Migratory Words

Computational Tools for Investigative Political Research

Can computational systems help investigative journalists to identify patterns of intellectual influence that shape the creation of public policy?

Hyunwoo Park
Gopal Vaswani
Jeremy Whitaker
Michael Manoochehri

Advisor: Eric Kansa

Partners: Rob Ennals, Intel Research
Laurent El Ghaoui, EECS
Saheli Datta

Migratory Words

Computational Tools for Investigative Political Research

The Importance of Journalism

Migratory Words

Computational Tools for Investigative Political Research

"Statements by more than a dozen lawmakers were ghostwritten, in whole or in part, **by Washington lobbyists**...The lobbyists, employed by Genentech and by two Washington law firms, were remarkably successful in getting the statements printed in the Congressional Record under the names of different members of Congress. Genentech, a subsidiary of the Swiss drug giant Roche, estimates that **42 House members picked up some of its talking points** — 22 Republicans and 20 Democrats, an unusual bipartisan coup for lobbyists."

Robert Pear
N.Y.T, Nov. 14th, 2009

Migratory Words

Computational Tools for Investigative Political Research

This is a complex intersection where private interests clash over ideas and seek influence public policy.

Migratory Words

Computational Tools for Investigative Political Research

Corporate Speech

PR Documents

Lobbyists

Expert Knowledge

Think Tanks

Academia

Government Records

Congressional

Presidential

Hearings

Migratory Words

Computational Tools for Investigative Political Research

Corporate Speech

Constructed speech designed to influence

Expert Knowledge

A bridge between academia and policy

Government Records

Policy and the opinion of policy makers

Corporate Speech	Campaign Finance Law, penalties
Expert Knowledge	Should behave like a 501c3
Government Records	Accountable to "informed constituency"

Migratory Words

Computational Tools for Investigative Political Research

≠

IDENTITY SIGNALS

Health Care Takeover

Corporate Welfare

Military Industrial Complex

Death Tax

Migratory Words

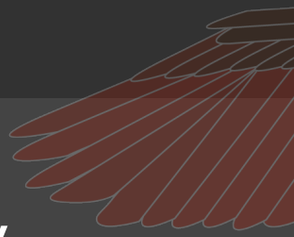
Computational Tools for Investigative Political Research

We are, as language animals, to a substantial degree simply

permuters of substrings

already well established within the history of the whole language

(Wilks, n.d.).



Migratory Words

Computational Tools for Investigative Political Research

"Interestingness"

memorable sticky emotional

Drill Baby, Drill

Take for example:

> "One of the reasons I have long supported the U.S. biotechnology industry is that it is a homegrown success story that has been an engine of job creation in this country. Unfortunately, many of the largest companies that would seek to enter the biosimilar market have made their money by outsourcing their research to foreign countries like India."
>
> Robert Pear, 2009

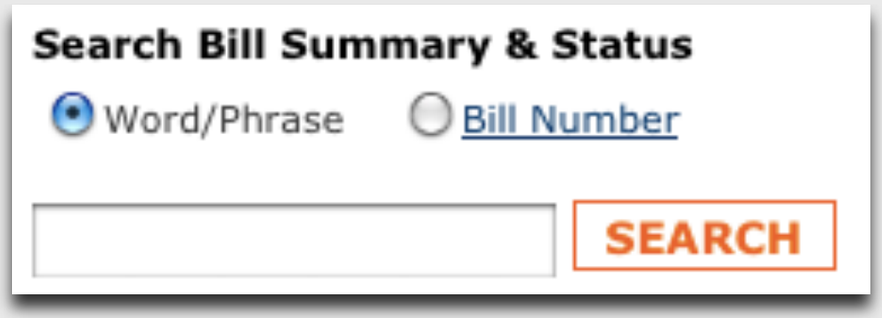
Migratory Words

Computational Tools for Investigative Political Research

How did Robert Pear and the NYT Make the Link?

Given Insider Text

Intuition-based Identification



Line by Line Search

Journalism's Current Methods

using networks of people

playing details against each other

identifying difficult patterns

aggregating different sources

generating trust with sources

applying human intuition as the investigator

understanding what is interesting to readers

leading the reader through a compelling story-line

(lots of tacit knowledge)

