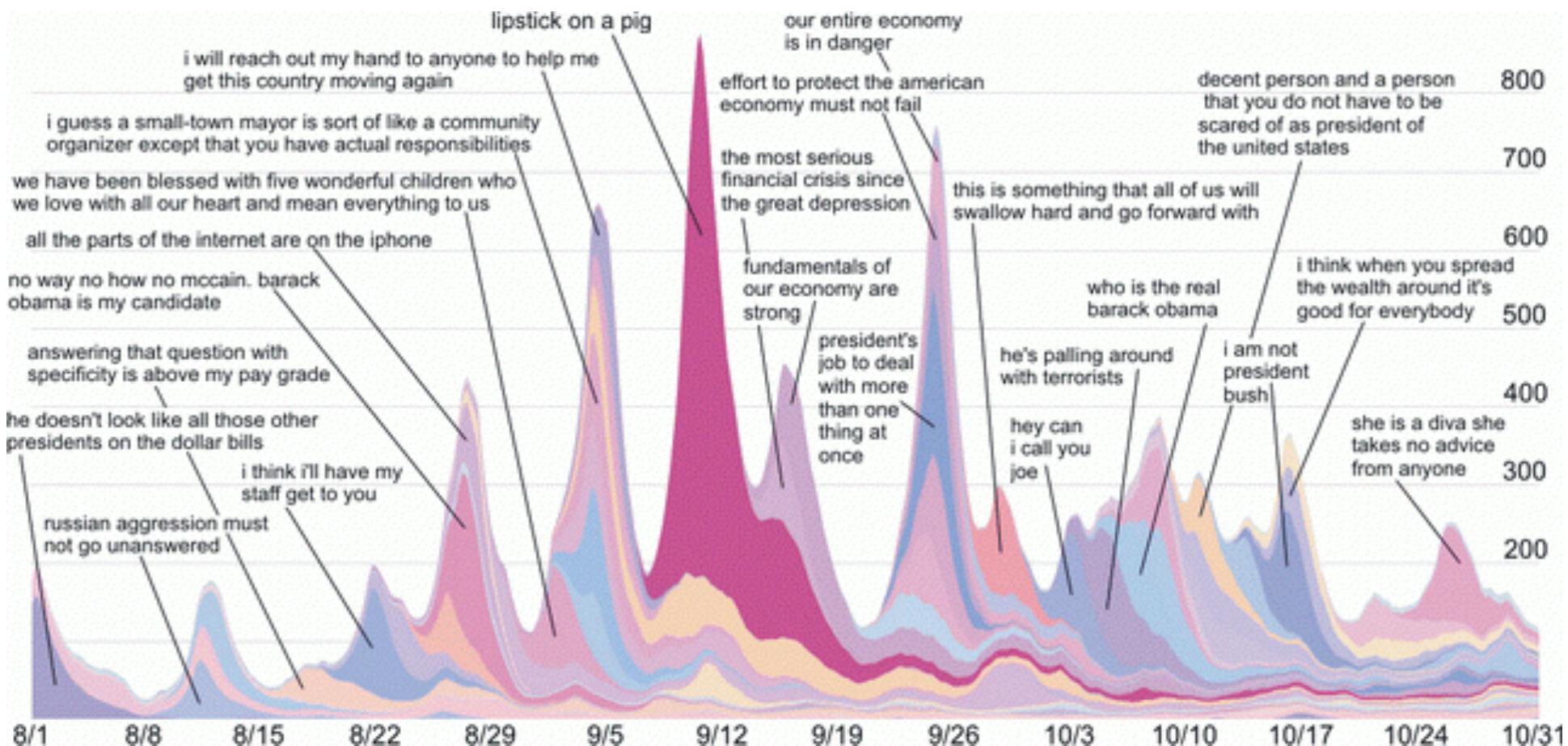
# Temporality of words, by hour

<http://www.tweetolife.com/hour/>

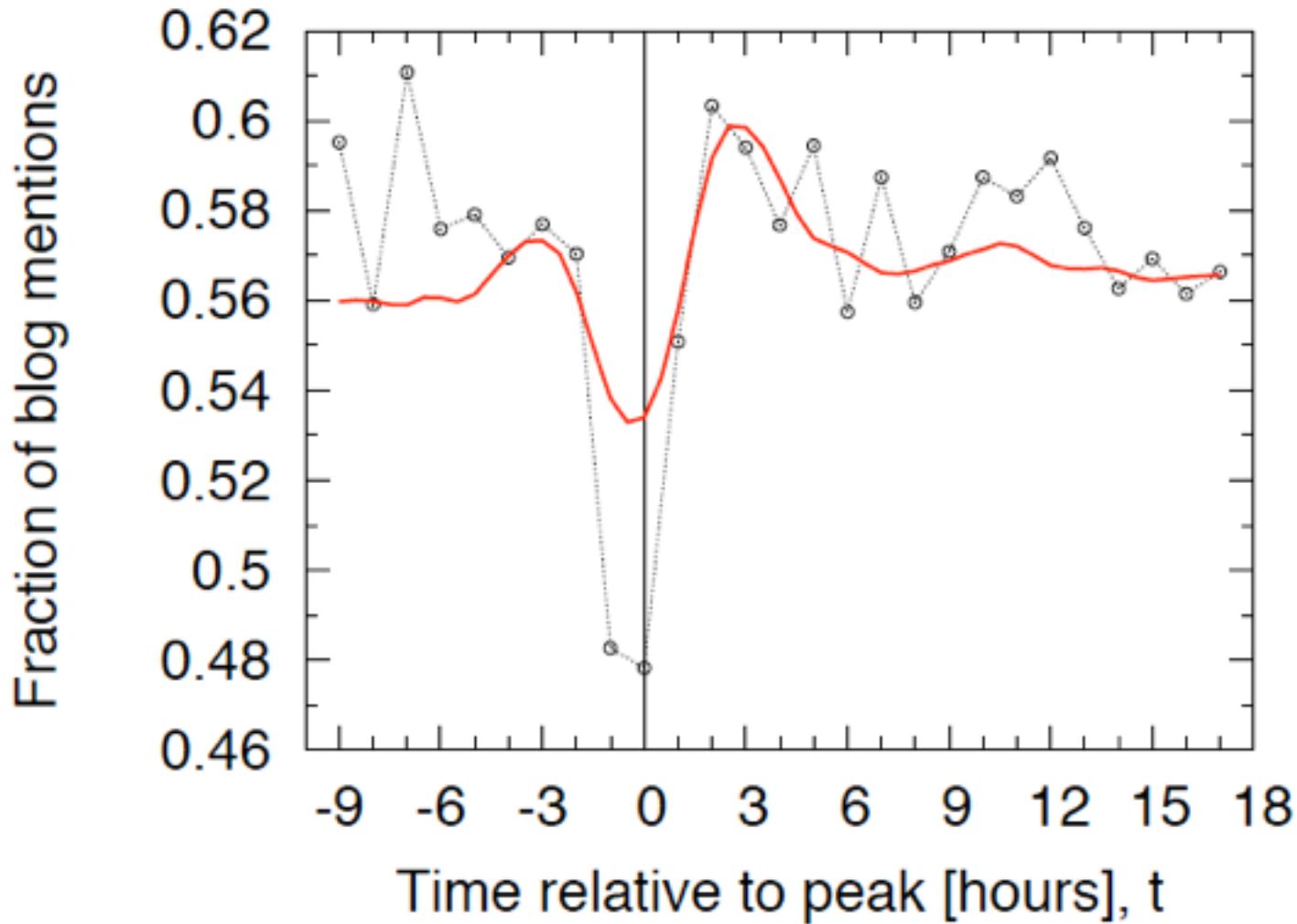


# Meme me up

<http://memetracker.org/>



# The pulse of the media



# Visualizing discussions: Wikipedia deletions [<http://notabilia.net/>]





# Analysis of debates

Shift

Speaker Cc Transcript

IFILL  
PALIN  
BIDEN

36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

Computational model identifies turn 48 in the debate as clearly indicating topic control.

