

Natural Language Processing: Data, Algorithms, and Knowledge

BEARS 2011



Dan Klein

Computer Science Division

University of California, Berkeley



Language Technologies



Goal: Deep Understanding

- Requires context, linguistic structure, meanings...

Reality: Shallow Matching

- Requires robustness and scale
- Amazing successes, but fundamental limitations



- [Re: A camel is a horse designed by committee](#) **SR** 16/April/04
 - [Re: A camel is a horse designed by committee](#) **Henry** 18/April/04



Factoids and Limitations

[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#) [more »](#)

any US states' capitals are also their largest cities?

Search

Web

Your search - **How many US states' capitals are also their largest cities?** - did not match any documents.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

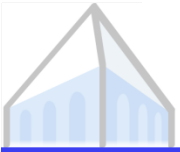
[Google Home](#) - [Business Solutions](#) - [About Google](#)

[capital of Wyoming: information from Answers.com](#)

Note: click on a word meaning below to see its connections and related words.
The noun **capital** of **Wyoming** has one meaning: Meaning #1 : the **capital**.
www.answers.com/topic/capital-of-wyoming - 21k - [Cached](#) - [Similar pages](#)

[Cheyenne: Weather and Much More From Answers.com](#)

Chey·enne (shī-ăn ' , -ěăn ') The **capital** of **Wyoming**, in the southeast part of the state near the Nebraska and Colorado borders.
www.answers.com/topic/cheyenne-wyoming - 74k - [Cached](#) - [Similar pages](#)



Distilling Information

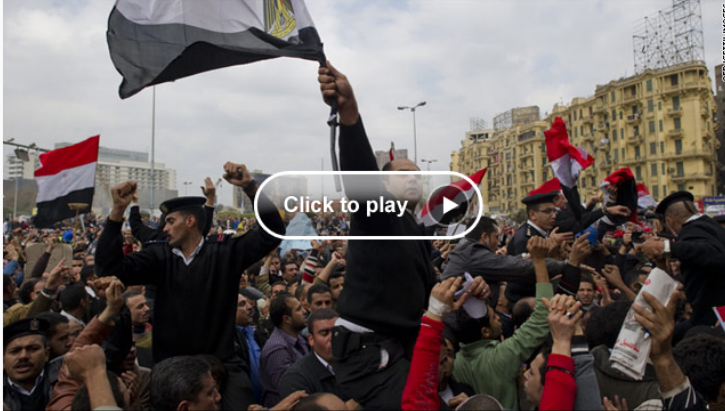
EDITION: U.S. | INTERNATIONAL | MÉXICO
Set edition preference

CNN World

Home Video NewsPulse U.S. **World** Politics Justice Entertainment Tech Health

Egypt's military dissolves Parliament, suspends constitution

By the CNN Wire Staff
February 13, 2011 2:44 p.m. EST



Click to play

STORY HIGHLIGHTS

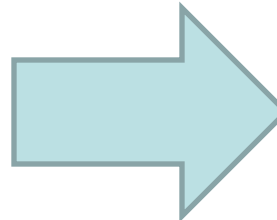
- **NEW:** Banks are shuttered until Wednesday as protests force top banker's resignation
- **NEW:** ElBaradei urges generals to "come out of their headquarters"
- **NEW:** Stock exchange to freeze transactions from officials being investigated
- Egypt's ambassador says the military will run a "technocratic" government until elections

Cairo, Egypt (CNN) -- Egypt's military dissolved the country's Parliament and suspended its constitution Sunday following the ouster of longtime leader Hosni Mubarak, telling Egyptians it would be in charge for six months or until elections can be held.

The Supreme Council of the Armed Forces said it would appoint a committee to propose changes to the Constitution, which would then be submitted to voters. The council will have the power to issue new laws during the transition period, according to a communique read on state television.

Sameh Shoukry, Egypt's ambassador to the United States, said Sunday that the generals have made restoring security and reviving the economy its top priorities.

"This current composition is basically a technocratic government to run the day-to-day affairs, to take care of the security void that has



STORY HIGHLIGHTS

- **NEW:** Banks are shuttered until Wednesday as protests force top banker's resignation
- **NEW:** ElBaradei urges generals to "come out of their headquarters"
- **NEW:** Stock exchange to freeze transactions from officials being investigated
- Egypt's ambassador says the military will run a "technocratic" government until elections




Clouds of Information

EDITION: U.S. | INTERNATIONAL | MÉXICO
Set edition preference

Home Video NewsPulse U.S. World

Egypt's military dissolves parliament, suspends constitution

By the CNN Wire Staff
February 13, 2011 2:44 p.m. EST



Click

Egypt suspends constitution

STORY HIGHLIGHTS

- **NEW:** Banks are shuttered until Wednesday as protests force top banker's resignation
- **NEW:** ElBaradei urges generals to "come out of their headquarters"
- **NEW:** Stock exchange to freeze transactions from officials being investigated
- Egypt's ambassador says the military will run a "technocratic" government until elections

Cairo, Egypt (CNN) — The Egyptian military consolidated its control over what it has called a democratic transition from nearly three decades of President [Hosni Mubarak's](#) authoritarian rule, dissolving the feeble Parliament, suspending the Constitution and calling for elections in six months in sweeping steps that echoed protesters' demands.