Migratory Words

Computational Tools for Investigative Political Research

Augmentation

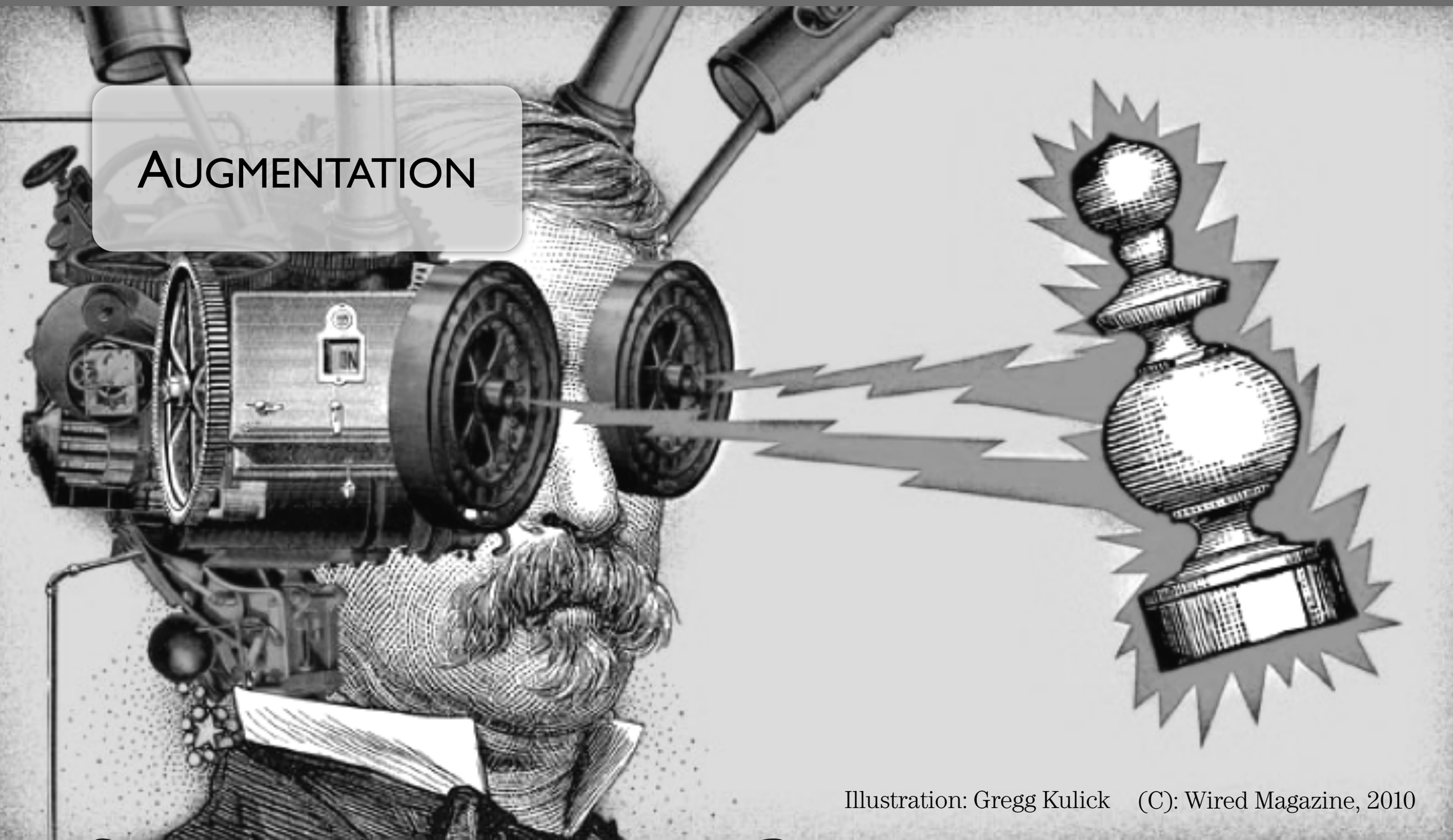
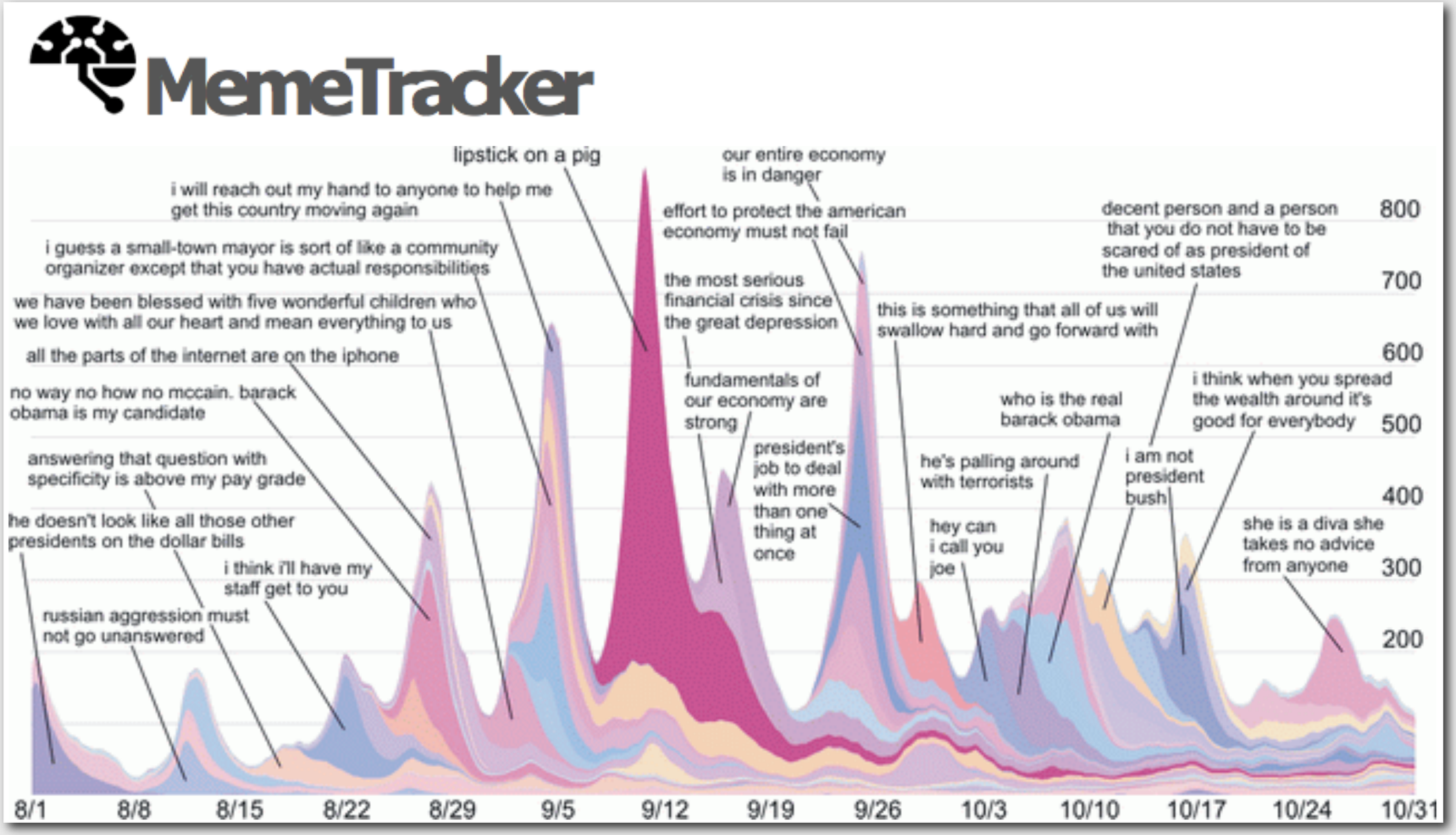