47 IFILL: Governor Palin, is that so?

PALIN: That is not so, but because that's just a quick answer, I want to talk about, again, my record on energy versus your ticket's energy ticket, also. I think that this is important to come back to, with that energy policy plan again that was voted for in '05. When we talk about energy, we have to consider the need to do all that we can to allow this nation to become energy independent. It's a nonsensical position that we are in when we have domestic supplies of energy all over this great land. And East Coast politicians who don't allow energy-producing states like Alaska to produce these, to tap into them, and instead we're relying on foreign countries to produce for us. We're circulating about \$700 billion a year into foreign countries, some who do not like America -- they certainly don't have our best interests at heart -- instead of those dollars circulating here, creating tens of thousands of jobs and allowing domestic supplies of energy to be tapped into and start flowing into these very, very hungry markets. Energy independence is the key to this nation's future, to our economic future, and to our national security. So when we talk about energy plans, it's not just about who got a tax break and who didn't. And we're not giving oil companies tax breaks, but it's about a heck of a lot more than that. Energy independence is the key to America's future.

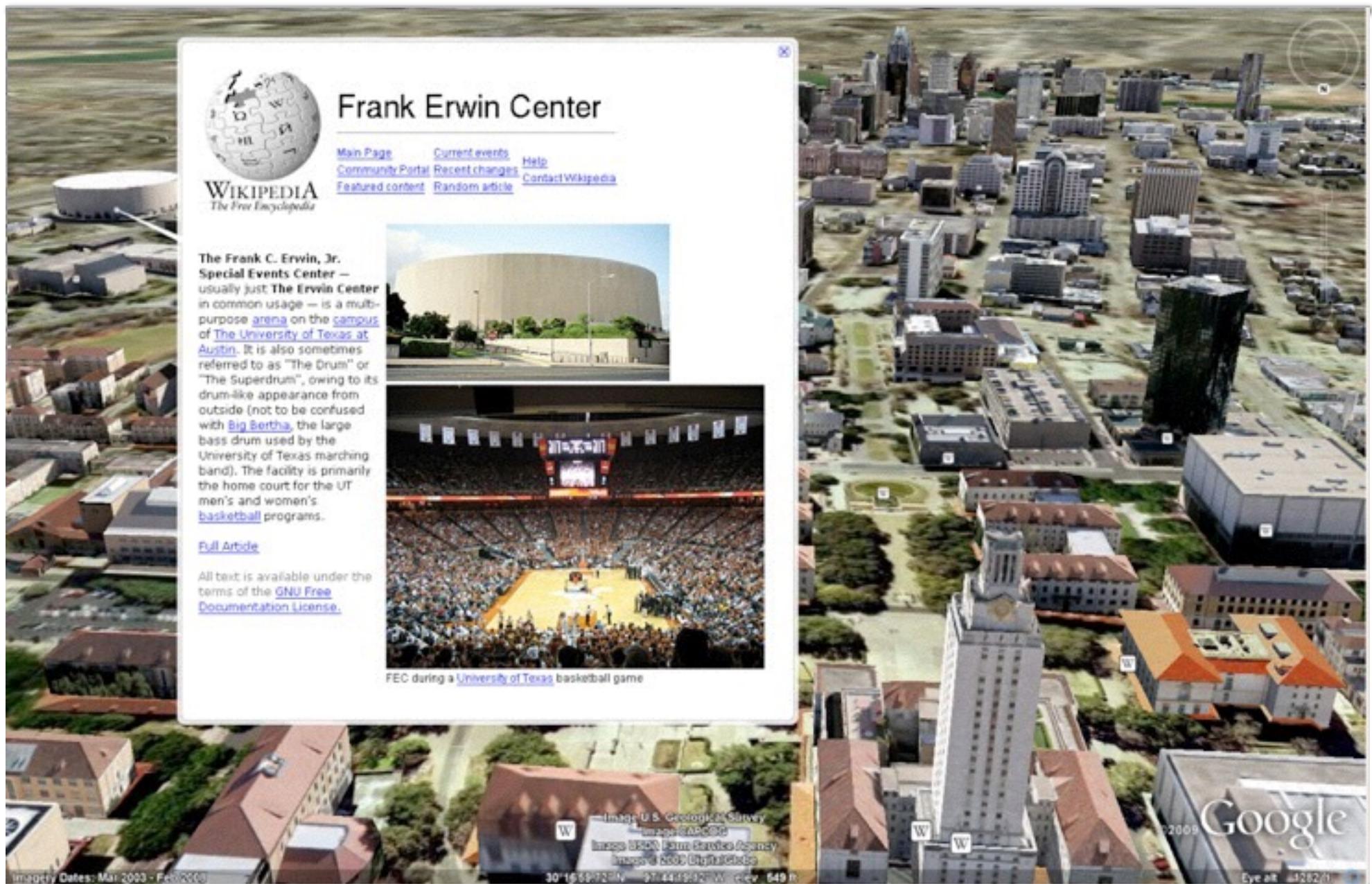
48 IFILL: Governor, I'm happy to talk to you in this next section about energy issues. Let's talk about climate change. What is true and what is false about what we have heard, read, discussed, debated about the causes of climate change?

Topic Cloud

oil street economy mccain people wall washington home	tax taxes jobs billion cut america percent business	people vice senate party american president record maverick	energy oil nuclear clean climate coal alternative power	country care health thank america times eight policies
trade free agreement country south priorities families agreements	campaign people running talking american campaigns tough congressman	pakistan afghanistan troops support government qaida bin taliban	voted support rights senate issue constitution roe wade	security judgment national world wrong nation iraq issues
genocide security darfur tell bosnia marines national force	question tonight thank senator presidential debate mccain obama	nuclear iran israel weapons si north administration ahmadinejad	senator mccain obama president time united people american	iraq afghanistan troops strategy war military surge obama
john mccain power	health care insurance plan	education school college	russia russia russians	spending billion environment



# Georeferenced content in Wikipedia



The image shows a composite view of a city's skyline from an aerial perspective. A green overlay covers the entire image, representing georeferenced content. In the top left corner, there is a screenshot of a Wikipedia page for the "Frank Erwin Center". The page includes the Wikipedia logo, a navigation bar with links like "Main Page", "Current events", "Community Portal", "Recent changes", "Help", "Featured content", "Random article", and "Contact Wikipedia". The main content area describes the Frank C. Erwin, Jr. Special Events Center, also known as "The Erwin Center", as a multi-purpose arena on the campus of The University of Texas at Austin. It is described as having a drum-like appearance from outside. Below the text are two images: one of the exterior of the arena and another of the interior during a basketball game. At the bottom of the Wikipedia screenshot, there is a copyright notice: "All text is available under the terms of the [GNU Free Documentation License](#)". The bottom right corner of the image features a "Google" watermark.