The Supreme Council of the Armed Forces, read on television, effectively put Egypt under direct military authority, thrusting the country into territory uncharted since republican Egypt was founded in 1952. Though enjoying popular support, the military must now

Sameh Shoukry, Egypt's ambassador to the United States, said Sunday that the generals have made restoring security and reviving the economy its top priorities.

"This current composition is basically a technocratic government to run the day-to-day affairs, to take care of the security void that has

Only at BlackBerry
Get the ScoreMobile Super App


Egyptian Military Dissolves Parliament



Protesters resisted being removed from Tahrir Square by Egyptian soldiers on Sunday.

By ANTHONY SHADID
Published: February 13, 2011

Multimedia




Tahrir Square
Area of

Mideast Unrest Spreads

Protests Target Iran, Bahrain, Libya; Egypt Dissolves Parliament

Article Video Slideshow

By MARGARET COKER, MATT BRADLEY and TAMER EL



Officials removed a portrait of ousted Egyptian President Hosni Mubarak at the main Cabinet building in Cairo on Sunday.

CAIRO—As Egypt's new military leadership suspended the constitution, dissolved parliament and promised fresh elections, demands for similar political reform swept across the Arab world—from Libya to Iran—following the resignation of President Hosni Mubarak.

Egypt's dramatic moves incorporate many demands issued during the mass demonstrations by

THE WALL STREET JOURNAL. MIDDLE EAST

Monday, February 14, 2011 New York 40°/30°

U.S. Edition Home Today's Paper Video Steps Journal C

World U.S. New York Business Markets T


Asia China Hong Kong Japan India Europe U.K. R

TOP STORIES IN World Massive Population Lifts Nation's Growth

MIDDLE EAST NEWS FEBRUARY 14, 2011

Article Video Slideshow

By MARGARET COKER, MATT BRADLEY and TAMER EL



Officials removed a portrait of ousted Egyptian President Hosni Mubarak at the main Cabinet building in Cairo on Sunday.

CAIRO—As Egypt's new military leadership suspended the constitution, dissolved parliament and promised fresh elections, demands for similar political reform swept across the Arab world—from Libya to Iran—following the resignation of President Hosni Mubarak.

Egypt's dramatic moves incorporate many demands issued during the mass demonstrations by

USA TODAY | World

U.S. Edition Home Today's Paper Video Steps Journal C

World U.S. New York Business Markets T


Asia China Hong Kong Japan India Europe U.K. R

TOP STORIES IN World Massive Population Lifts Nation's Growth

MIDDLE EAST NEWS FEBRUARY 14, 2011

Article Video Slideshow

By MARGARET COKER, MATT BRADLEY and TAMER EL



Officials removed a portrait of ousted Egyptian President Hosni Mubarak at the main Cabinet building in Cairo on Sunday.

CAIRO—As Egypt's new military leadership suspended the constitution, dissolved parliament and promised fresh elections, demands for similar political reform swept across the Arab world—from Libya to Iran—following the resignation of President Hosni Mubarak.

Egypt's dramatic moves incorporate many demands issued during the mass demonstrations by

... 27,000+ more

TWITTER

SIGN IN TO E-MAIL

PRINT

SINGLE PAGE

REPRINTS

SHARE

NOW PLAYING EVERYWHERE



Cross-Document Summaries

Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a \$2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case, the judge set bail at \$40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail. The Mean Girls star is due back in court on Feb. 23, an important hearing in which Lohan could opt to end the case early.

3news.co.nz

Pickering murder accused warned he would kill - awards

Home National World Sport Entertainment Business Politics Technology EnvironmentSci

You are here » Entertainment » Story 3 News Chase | 3 News in a minute

Full Story View Video

Lindsay Lohan pleads not guilty to theft

0 comments | Post Comment

THU, 10 FEB 2011 12:02 PM

Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a US\$2,500 necklace - a case that could return the troubled starlet to jail rather than the big screen.

She entered the plea through her attorney Shawn Chapman Holley after a judge scolded her about following the law.

"Look around this room," Superior Court Judge Keith Schwartz told Lohan. "Everybody else has to follow the law. You're no different than anyone else. So please, don't push your luck."

Schwartz set bail at US\$40,000. He said it appeared Lohan had violated her probation in a 2007 drunken driving case and warned that if she violated the law while free on bail he would have her arrested and held.

Lohan spoke little during the hearing except to acknowledge her name and that she understood the charge and the possible consequences.

The Mean Girls star was met by a swarm of photographers

Get a quote now.
Save it, then buy online later.

WA today
LIFE & STYLE

Choose Qantas & enjoy:
Genuine Aussie Hospitality
Complimentary food & wine
More than 400 on-board
The largest domestic fleet

Beauty Wellbeing Fashion People Home & Style Life Matters Weddings Shopping Culture

You are here Home » Life & Style » People » Article

Lindsay Lohan charged with theft

February 10, 2011

Join the conversation
You're the only person reading this now. Tell your friends

Top Life & Style articles

1. Tackling the trial weight loss hurdle
2. Stretching the friendship
3. True romance
4. Women push for reforms under bulk bill reform
5. Victoria Beckham finds her style

More Life & Style articles

Story Tools

- Share on Facebook
- Email this story
- Print this story

WA Today Jobs

Assistant Property Man... \$45,000 - \$50,000
East Perth, WA 6004
Adesso Services

THE DAILY CALLER

Log In | Register

HOME POLITICS US ENTERTAINMENT CRIME WORLD SPORTS BUSINESS TECH LIFE VIDEO

Judge tells Lohan she's no star in his courtroom

By admin | Published: 8:11 PM 02/09/2011 | Updated: 10:05 AM 02/10/2011

LOS ANGELES (AP) — Lindsay Lohan walked into a courtroom to face a felony grand theft charge looking like a million dollars, only to be told by a judge she was no different than anyone else.