Illustration: Gregg Kulick (C): Wired Magazine, 2010

Migratory Words

Computational Tools for Investigative Political Research



Migratory Words

Computational Tools for Investigative Political Research

Google trends

iceland volcano, oil spill

Search Trends

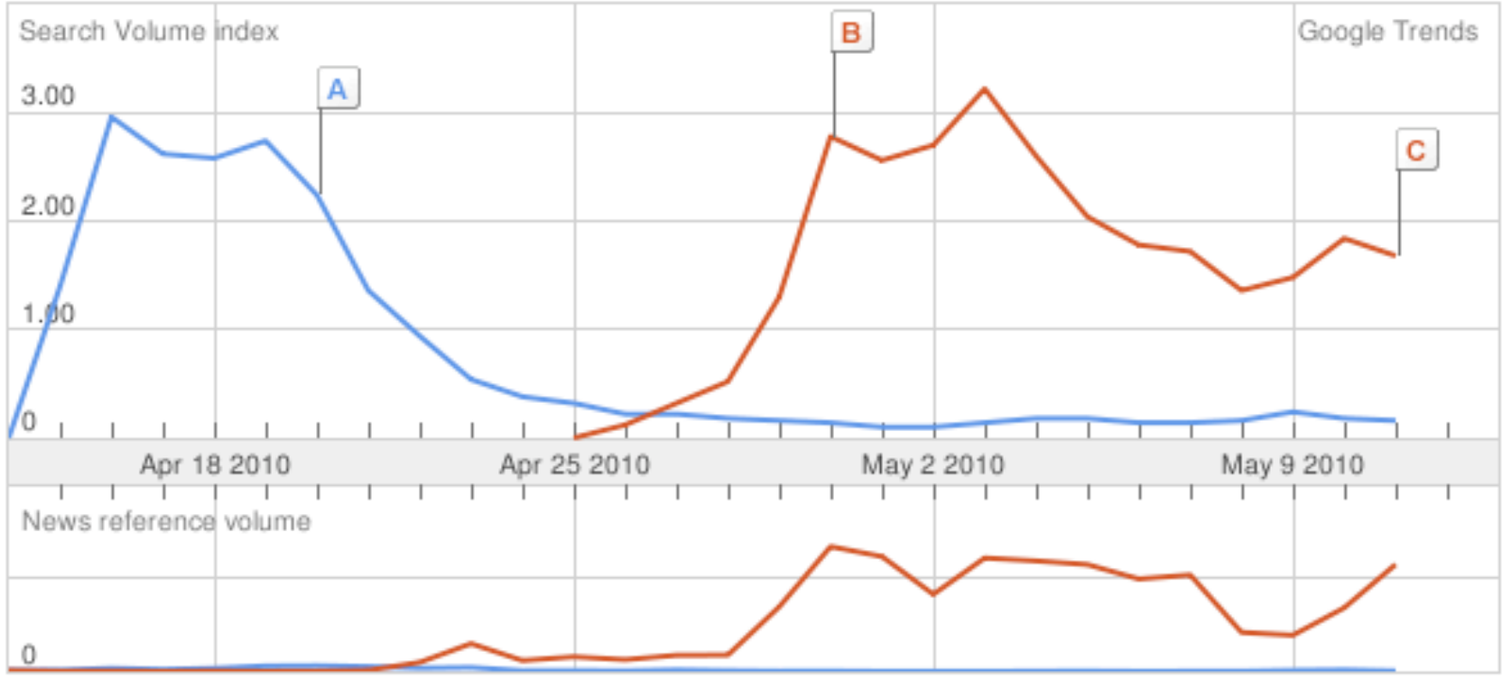
Tip: Use commas to compare multiple search terms.

Searches Websites

United States

Scale is based on the average traffic of **oil spill** from United States in the last 30 days. Learn more

iceland volcano 0.74 **oil spill** 1.00



A Iceland volcano
Daily Mail - Apr 20 2010

B Gulf Coast oil spill comes ashore
Newsday (subscription) - Apr 30 2010

C Gulf oil spill swells to 4m gallons
Stuff.co.nz - May 11 2010

More news results »

Rank by oil spill

The "Migratory Words" Tool

They said that the health care bill passed by the House of Representatives "keeps in place the longstanding and widely supported federal policy against government funding of elective abortions and plans that include elective abortions" and "ensures that where federal funds are involved, people are not required to pay for other people's abortions." The Senate bill does not maintain this commitment.