Over 1 million coordinate links have been annotated.

# Georeferenced Twitter



storm since:2009-08-26 near:78712 wti

Realtime results for **storm since:2009-08-26**  
**near:78712 within:15mi**



[nannermagrill](#): just danced in the rain; now drying off on the porch and watch the **storm**. the PERFECT birthday gift!!!!  
Austin, TX  
Aug 27, 2009 02:31 AM GMT · from TweetDeck · [Reply](#) · [View Tweet](#)



[katemorris](#): To South Austin PPL, get ready, this **storm** is rough but we need rain  
Austin, TX  
Aug 27, 2009 02:31 AM GMT · from Seesmic · [Reply](#) · [View Tweet](#)



[jameswheadrick](#): And of course the **storm** dies as soon as I set up my cam  
Austin, TX  
Aug 27, 2009 02:29 AM GMT · from txt · [Reply](#) · [View Tweet](#)



[crashheatrice](#): @Buchanan007 the **storm** coming in is bringing cold weather  
iPhone: 30.326029,-97.729286  
Aug 27, 2009 02:20 AM GMT · from TwitterFan · [Reply](#) · [View Tweet](#) · [Show Conversation](#)



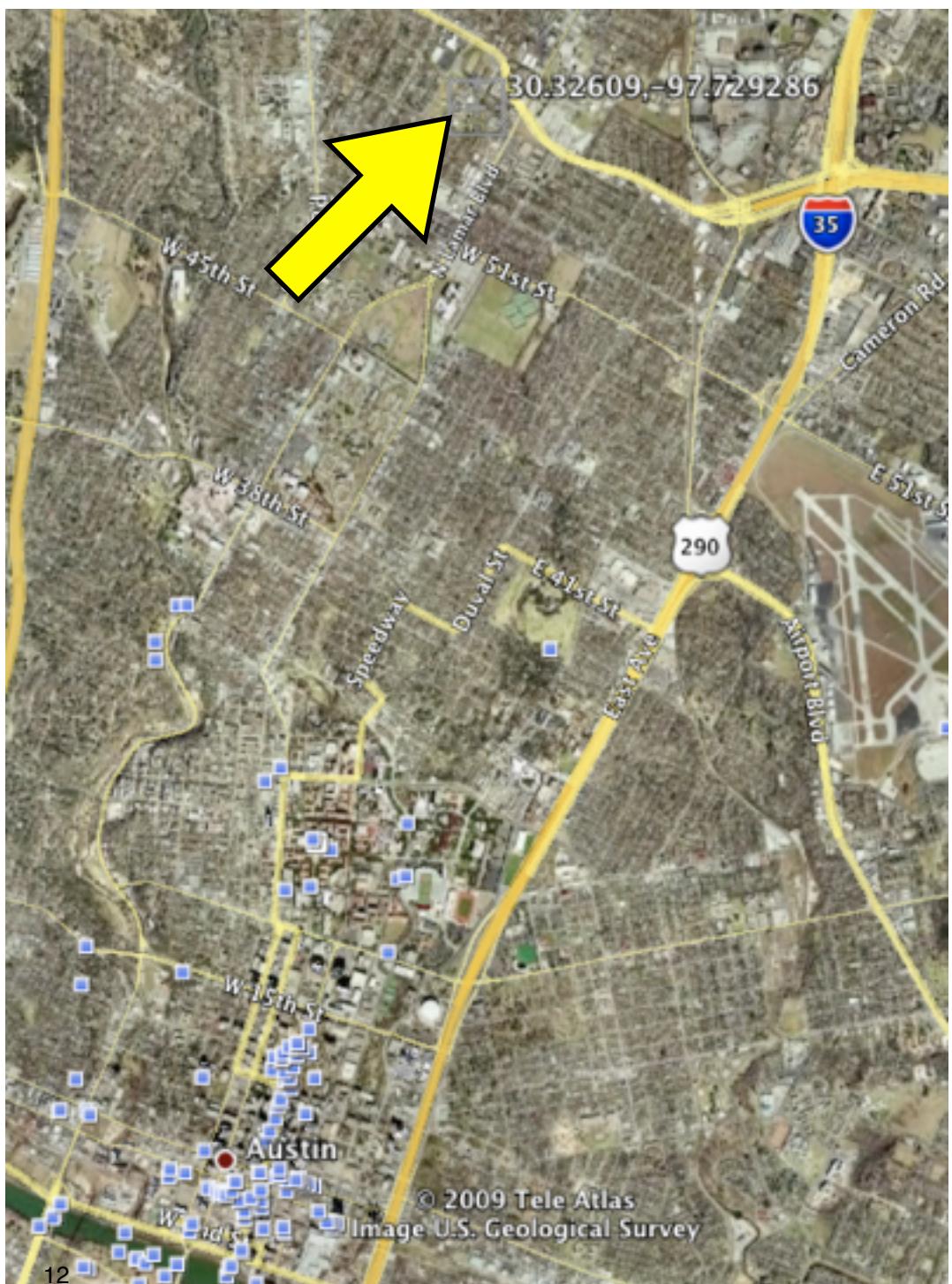
[sgharms](#): At the showdown, er, little woodrow's on Guadalupe listening to the approaching **storm**...and weezer.  
Austin, Tejas  
Aug 27, 2009 02:26 AM GMT · from Tweetie · [Reply](#) · [View Tweet](#)



[locustfist](#): the **storm** is just skirting by avery ranch. cool lightning though.  
Austin, TX  
Aug 27, 2009 02:26 AM GMT · from TweetDeck · [Reply](#) · [View Tweet](#)



[Chippah56](#): The rain **storm** knock'd out my power





u/hrdi/



WIKIPEDIA  
*The Free Encyclopedia*

## Hôtel des Mille Collines

[Main Page](#)  
[Community Portal](#)  
[Featured content](#)

[Current events](#)  
[Recent changes](#)  
[Random article](#)

[Help](#) [Contact Wikipedia](#)

The **Hôtel des Mille Collines** is a large hotel in [Kigali](#), the capital city of [Rwanda](#). It became famous as the building in which more than a thousand people took refuge during the [Rwandan Genocide](#) of 1994. The story of the hotel and its manager at that time, [Paul Rusesabagina](#), was used as the basis of the film *[Hotel Rwanda](#)*.

## Full Article

All text is available under the terms of the [GNU Free Documentation License](#).



## The front of the hotel

Kicukiro, Kigali, Rwanda



00:01:36		Baratuzana abandi twari kumwe babicira hano ndumva bararengaga barindwi. Bamaze kubica ubwo njye nahise mbacika, nyura hagati y'amazu hirya aha ngera ku muhanda wa hano hepfo w'igitaka Ubwo interaharmwe zarengaga nk'icum, zahise zimbohoza zimaze kumbohoza zirandarana buracya ndagenda nsubira hepfo y'aho twabaga.
00:02:11	Yves	Ahangaha mwatubwiye ko haguye abantu barenga barindwi.
00:02:15		Haguye abantu barenga barindwi.