Lohan's arraignment on a charge that she stole a \$2,500 necklace from an upscale jeweler wasn't the first time a judge threatened to throw the troubled starlet in jail. But it was the first time a judge wielded enough power to keep her locked up for a long time.

"You're in a different situation now that a felony has been filed," Los Angeles Superior Court Judge Keith Schwartz said after the actress pleaded not guilty Wednesday.

"Everybody else has to follow the law," Schwartz said, noting that he was giving the actress a tamer version of a lecture he'd delivered to her attorney behind closed doors and away from the dozens of assembled reporters.

"You're no different than anyone else. So please, don't push your luck."

Testing the limits — in the courtroom rather than the big screen — has been Lohan's calling card in recent months.

The "Mean Girls" star is due back in court on Feb. 23, an important hearing in which Lohan could opt to end the case early. Her attorney, Shawn Chapman Holley, indicated Wednesday that the actress was interested in an early disposition program if the terms are right.



More Data: Machine Translation

| | |
|------------|---|
| SOURCE | Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante. |
| HUMAN | That would be an interim solution which would make it possible to work towards a binding charter in the long term . |
| 1x DATA | [this] [constituerait] [assistance] [transitoire] [who] [permettrait] [licences] [to] [terme] [to] [a] [charter] [to] [value] [contraignante] [.] |
| 10x DATA | [it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to a] [charter] [to] [value] [binding] [.] |
| 100x DATA | [this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding] [.] |
| 1000x DATA | [that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter] [.] |



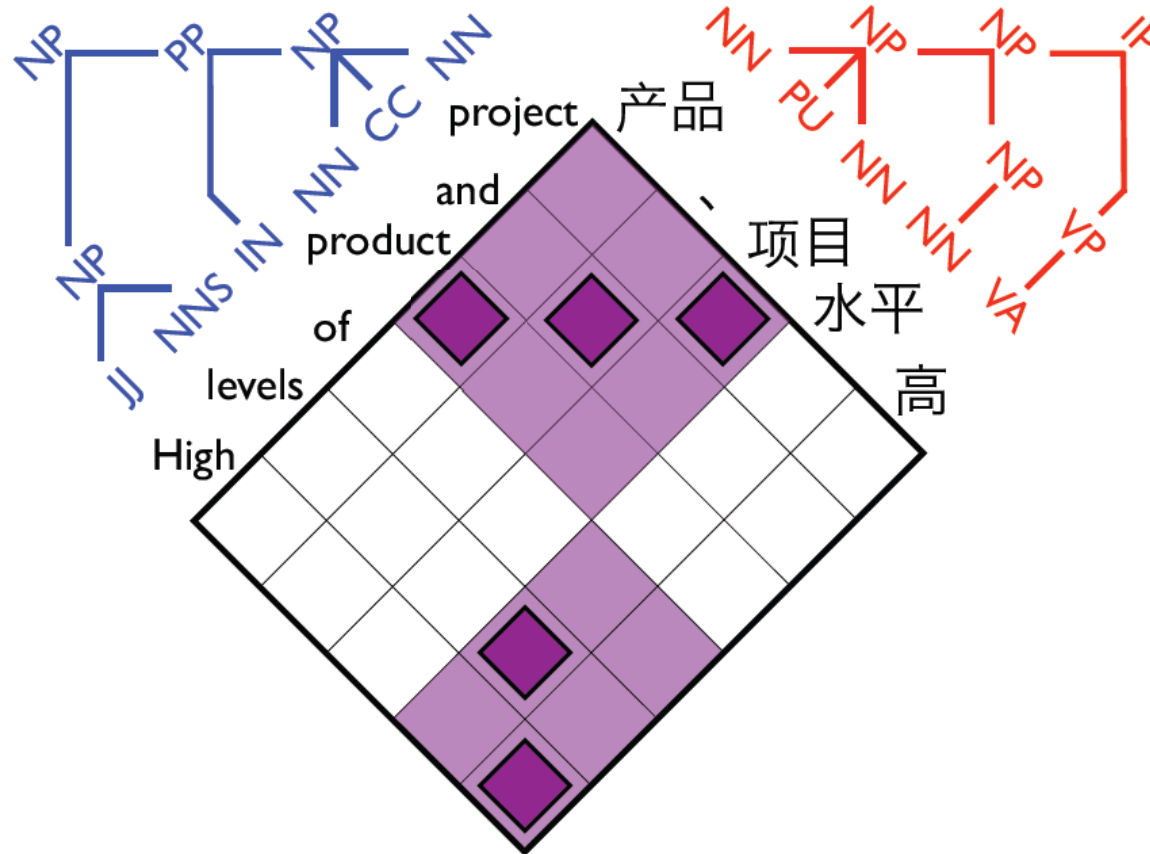
Data By Itself Isn't Enough!