In the Senate version, "federal funds will help subsidize, and in some cases a federal agency will facilitate and promote, health plans that cover elective abortions," the bishops said. "All purchasers of such plans will be required to pay for other people's abortions in a very direct and explicit way, through a separate premium payment designed solely to pay for abortion. There is no provision for individuals to opt out of this abortion payment in federally subsidized plans, so people will be required by law to pay for other people's abortions." The public consensus against abortion funding, said the bishops, "is borne out by many opinion surveys, including the new Quinnipiac University survey December 22 showing 72 percent opposed to public funding of abortion in health care reform legislation."

"This bill also continues to fall short of the House-passed bill in preventing governmental discrimination against health care providers that decline involvement in abortion," the bishops said. And it also "includes no conscience protection allowing Catholic and other institutions to provide and purchase health coverage consistent with their moral and religious convictions on other procedures.

U.S. Conference of Catholic Bishops
Dec. 22 2009

They said that the health care bill passed by the House of Representatives "keeps in place the longstanding and widely supported federal policy against government funding of elective abortions and plans that include ***elective abortions****" and "ensures that where federal funds are involved, people are not required to pay for other people's abortions." The Senate bill does not maintain this commitment.*

*In the Senate version, "****federal funds will help subsidize****, and in some cases a federal agency will facilitate and promote,* ***health plans that cover elective abortions****," the bishops said. "All purchasers of such plans will be required to pay for other people's abortions in a very direct and explicit way, through a separate premium payment designed solely to pay for abortion. There is no provision for individuals to opt out of this abortion payment in* ***federally subsidized plans****, so people will be required by law to pay for other people's abortions." The public consensus* ***against abortion funding****, said the bishops, "is borne out by many opinion surveys, including the new Quinnipiac University survey December 22 showing 72 percent opposed to public funding of abortion in health care reform legislation."*

"This bill also continues to fall short of the House-passed bill in preventing governmental ***discrimination against health care providers*** *that decline involvement in abortion," the bishops said. And it also "includes no* ***conscience protection*** *allowing Catholic and other institutions to provide and purchase health coverage* ***consistent with their moral and religious convictions*** *on other procedures.*

U.S. Conference of Catholic Bishops
Dec. 22 2009

Methods: Corpus Construction

CORPUS INDEXING

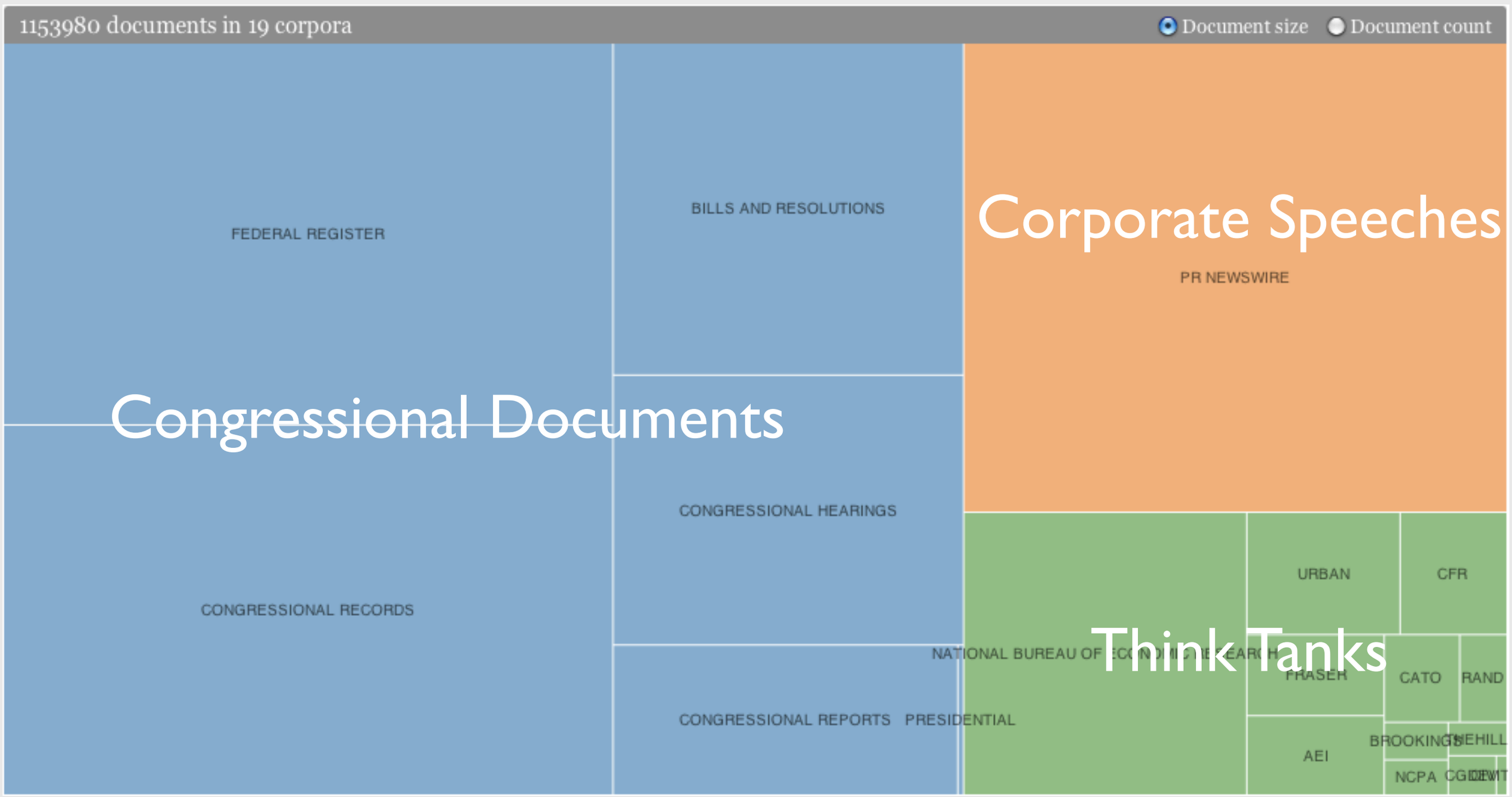
Indexed over 1.2 million documents

Over 1.7 billion terms

19 repositories of publicly available text

Migratory Words

Computational Tools for Investigative Political Research



Migratory Words

Computational Tools for Investigative Political Research

PRNewsWire (HTML)