# Crisis response: Haiti earthquake

ID: 4025

TITLE: Water needed f

DATE: 2010-01-28 22:38

LOCATION: Rte Delm

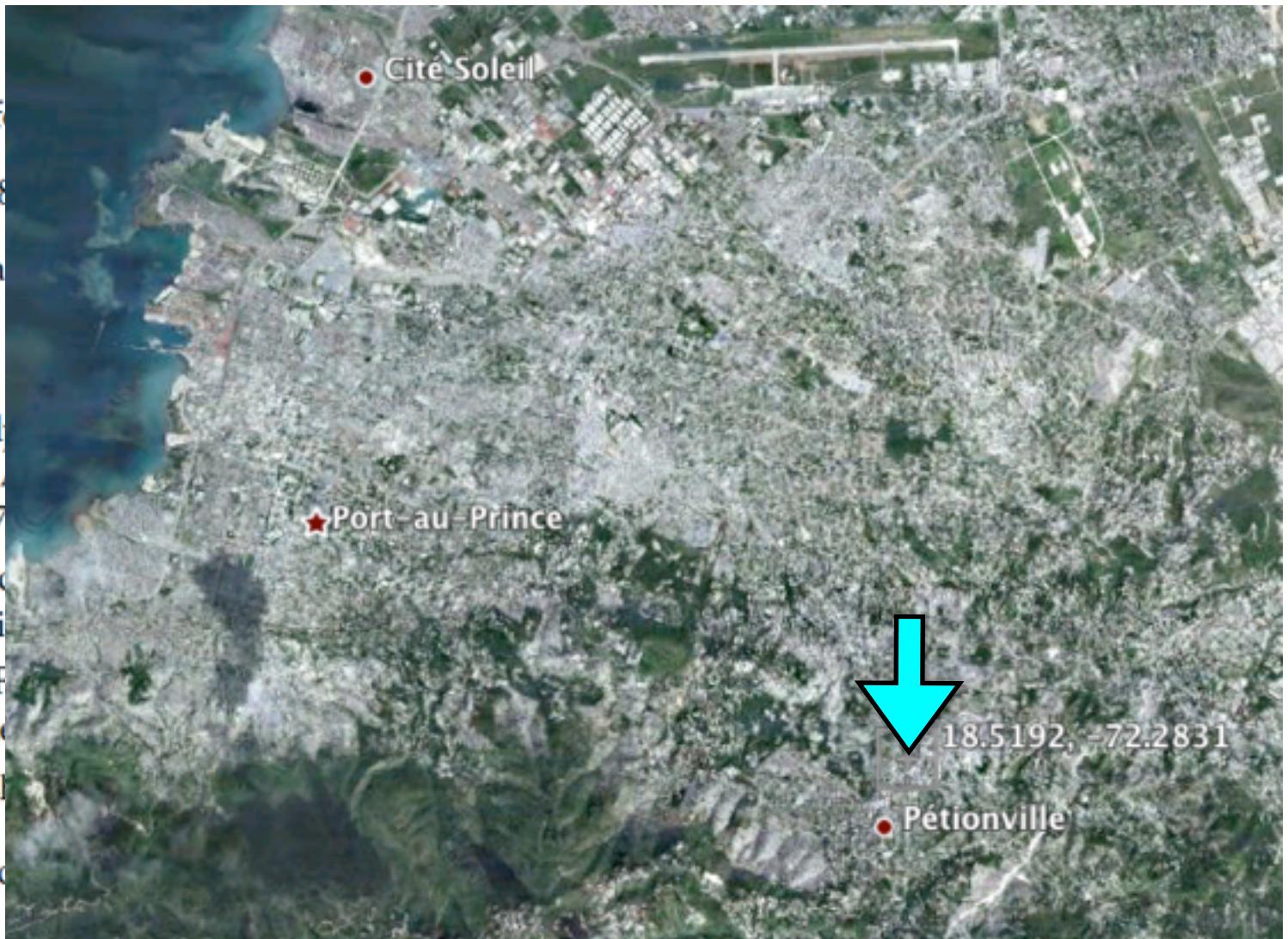
## DESCRIPTION:

- *English*: Need help water and food. might be number 27
- *Creole*: Jai besoin d'eau et de la nourriture.ais
- *Comments*: In the Pétionville area but she doesn't mention her name
- *Ushahidi ID*: IDUs

CATEGORY: 2. Urgenc

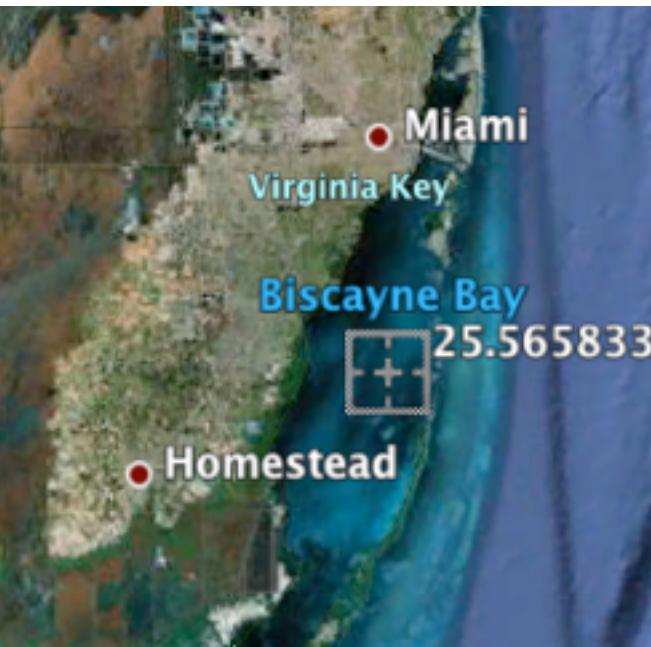
LATITUDE: 18.5192

LONGITUDE: -72.2831





# On a lighter note... they're on a boat!



Oh sh\*t, get your towels ready  
It's about to go down  
Everybody in the place hit the f\*cking deck  
But stay on your motherf\*cking toes  
We running this, let's go

I'm on a boat, I'm on a boat  
Everybody look at me  
'Cause I'm sailing on a boat  
I'm on a boat, I'm on a boat  
Take a good hard look  
At the motherf\*cking boat

I'm on a boat motherf\*cker, take a look at me  
Straight flowing on a boat on the deep blue sea  
Busting five knots, wind whipping out my coat  
You can't stop me motherf\*cker, 'cause I'm on a boat

Take a picture, trick, I'm on a boat, bitch  
We drinking Santana champ 'cause it's so crisp  
I got my swim trunks and my flippie-floppies  
I'm flipping burgers, you at Kinko's  
Straight flipping copies

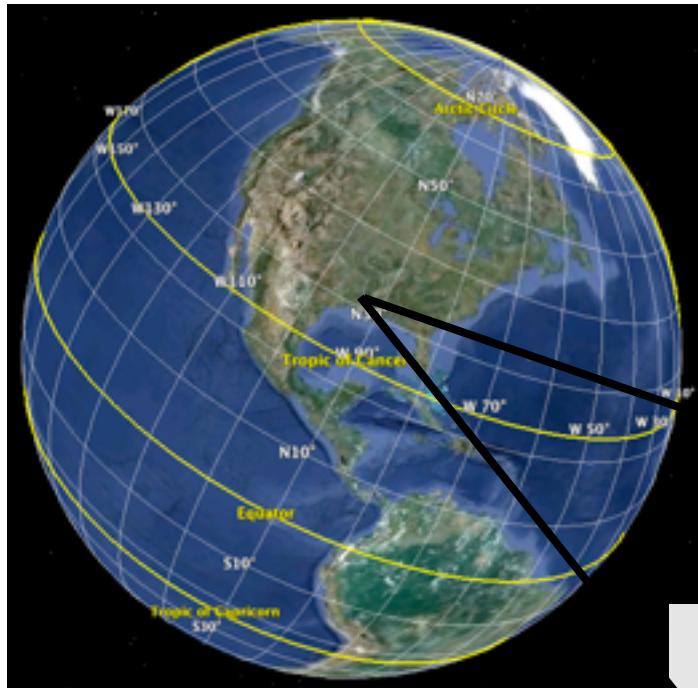
I'm riding on a dolphin, doing flips and sh\*t  
This dolphin's splashing, getting everybody all wet  
But this ain't Seaworld, this is real as it gets  
I'm on a boat, motherf\*cker, don't you ever forget

towels  
deck  
motherf\*cking  
boat  
sailing  
sea  
knots  
bitch  
swim  
Kinko's  
dolphin  
Seaworld