CLASSIC SOUPS

| CLASSIC SOUPS | | | Sm. | Lg. |
|---------------|-----|---|------|------|
| 清 燉 雞 湯 | 57. | House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot) | 1.50 | 2.75 |
| 雞 飯 湯 | 58. | Chicken Rice Soup | 1.85 | 3.25 |
| 雞 麵 湯 | 59. | Chicken Noodle Soup | 1.85 | 3.25 |
| 廣 東 雲 吞 | 60. | Cantonese Wonton Soup..... | 1.50 | 2.75 |
| 蕃 茄 蛋 湯 | 61. | Tomato Clear Egg Drop Soup | 1.65 | 2.95 |
| 雲 吞 湯 | 62. | Regular Wonton Soup | 1.10 | 2.10 |
| 酸 辣 湯 | 63. | Hot & Sour Soup | 1.10 | 2.10 |
| 蛋 花 湯 | 64. | Egg Drop Soup..... | 1.10 | 2.10 |
| 雲 蛋 湯 | 65. | Egg Drop Wonton Mix..... | 1.10 | 2.10 |
| 豆 腐 菜 湯 | 66. | Tofu Vegetable Soup | NA | 3.50 |
| 雞 玉 米 湯 | 67. | Chicken Corn Cream Soup | NA | 3.50 |
| 蟹 肉 玉 米 湯 | 68. | Crab Meat Corn Cream Soup..... | NA | 3.50 |
| 海 鮮 湯 | 69. | Seafood Soup..... | NA | 3.50 |




Analysis and Alignment



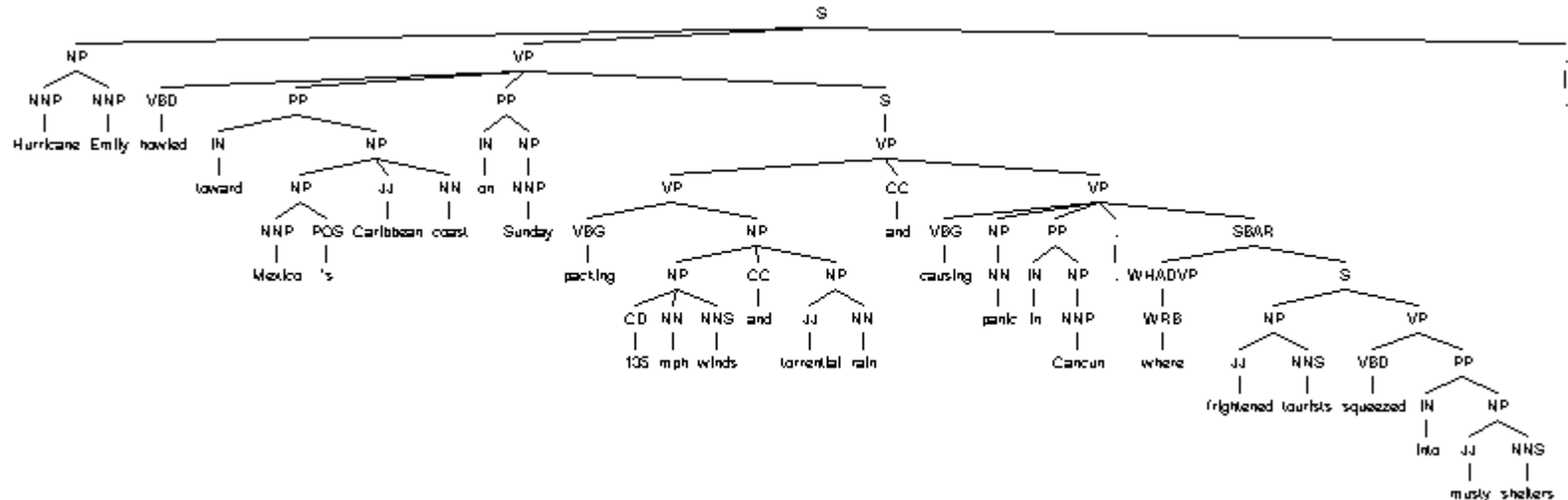


Data and Knowledge

- Classic knowledge representation worry: How will a machine ever know that...
 - Barack Obama is the president?
 - Beige looks like this: 
 - Chairs are solid?
- Answers:
 - 1980: write it all down
 - 2000: get by without it
 - 2020: learn it from data