All NewsReleases - Sept 09 - Mar 10 (6mo)
Including:
- All Health News
- All Financial News
- All Current Events

LouisDB

Bills and Resolutions
Congressional Reports
Congressional Record
Congressional Hearings
Federal Register
Presidential Documents

Corporate Text via "Hill Archive" HTML

All articles provided by "Hill" White-Paper database
Including:
- Health Care
- Health Cate Reform
- Public Health
- Finance

PRNewsWire via data dump

All NewsReleases - 2006- 2007 (2yrs)

PR Feeds

PR Newswire
United Business Media

Government Text

LOUISdb
Library of Unified Information Services

Position Papers

THE HILL
WHITE PAPER DATABASE

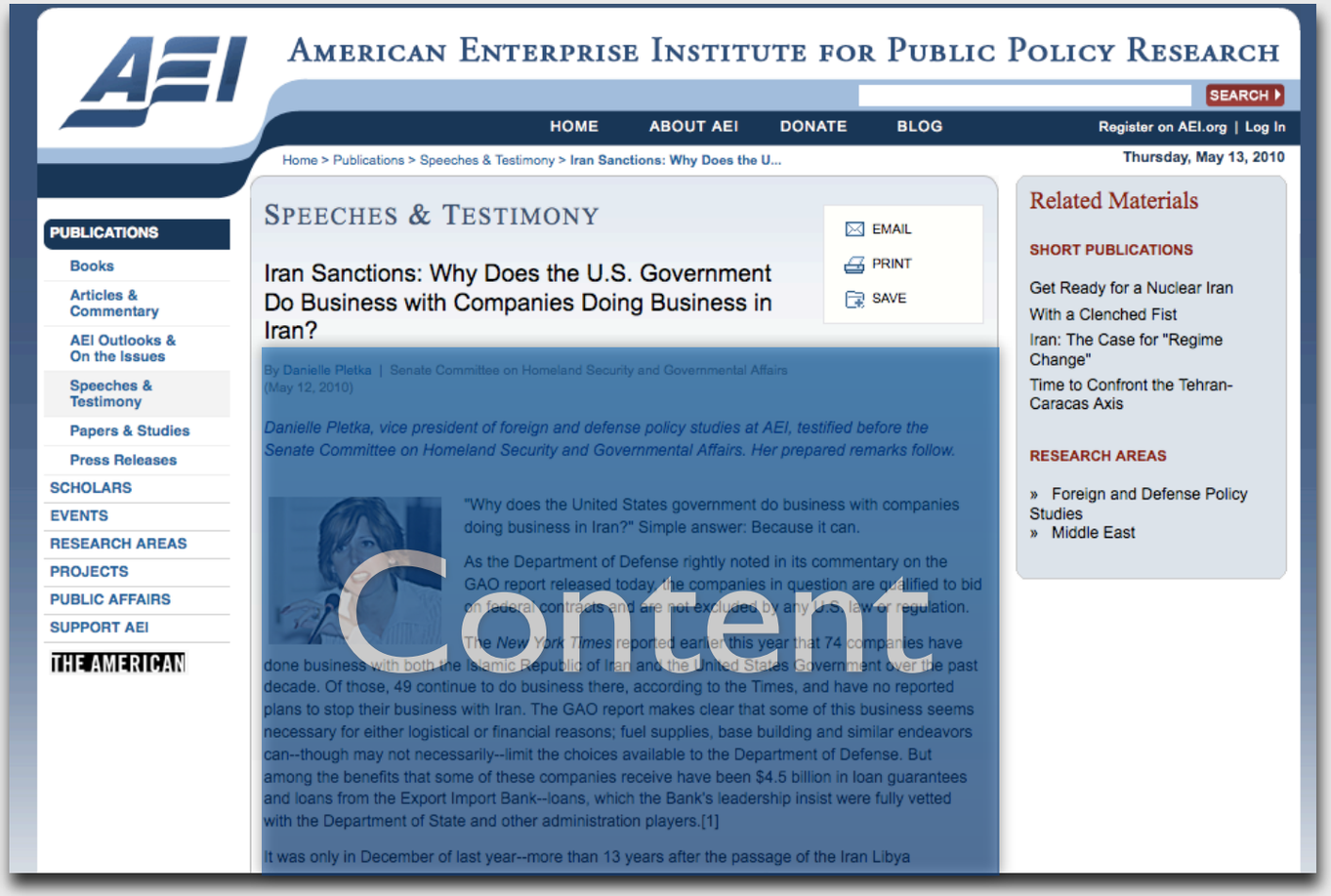
Think Tanks

CATO INSTITUTE · RAND · NBER · Urban Institute · NCPA · FRASER INSTITUTE · Center for Global Development · JOHN W. POPE CIVITAS INSTITUTE · AEI · BROOKINGS