# Where's a word on Earth?



# Automatic geolocation



**twitter** storm since:2009-08-26 near:78712 wti Search

Realtime results for **storm** since:2009-08-26 near:78712 within:15mi 0.33 seconds

**nannermagirl:** just danced in the rain; now drying off on the porch and watching the **storm**. the PERFECT birthday gift!!!!  
Austin, TX  
Aug 27, 2009 02:31 AM GMT - from TweetDeck · [Reply](#) · [View Tweet](#)

**katemorris:** To South Austin PPL, get ready, this **storm** is rough but we needed the rain  
Austin, TX  
Aug 27, 2009 02:31 AM GMT - from Seismic · [Reply](#) · [View Tweet](#)

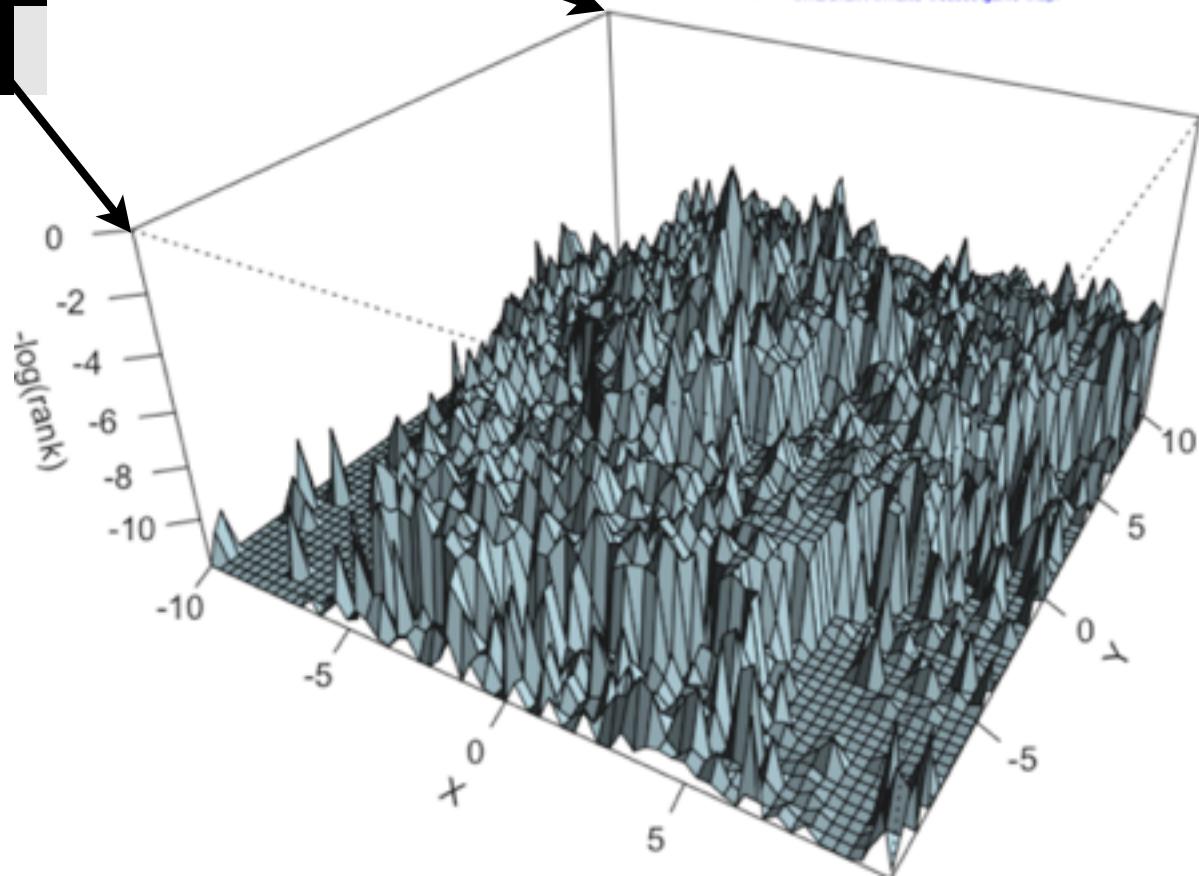
**jameswheadrick:** And of course the **storm** dies as soon as I set up my camera  
Austin, TX  
Aug 27, 2009 02:29 AM GMT - from txt · [Reply](#) · [View Tweet](#)

**crashantiqoxx:** @Buchanan007 the **storm** coming in is bringing cold weather  
iPhone: 30.320028,-97.729286  
Aug 27, 2009 02:28 AM GMT - from TwitterFor · [Reply](#) · [View Tweet](#) · [Show Conversation](#)

**sqharm:** At the showdown, er, little woodrow's on Guadalupe listening to the approaching **storm**...and weezier.  
Austin, Tejas  
Aug 27, 2009 02:26 AM GMT - from Tweeter · [Reply](#) · [View Tweet](#)

**locustifist:** the **storm** is just skirting by every ranch. cool lightning though.  
Austin, TX  
Aug 27, 2009 02:26 AM GMT - from TweetDeck · [Reply](#) · [View Tweet](#)

**Chippah56:** The rain **storm** knock'd out my power



Philip Flip Kromer @mflip 6h  
I'm at Opal Divine's Freehouse (Austin, TX) w/ 2 others  
[4sq.com/yKfbTB](http://4sq.com/yKfbTB) · [View media](#)

Philip Flip Kromer @mflip 6h  
Heading out w/ @bill\_michels and @tbell to Data Drinks at Opal Divine's 700 W 6th [bit.ly/datadrinks12](http://bit.ly/datadrinks12) - No free booze, just smart geeks

Philip Flip Kromer @mflip 6h  
Big data meetup - Hilton 6th floor [bit.ly/hilton\\_austin\\_6th\\_floor\\_meeting\\_rm\\_A](http://bit.ly/hilton_austin_6th_floor_meeting_rm_A) [4sq.com/xkyOIZ](http://4sq.com/xkyOIZ) · [View media](#)