Deeper Linguistic Analysis

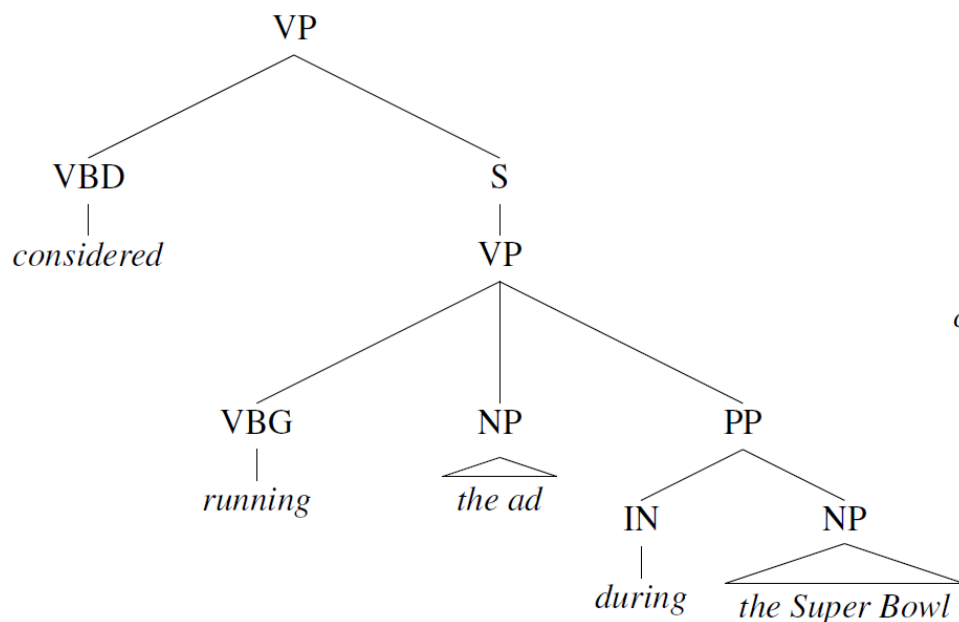


Hurricane Emily howled toward Mexico 's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters .



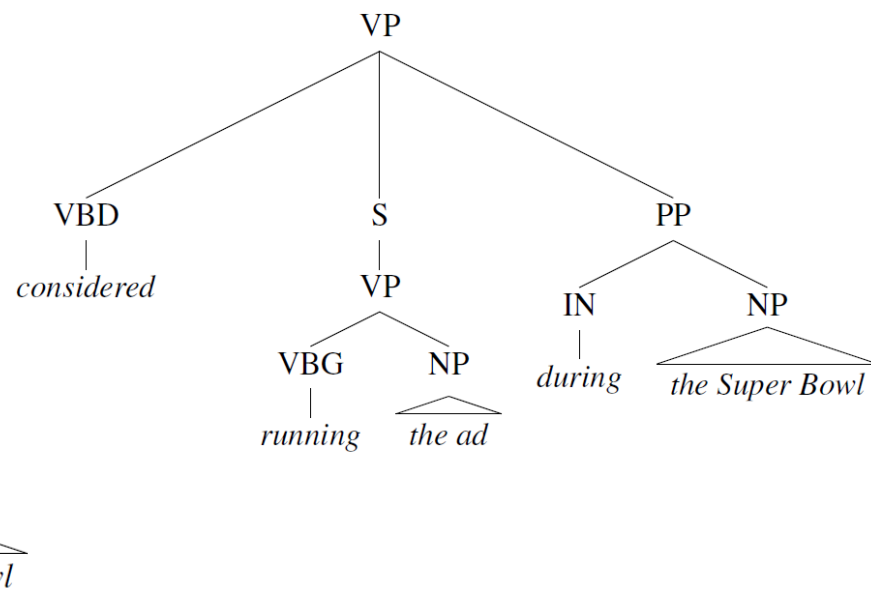
Data and Knowledge: Parsing

They considered running the ad during the Super Bowl.



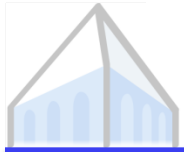
*running * during: 3k*

running it during: 239



*considered * during: 2k*

considered it during: 112



Deeper Understanding: Reference

Q: Who signed the Serve America Act?

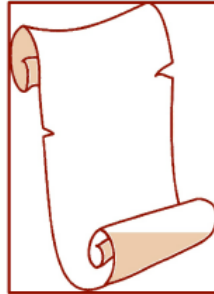
A: Barack Obama

Los Angeles Times

President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.



Names vs. Entities



President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.



Example Errors

Input

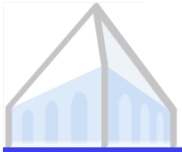
America Online announced on Monday that the company plans to update its instant messaging service.

Correct

America Online the company its
instant messaging service

Guess

America Online
the company its
instant messaging service



Discovering Knowledge

America Online ← → **company**

America Online, LLC (commonly known as **AOL**) is an **American global Internet services and media company** operated by **Time Warner**. It is headquartered at 770 Broadway in **Midtown Manhattan, New York City**.^{[2][3]} Founded in 1983 as **Quantum Computer Services**, it has **franchised** its services to companies in several nations around the world or set up **international versions** of its services.^[4]