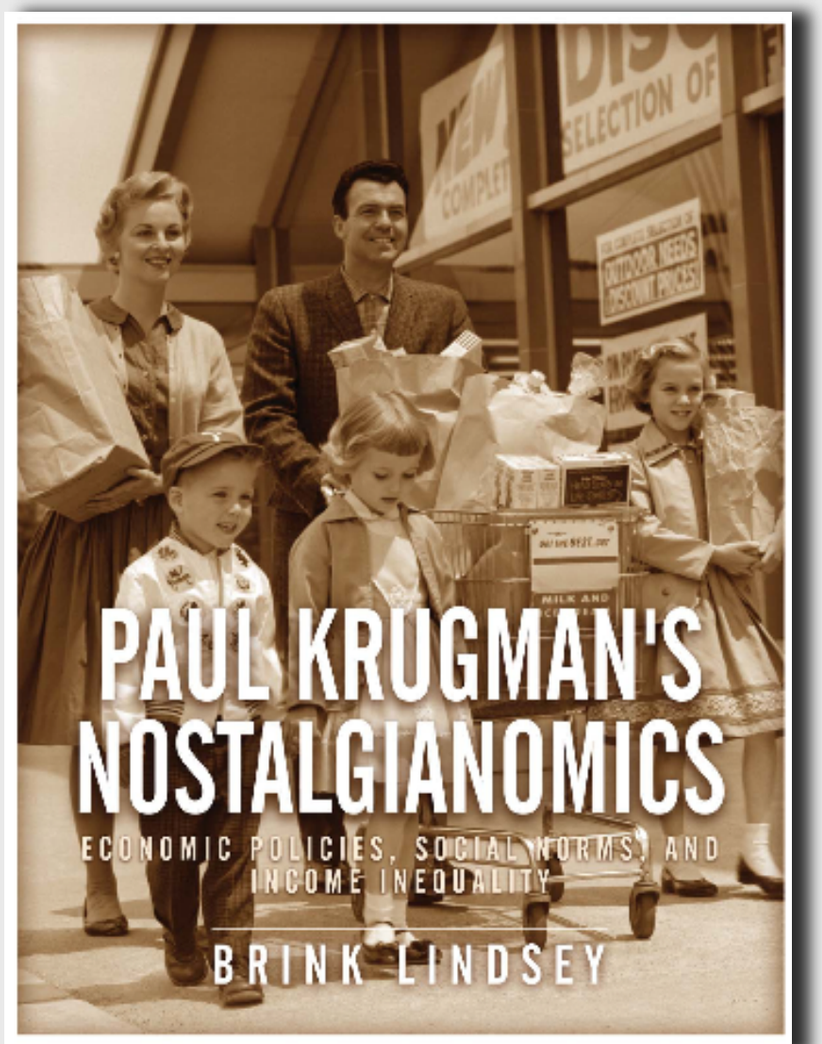
Think Tank (HTML and PDF)

RAND
CATO
Fraser
Center for Global Development
Brookings
NCPA
Civitas
Center for Foreign Relations
American Enterprise Institute
NBER
Urban Institute

DIFFERENT PARTS OF HTML



EVEN WORSE. PDFS.



offshore specific parts of the production process have magnified the downward pressure of trade on low-end wages. He concludes, however, that data on trade flows are not sufficiently detailed to permit a reliable quantitative estimate.[79]

The huge wave of immigration over the past generation, the result of a policy change catalyzed by changing social norms, has also exerted mild downward pressure on the wages of native-born low-skilled workers. Most estimates show a small effect: at the high end, Harvard economist George Borjas found that immigration between 1980 and 2000 depressed the wages of high-school dropouts by around 9 percent.[80]

The more dramatic impact of immigration on *measured* inequality, however, has come from its effect on the composition of the American workforce. Specifically, immigration has substantially increased the number of less-skilled workers, thereby increasing apparent inequality by depressing average wages at the low end of the income distribution. According to American University and Urban Institute economist Robert Lerman, excluding recent immigrants from the analysis would eliminate roughly 30 percent of the increase in adult male annual earnings inequality between 1979 and 1996.[81]

Interestingly, although the large influx of unskilled immigrants has made American inequality statistics look worse, it has actually reduced inequality for the people involved. After all, immigrants experience large wage gains as a result of relocating to the United States, thereby reducing the wage gap between them and top earners in this country. According to Lerman, if trends in inequality are recalculated to include, at the beginning of the period, recent immigrants and their native-country wages, a very different picture emerges. Instead of the 90/10 wage ratio increasing by 16.6 percent between 1979 and 1996, it actually fell by 4.7 percent. Thus, the result of immigration has been to reduce human inequality while increasing national inequality.[82]

Although the large influx of unskilled immigrants has made American inequality statistics look worse, it has actually reduced inequality for the people involved.

Changing social norms led, not only to a large infusion of foreign-born workers into the labor force, but to a large infusion of American women as well. How did the transformation of social attitudes about the role of women in society affect the inequality picture? The massive expansion of opportunities for women resulted in significant gains for gender equality: the female-male earnings ratio shot up from 0.30 in the mid-1960s to 0.57 in 2002.[83]

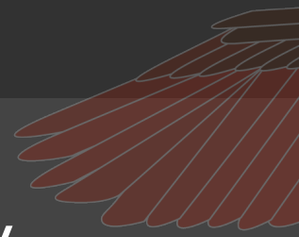
In terms of overall income inequality, though, gains for women have ended up widening rather than narrowing income differences. Since women's incomes have been rising faster than men's in recent decades, the shift from one- to two-earner households might have reduced inequality by supplementing lagging male earnings with more rapidly increasing female earnings. But because of the prevalence of "assortative mating"—that is, the tendency of people to choose spouses with similar educational and socioeconomic backgrounds—the rise in dual-income couples has actually exacerbated household income inequality. Between 1979 and 1996, the proportion of working-age men with working wives rose by approximately 25 percent among those in the top fifth of the male earnings distribution, and their wives' total earnings rose by over 100 percent.[84] According to a 1999 estimate by Gary Burtless of the Brookings Institution, this unanticipated consequence of the women's movement accounted for about 13 percent of the total rise in income inequality since 1979.[85]

What about rapidly rising incomes at the top of the pay scale? Is there any evidence that changes in policy and social norms have contributed to the rise of "winner-take-all" or "superstar" markets?