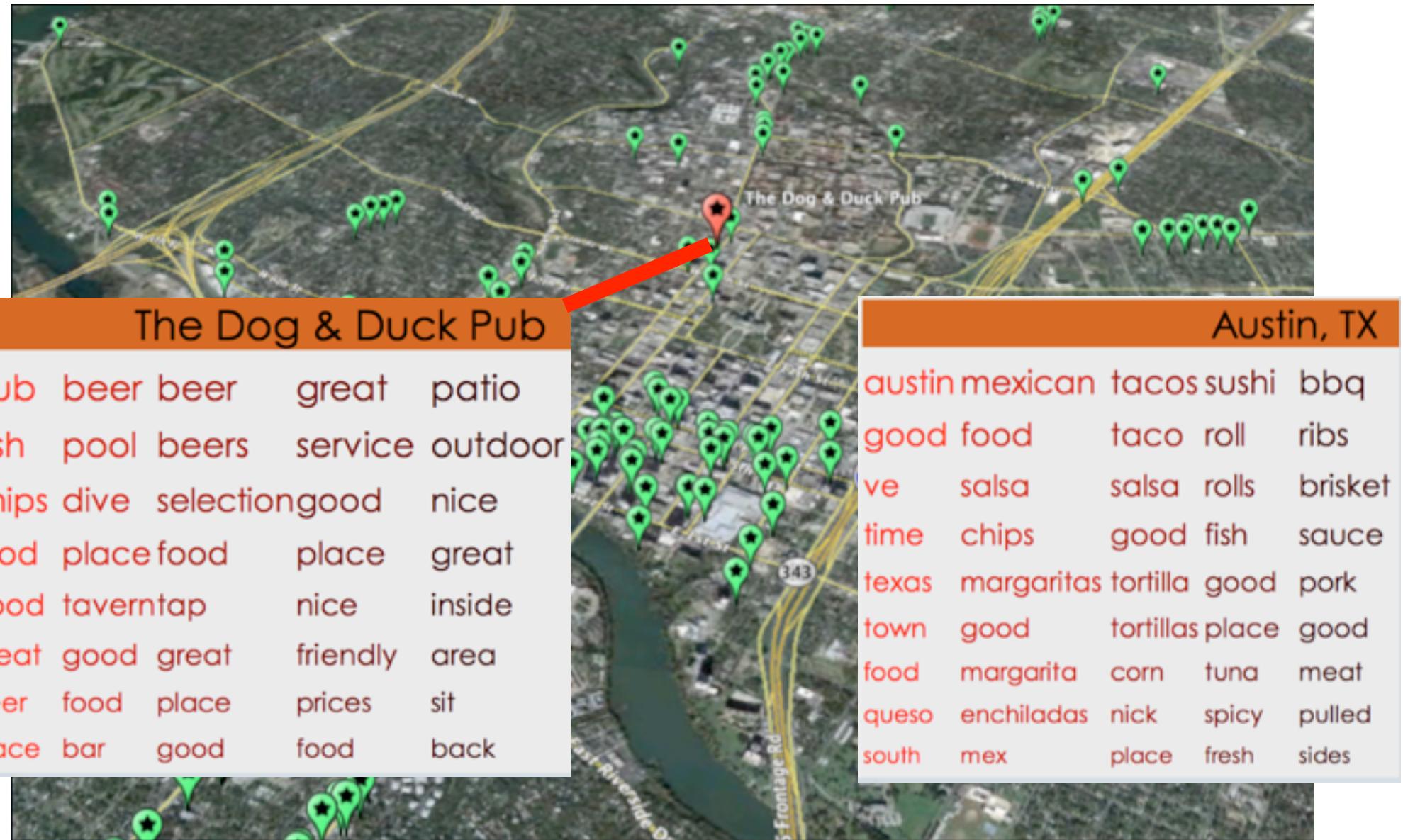
Philip Flip Kromer @mflip 10h  
@Hippolyta @baratunde not to worry, we're concentrated on making humans smarter - Human + machine way beats machine: [collisiondetection.net/marchives/20...](http://collisiondetection.net/marchives/20...) · [In reply to Hippolyta](#)

Damon Clinkcales @damon 11h  
Secret [tip](#) for coffee lovers.... [@HoundstoothATX](#) is downtown on sixth street this week. Thank you [@SoundCloud](#)! You chose very well.  
13 Retweeted by Philip Flip Kromer

Philip Flip Kromer @mflip 11h  
@skellan 630 Opal Divine's data drinks? /cc @antgoldbloom  
@manhalki @matto @caseorganic @Qpi



# Combining georeferencing with word clouds



2010 project by Andy Luong (UT CS undergrad): topic modeling and visualization of 13,000 restaurants in 42 North American cities based on their Yelp reviews.

# What does this sentence mean?



[<http://www.clker.com/clipart-green-eyes-3.html>]

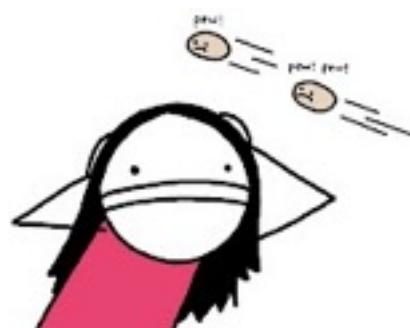
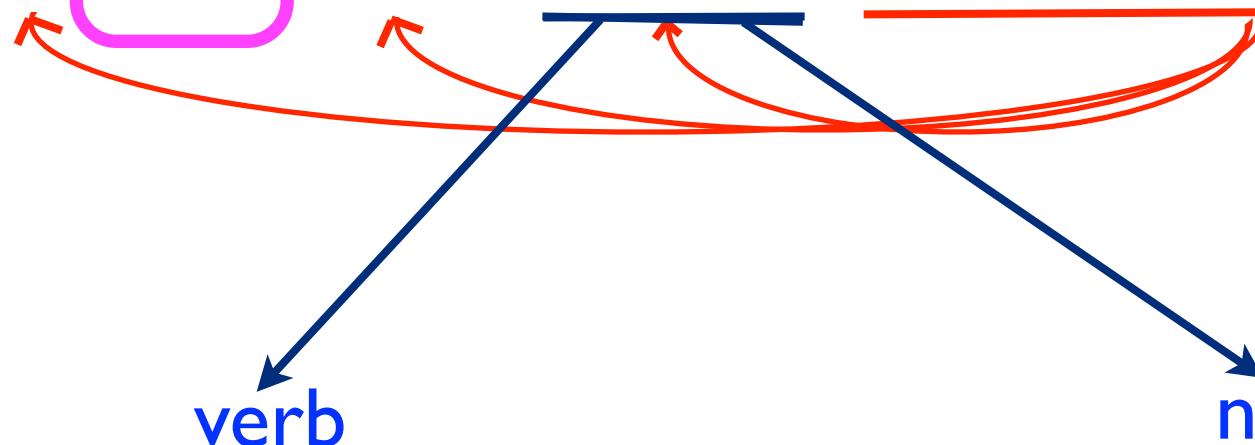


[<http://www.clker.com/clipart-3163.html>]

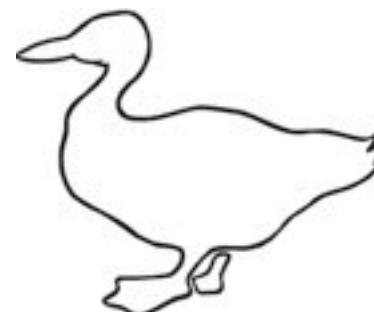


[<http://www.simonpalfrader.com/category/tournament-poker>]

I **saw** her **duck** with a **telescope**.