America Online



Type

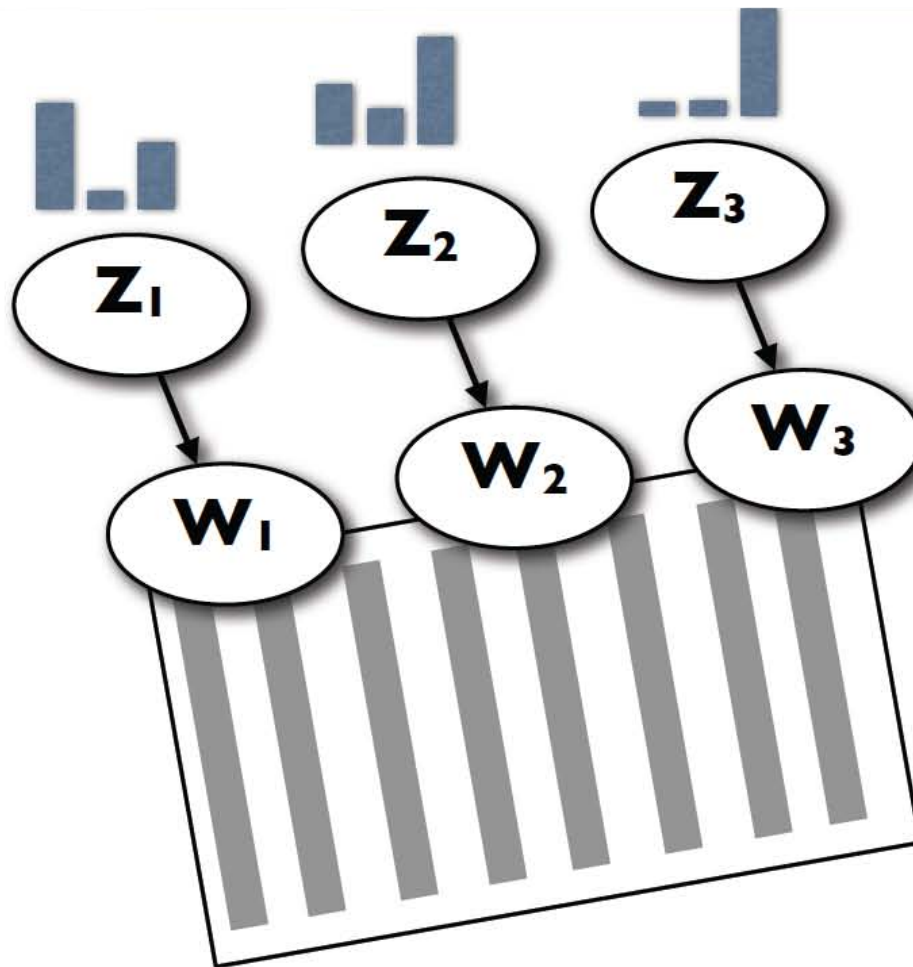
Subsidiary of Time Warner

Founded

1983 as *Quantum Computer Services*



Unsupervised Learning

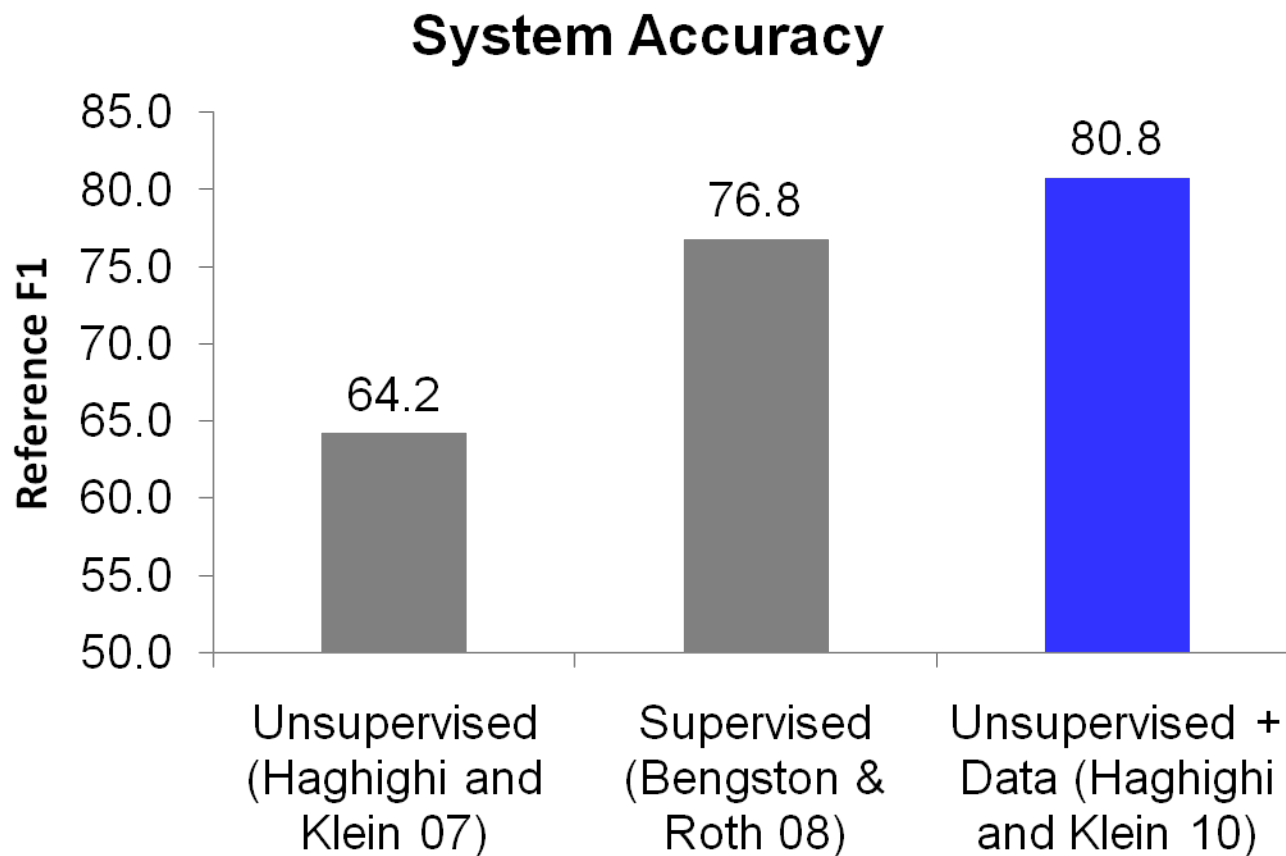


Parameters

| | | |
|------|------|------|
| 0.20 | 0.45 | 0.35 |
| 0.55 | 0.28 | 0.17 |
| 0.31 | 0.51 | 0.18 |
| 0.06 | 0.70 | 0.24 |
| 0.62 | 0.13 | 0.15 |