Economic logic suggests that the general throttling of competitive pressures during the Treaty of Detroit era probably did suppress high-end earnings by reducing competition among employers for the most talented executives and other key employees. After all, to the extent that trade barriers abroad and price and entry controls at home weakened the connection between firms' productivity and their bottom line, the importance of attracting and

DIFFERENT METADATA SCHEMES

corpus	doc_id	trackback	published_time	publisher	author	title	keywords	categories
bills	[filename]	<id>	<published>			<title>		
cr	[filename]	<id>	<published>			<title>		
fr	[filename]	<id>	<published>			<title>		
hearings	[filename]	<id>	<published>			<title>		
presidential	[filename]	<id>	<published>			<title>		
reports	[filename]	<id>	<published>			<title>		
aei	[filename]	URL:	DC.date:	publisher:	author:	DC.title:	keywords:	taxonomy:
cato	[filename]	URL:	date: or date_parsed: (first)		author:	title:	keywords:	sector:
cgdev	[filename]	URL:	date:	copyright:	authors:	title:	keywords:	sector:
fraser	[filename]	URL:	Date Published:		Author(s):			sector:
thehill	[filename]	URL:		organization:		title:	keywords:	sector:



IT WAS JUST CUTE.

Thursday, August 13, 2010

Aug 13

2010-08-13

20100813

2010

10-8-13

08-13-2010

Methods: Filter on `<interesting>`?

Migratory Words

Computational Tools for Investigative Political Research

President Obama opens new area

By Juliet Eilperin and Anne E. Kornblut
Washington Post Staff Writer
Thursday, April 1, 2010

VIDEO

ationwide, a solid majority of Americans support broader energy
xploration: In the latest national survey on the subject, 63 percent of
mericans told Pew pollsters in February that they favor allowing
ffshore oil and gas drilling in U.S. waters.

bama emphasized Wednesday that the policy fits within his ove
amework for clean-energy development. The nation will "need t
arness traditional sources of fuel even as we ramp up production of
ew sources of renewable, homegrown energy," he said at Joint Base
ndrews Naval Air Facility Washington, formerly Andrews Air Force
ase. "So today we're announcing the expansion of offshore oil and gas
xploration -- but in ways that balance the need to harness domestic
nergy resources and the need to protect America's natural resources."

Independence

View All Items in This Story

In what could represent the biggest expansion o
offshore energy exploration in half a century, Obama

Resize Print E-mail

26 Results found

Most Interesting:

Expansion of offshore oil See Details Add to favori
Found in: 2 documents | Present in the input text

Harness domestic See Details Add to favori
Found in: 3 documents | Present in the input text

Timeline view Corpora view view details

raditional sources of fuel See Details Add to favori
Found in: 6 documents | Present in the input text

Production of new sources See Details Add to favori
Found in: 7 documents | Present in the input text

Obama emphasized See Details Add to favori
Found in: 10 documents | Present in the input text

Expansion of offshore See Details Add to favori
Found in: 14 documents | Present in the input text

Announcing the expansion See Details Add to favori
Found in: 28 documents | Present in the input text

New sources of renewable See Details Add to favori
Found in: 33 documents | Present in the input text

TOKENIZE

rity of Americans support broader energy
national survey on the subject, 63 percent of
ters in February that they favor allowing more
ng in U.S. waters.

nesday that the policy fits within his overall
gy development. The nation will "need to
s of fuel even as we ramp up production of
, homegrown energy," he said at Joint Base
ty Washington, formerly Andrews Air Force
nouncing the expansion of offshore oil and gas
that balance the need to harness domestic
need to protect America's natural resources."

```
('ramp', 'up')
('up', 'production')
('production', 'of')
('of', 'new')
('new', 'sources')
('sources', 'of')
('of', 'renewable')
('renewable', ',')
(',', 'homegrown')
('homegrown', 'energy')
('energy', ',')
(',', 'he')
('he', 'said')
('said', 'at')
```

ramp up
up production
~~production of~~
~~of new~~
new sources
~~sources of~~
~~of renewable~~
~~renewable ,~~
~~, homegrown~~
homegrown energy
~~energy ,~~
~~, he~~
~~he said~~
~~said at~~

FILTER: REMOVE NOISE

- digits
- n-grams with many common words
- n-grams with really long terms (like URLs)
- etc...