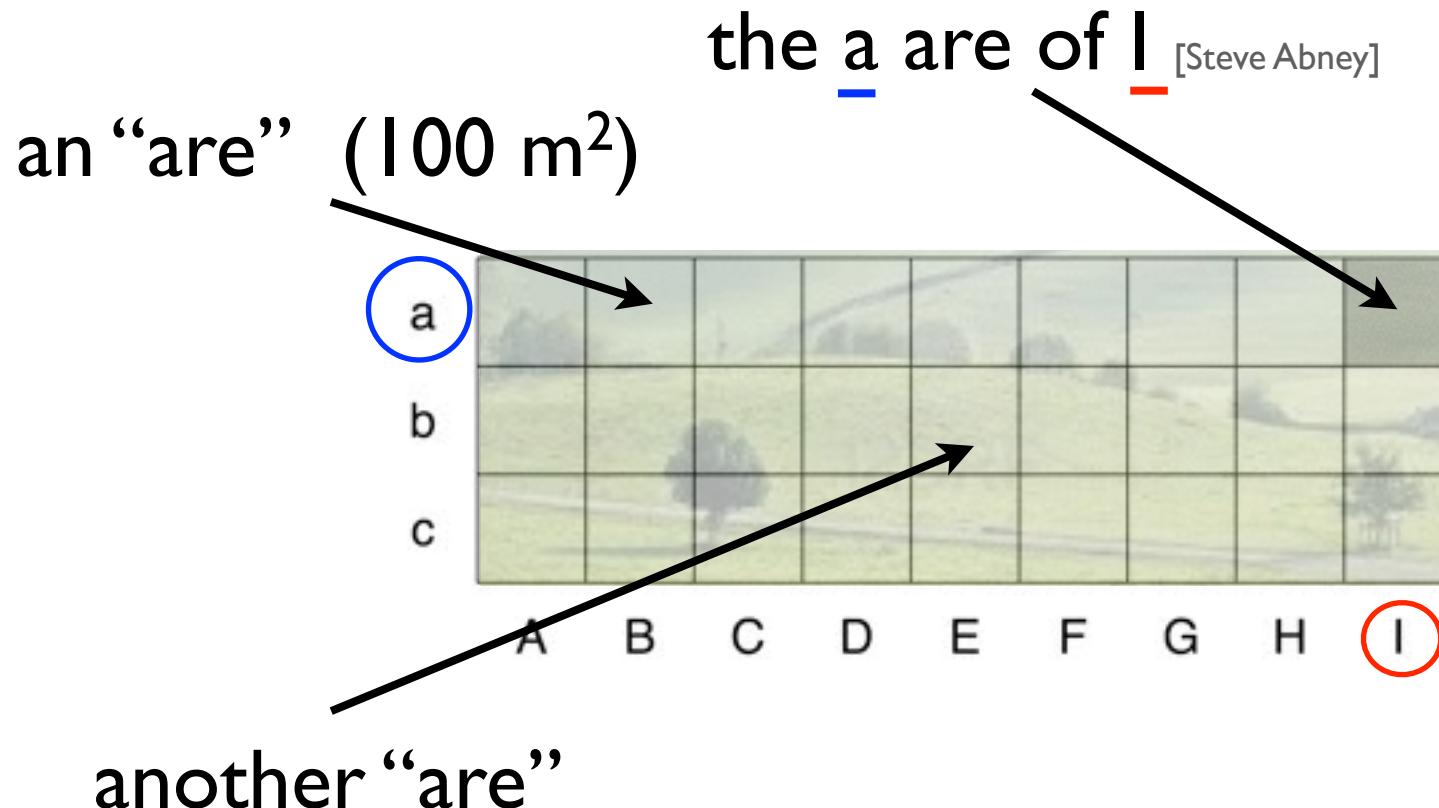


[<http://casablancapa.blogspot.com/2010/05/fore.htm>]



[<http://www.supercoloring.com/pages/duck-outline/>]

# Ambiguity is pervasive





# Watson



# Mistakes are consequential



## Causes of AIDS

Found 20 Causes from 116 records.

### Causes of AIDS (20)

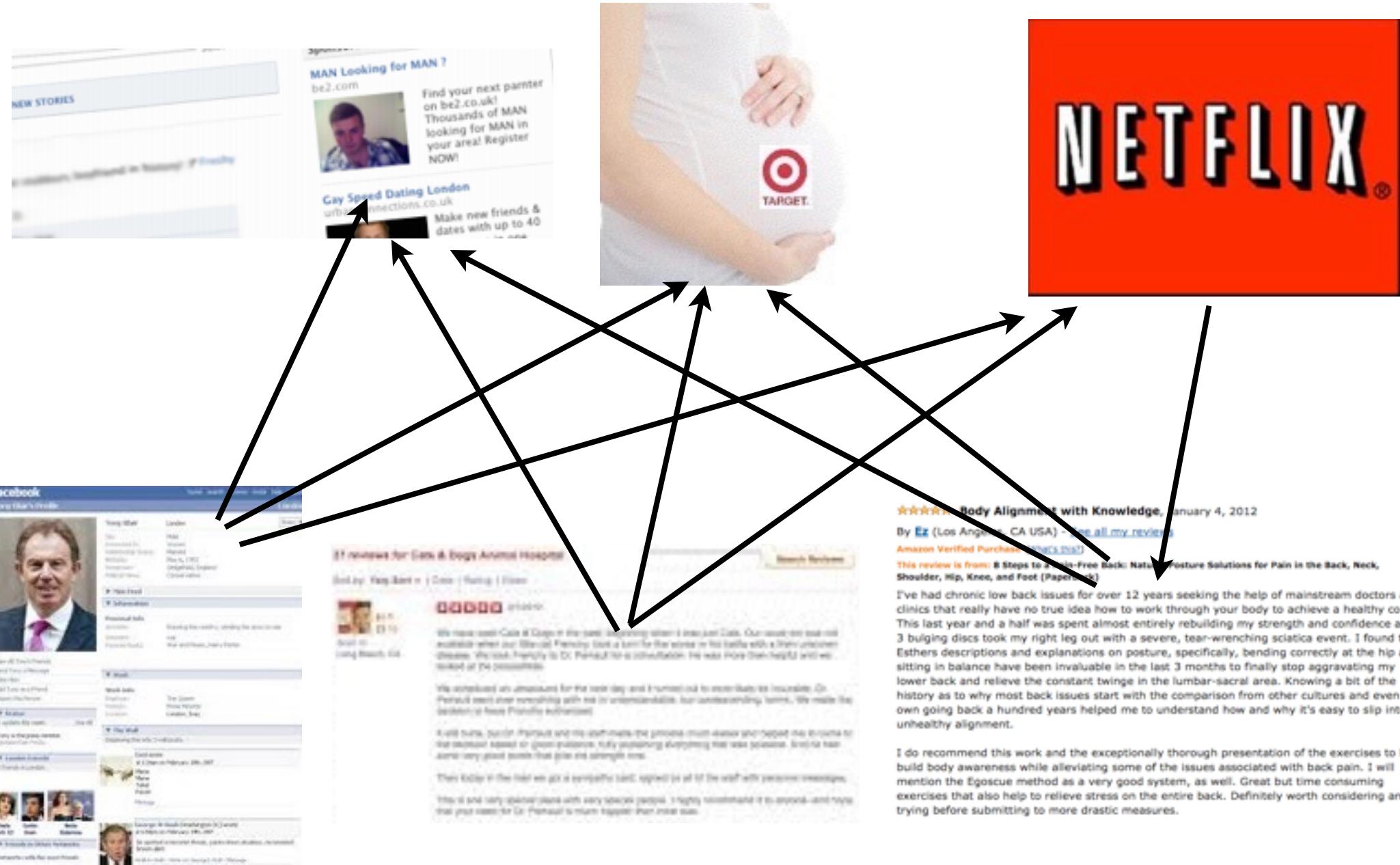
- + Contact infected partner
- + Sexual contact
- + Virus
- + Jew
- + Physiotherapy
- + Feature
- + Strong magnetic field
- + AUDIOVISUAL
- + Akhmed Zakayev
- + Bacille Calmette-Guerin

[View more Causes](#)

FAIL



# Being right can be consequential too



# People get language rong two

## "The Onion" Strikes Again!

[+ Add to Favorites](#)

[+ Add to My Site](#)



via

[REDACTED]  
funny facebook fails – "The Onion" Strikes Again!  
ghter machine? I hope it  
wasn't stolen from me by our well paid and highly educated politicians agin.  
Please forgive us Lord .