Coreference Systems





Cross-Document Identity



Dan Klein

Associate Professor
Computer Science Division
University of California at Berkeley

Contact Information

Email klein@cs.berkeley.edu
Mail Dan Klein, Soda Hall, Berkeley, CA 94720-1776
Phone (510) 643-0805 (email w/voicemail)


 [Contact Info](#)

 [Research](#)

 [Background](#)

 [Teaching](#)

 [Publications](#)

 [Personal](#)

Research

My research focuses on the automatic organization of natural language information. Some topics of interest to me are:

- Unsupervised language acquisition
- Machine translation
- Efficient algorithms for NLP
- Information extraction
- Linguistically rich models of language
- Integrating symbolic and statistical methods for NLP
- Organization of the web

[My group's web page \(the Berkeley Language Processing Group\).](#)

Our agent, the *Overmind*, won the *StarCraft AI competition*!

Background

Daniel Klein



dklein@gmu.edu

I am Professor of Economics at George Mason University. I have degrees from George Mason University and New York University, where I studied the classical liberal traditions of economics. My research focuses on economic principles, public policy issues, and the tradition of Adam Smith and Friedrich Hayek.

I've published research on policy issues including toll roads, carbon auto emission, credit reporting, and the Food and Drug Administration. I've also written on spontaneous order, coordination, the distribution of opportunity, the demand and supply of assurance, why governments exist, officials believe in the goodness of bad policy, why people fail to act, government intervention more than they should, and the relationship between liberty, dignity, and responsibility.

Lately my research dwells on Adam Smith. Russ Roberts and I have a multipart audio book club on *The Theory of Moral Sentiments*.

I participate in the new GMU Econ course sequence and field exam in **Smithian Political Economy (SPE)**. The first SPE field exam will be given in August 2011. ([Some sample field questions.](#)) Also, I lead an Adam Smith reading group and an *Invisible Hand Seminar* for graduate students.

I am the chief editor of *Econ Journal Watch*, an online journal dedicated to economic criticism from a Smith-Hayek viewpoint. I've contributed several papers on the character heterogeneity of economists. I push the

UCLA
BRUINS

#1 IN NCAA CHAMPION

THE OFFICIAL WEBSITE OF UCLA ATHLETICS

AUDIO

UCLABRUINS

SPORTS | DONATE | TICKETS | RECRUITS | FAN CENTRAL | CAMPS | ABOUT UCLA

ROSTER | SCHEDULE/RESULTS | STATISTICS | NEWS | ARCHIVES

BA

Like

Sign Up to see what your friends like.

Share


Print

Email

Text

RSS

23 Dan Klein



PROFILE

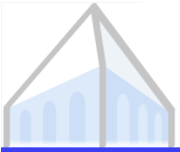
- Class: RS Sophomore
- Hometown: Los Alamitos, Calif.
- High School: Servite
- Height / Weight: 6-3 / 190
- Position: RHP
- Bats/Throws: R/R

Notes

Helped lead UCLA to two NCAA Regionals (2008, 2010) and the Bruins' third-ever trip to the College World Series in 2010...had a sensational season as the Bruins' closer in 2010...as a redshirt sophomore in 2010, recorded 10 saves, the second-highest single-season total in school history...selected in the third round (85th overall) of the 2010 MLB Draft by the Baltimore Orioles.

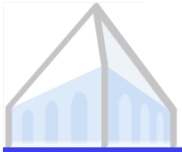
2010

Made 39 appearances (all in relief), tying UCLA's single-season appearances record...also tied for the Pac-10 lead in appearances with Arizona State's Mitchell Lambson...served as UCLA's closer, notching 10 saves, the second-highest total single-season total in program history and the third-most saves among Pac-10 pitchers in 2010...earned All-Pac-10 Team honors, going 6-1 with a team-leading 1.90 ERA...tallied 55 strikeouts and 11 walks in 52.0 innings, limiting the opposition to