SCORE FOR "UNLIKELINESS"

ramp up

up production

new sources

homegrown energy

up appears 1,049,124 times

homegrown appears 1,911 times

production appears 502,350 times

energy appears 1,106,644 times

COUNT

"ramp up"	"homegrown energy"
2,757 documents	67 documents

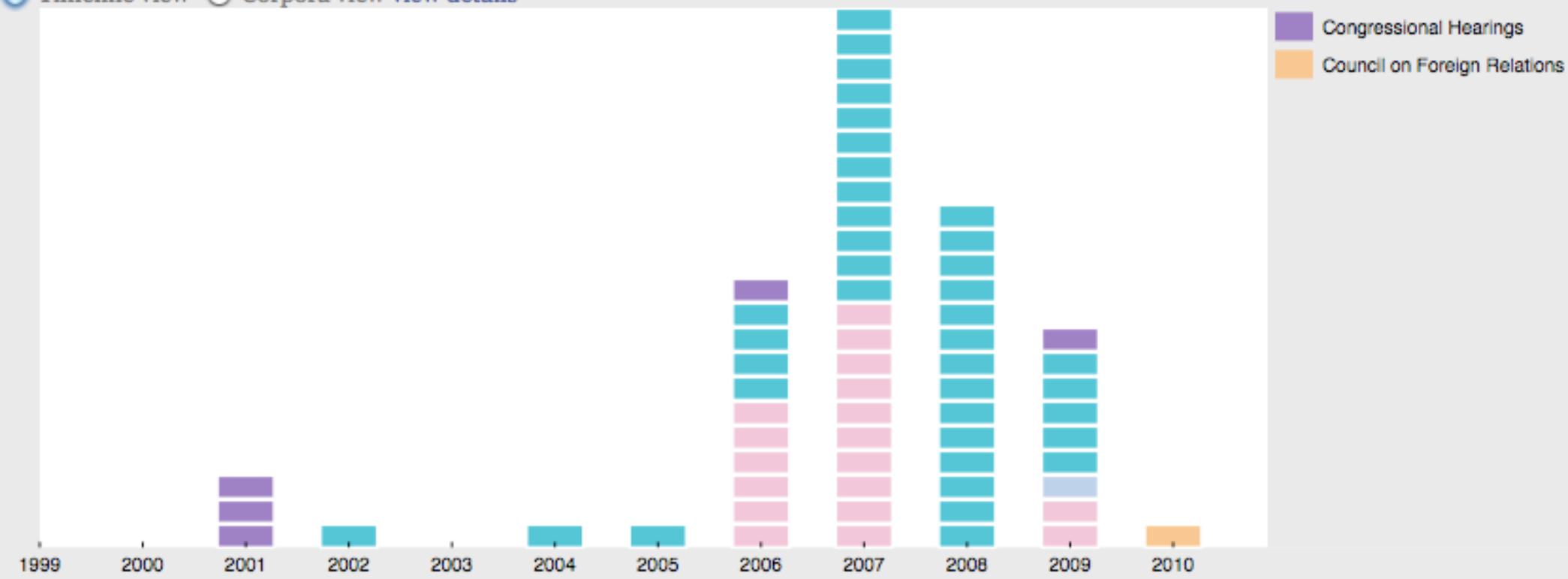
"HOMEGROWN ENERGY"

31 Results found

Most Interesting:

Homegrown energy See Details Add to favorites F
Found in: 67 documents | Present in the input text

Timeline view Corpora view view details



RENEWABLE FUELS FOR ENERGY SECURITY | Congressional Hearings (2001-07-06)
We believe that this is a program that is flexible and user friendly. We commend you and Senator Hagel for this bold and innovative step in moving our country toward **homegrown energy** sources. As you know, Mr. Chairman, for the last 8 to 10 years the U.S. soybean growers have invested their research and development and commercialization money in biodiesel.

IT'S A LITTLE MORE COMPLICATED

"Does the phrase have more that 50% stopwords?

"Does one term appear < 30 times in corpus?"

"Does a term have more two hyphens in it?"

"Does the n-gram end with a stopword?"

LIMITATIONS

SEMANTIC:

"biological nuclear national security missile regional stability firearms"

STATISTICAL:

"appropriations subcommittee in january"

Migratory Words

Computational Tools For Investigative Political Research

Can computational systems help investigative journalists to identify patterns of intellectual influence that shape the creation of public policy?

Type or paste your text into the search box and let Migratory Words find the most interesting language as it has evolved and changed over time

Read more ...

Type a phrase you want to investigate or paste a bulk text.

Show filters

Investigate

Top phrases

View all

Favorite phrases by users

- » Slash-and-burn and hit-and-run negative attack
- » Run roughshod over the constitution
- » Combating global warming
- » Moral and religious convictions

Council on Foreign Relations

- » Renewable-energy technologies
- » National-security information
- » Nationally-representative sample
- » Next-generation technologies
- » Intelligence-related activities
- » Information-technology industry

Cato Institute

- » Minority-serving institutions
- » Notice-and-comment procedures
- » Information-technology products
- » Energy-conservation measures
- » Small-government jurisdictions
- » Research-and-development spending

Fraser Institute

- » Full-time-equivalent employees
- » Administration environmental
- » Non-governmental organization
- » Regulatory responsibilities
- » Telecommunications networks
- » Environmental considerations

Disclaimer: The website runs best in Firefox, Chrome and Safari. We are working on making it Internet Explorer ready.

Future Potential

Studying the "Congressional Remix"

Social Network Integration, Twitter

Automated Content Gathering

Firefox Plug-in for Contextual Search

Semantic and Sentiment Analysis

Linguistic Analysis of Legal Text

Academic Ideological Mapping

ONGOING POLICY CHALLENGES: FAIR USE

Not for Profit

For Research Purposes
Public Service, Important for Discussion or Debate

Transformative Use

Caching

Link Back to Source

ONGOING POLICY CHALLENGES: TRANSPARENCY

"Targeted Transparency" - Kung et al
is not
"Naked Transparency" - Lessig

MigratoryWords.com

Questions?