**Planned Parenthood Opens \$8 Billion Abortionplex**  
[www.theonion.com](http://www.theonion.com)

TOPEKA, KS—Planned Parenthood announced Tuesday the grand opening of its long-planned \$8 billion Abortionplex, a sprawling abortion facility that will allow the organization to terminate unborn lives with an efficiency never before thought possible....



Planned Parenthood's Auschwitz. Who would have thought!



**Planned Parenthood Opens \$8 Billion Abortionplex**  
[www.theonion.com](http://www.theonion.com)

TOPEKA, KS—Planned Parenthood announced Tuesday the grand opening of its long-planned \$8 billion Abortionplex, a sprawling abortion facility that will allow the organization to terminate unborn lives with an efficiency never before thought possible....



An 8 BILLION DOLLAR ABORTIONPLEX??? This has to be stopped!!!!



**Planned Parenthood Opens \$8 Billion Abortionplex**  
[www.theonion.com](http://www.theonion.com)

TOPEKA, KS—Planned Parenthood announced Tuesday the grand opening of its long-planned \$8 billion Abortionplex, a sprawling abortion facility that will allow the organization to terminate unborn lives with an efficiency never before thought possible....



- COSY - Mobile Robot
- MIT Media labs - Ripley
- Microsoft Research - Situated Interaction



- Basic following behavior and supervised learning of





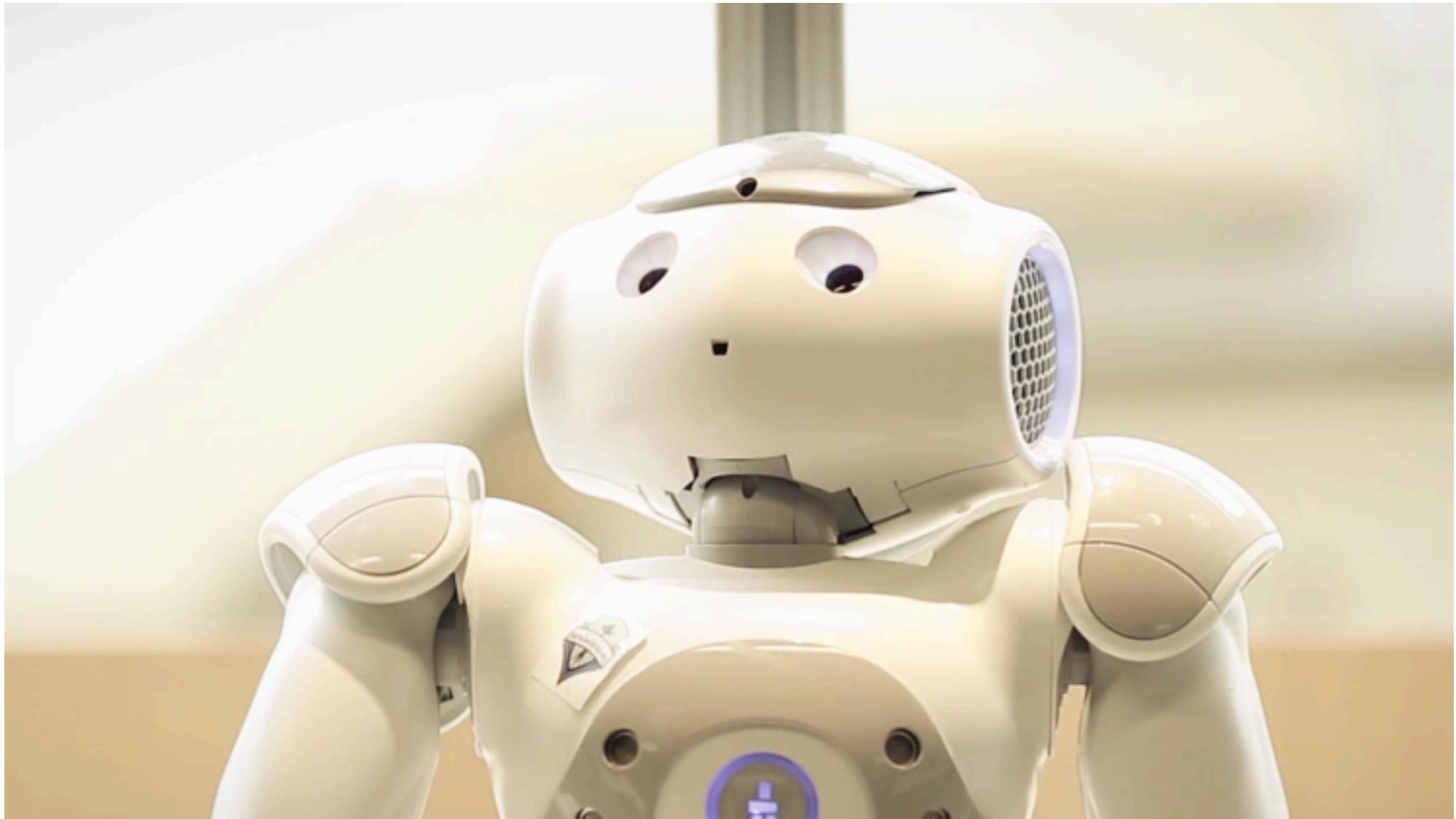
- Basic locative questions.





- When speech recognition fails...
  - [http://videolectures.net/aaai07\\_zender\\_htsn/](http://videolectures.net/aaai07_zender_htsn/)





<http://blogs.microsoft.com/firehose/2014/04/08/smart-elevator-robot-and-virtual-admin-will-help-guide-you-at-microsoft-research/>



- Instructor - Jason Baldridge
- TA - Hammal Al Bulushi
- Syllabus: [http://utcompling.github.io/  
language\\_and\\_computers\\_s16/syllabus](http://utcompling.github.io/language_and_computers_s16/syllabus)
- Schedule: [http://utcompling.github.io/  
language\\_and\\_computers\\_s16/schedule/](http://utcompling.github.io/language_and_computers_s16/schedule/)
- See the links page for links to materials in these slides  
and more:
- Canvas



- Who this course is geared toward and who it isn't.
  - This course is meant to be fun and topical, yet rigorous and specific.
- Other computational linguistics courses:
  - Introduction to Computational Linguistics
  - Natural Language Processing
- I welcome your input and feedback. (See something cool in the news? Tell me!)
- Questions?