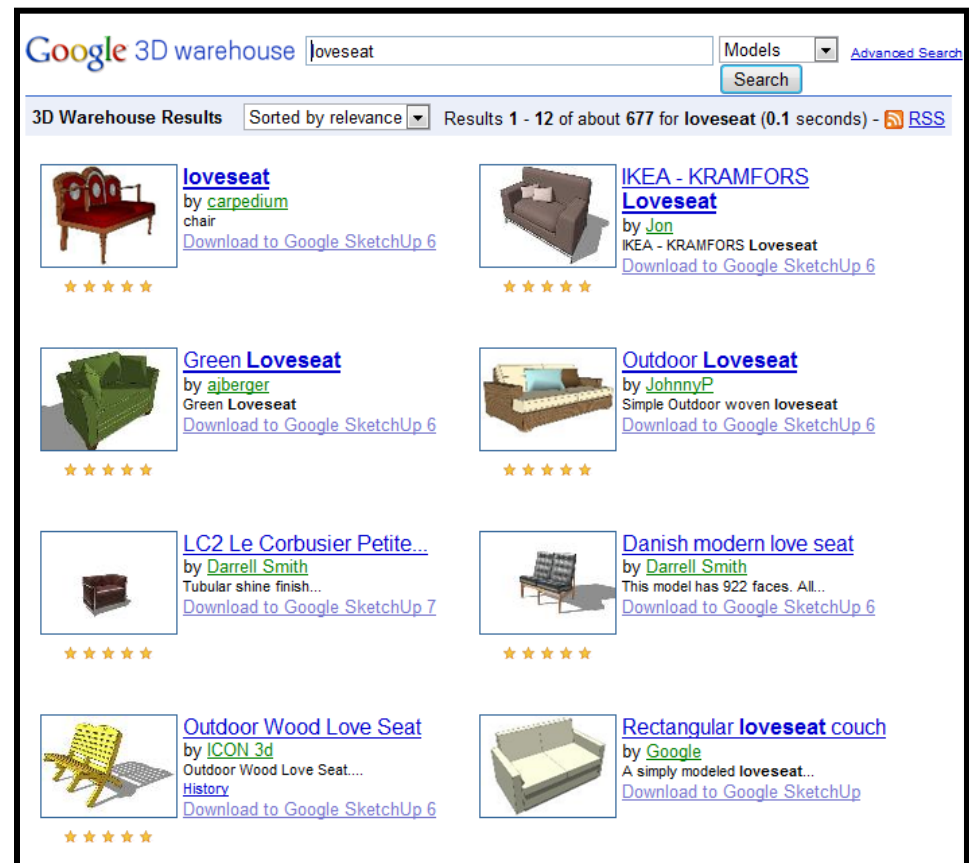
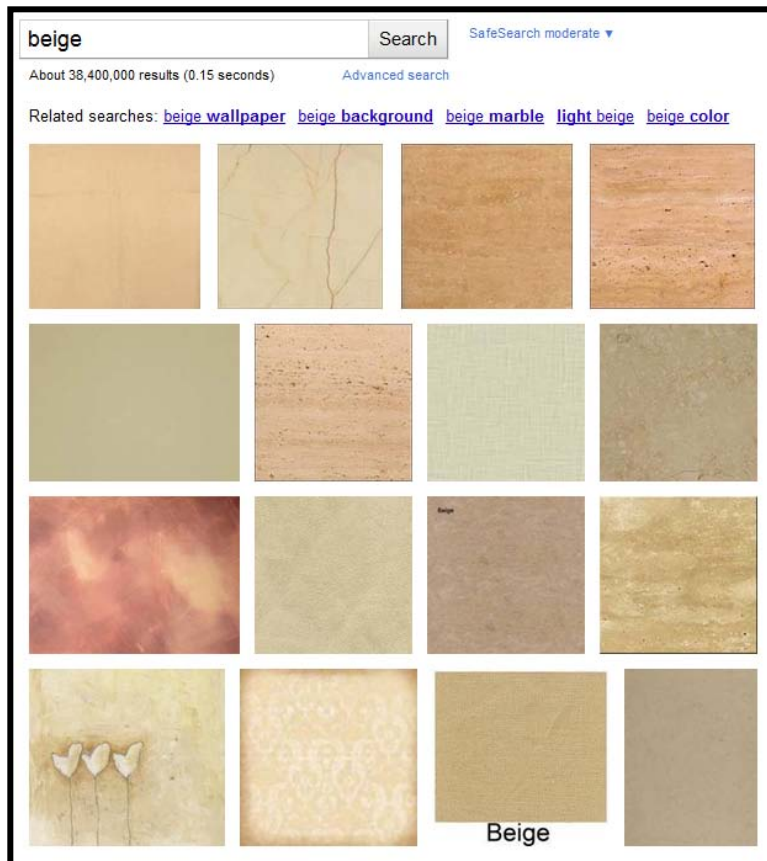
Grounded Language

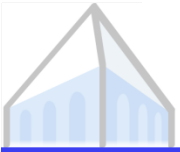




Grounding with Natural Data

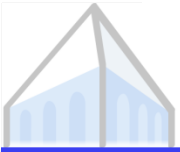
... on the beige loveseat.





Predictions

| Today | 2020 (likely) | 2020 (hopefully) |
|--|-----------------------------|-------------------------------------|
| Find information | Synthesize information | Infer information |
| Keywords and names | Entities | Concepts |
| Knowledge-free “structural” systems | Knowledge from text | Knowledge from grounded contexts |
| “Talk” to search engines | Talk to embedded devices | Talk to mobile robots |
| Superficial patterns | Deep understanding | Monologs → dialogs |



Conclusion

- Simple algorithms and large data have gotten us amazingly far!
- To go further, we need
 - Algorithms that work with deeper structure
 - Systems that are contextualized
 - Learning methods that turn data into knowledge

Thank you!

