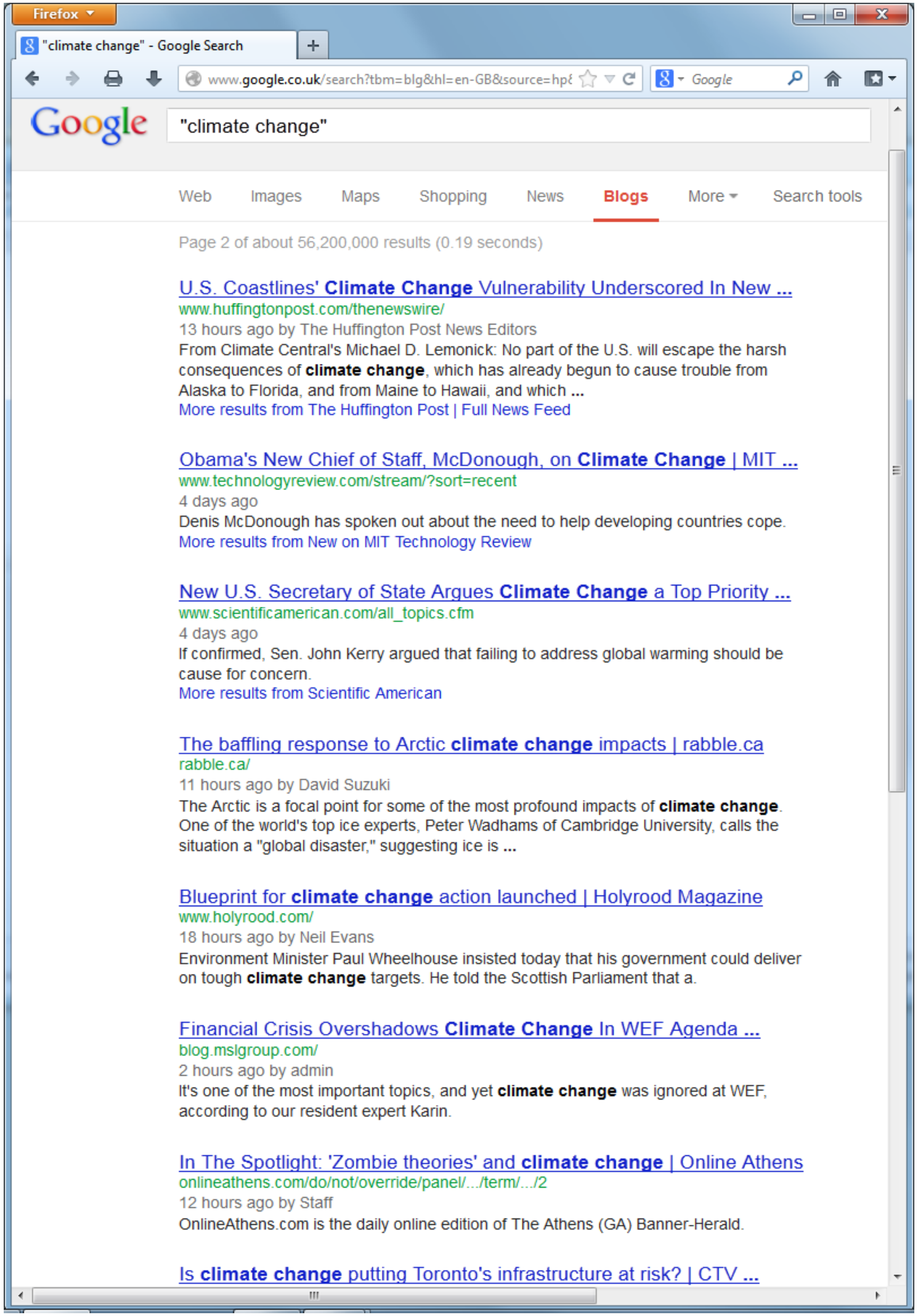
Key Statement Extraction in the NTAP project

Andrew Salway

Uni Computing, Bergen

NLP Group Seminar, University of Sheffield, 28th Feb 2013


Firefox
"climate change" - Google Search
www.google.co.uk/search?tbm=blg&hl=en-GB&source=hp&
Google
Google
"climate change"
Web
Images
Maps
Shopping
News
Blogs
More
Search tools

Page 2 of about 56,200,000 results (0.19 seconds)

U.S. Coastlines' **Climate Change** Vulnerability Underscored In New ...
www.huffingtonpost.com/thenewswire/
13 hours ago by The Huffington Post News Editors
From Climate Central's Michael D. Lemonick: No part of the U.S. will escape the harsh consequences of **climate change**, which has already begun to cause trouble from Alaska to Florida, and from Maine to Hawaii, and which **...**
More results from The Huffington Post | Full News Feed

Obama's New Chief of Staff, McDonough, on **Climate Change** | MIT ...
www.technologyreview.com/stream/?sort=recent
4 days ago
Denis McDonough has spoken out about the need to help developing countries cope.
More results from New on MIT Technology Review

New U.S. Secretary of State Argues **Climate Change** a Top Priority ...
www.scientificamerican.com/all_topics.cfm
4 days ago
If confirmed, Sen. John Kerry argued that failing to address global warming should be cause for concern.
More results from Scientific American

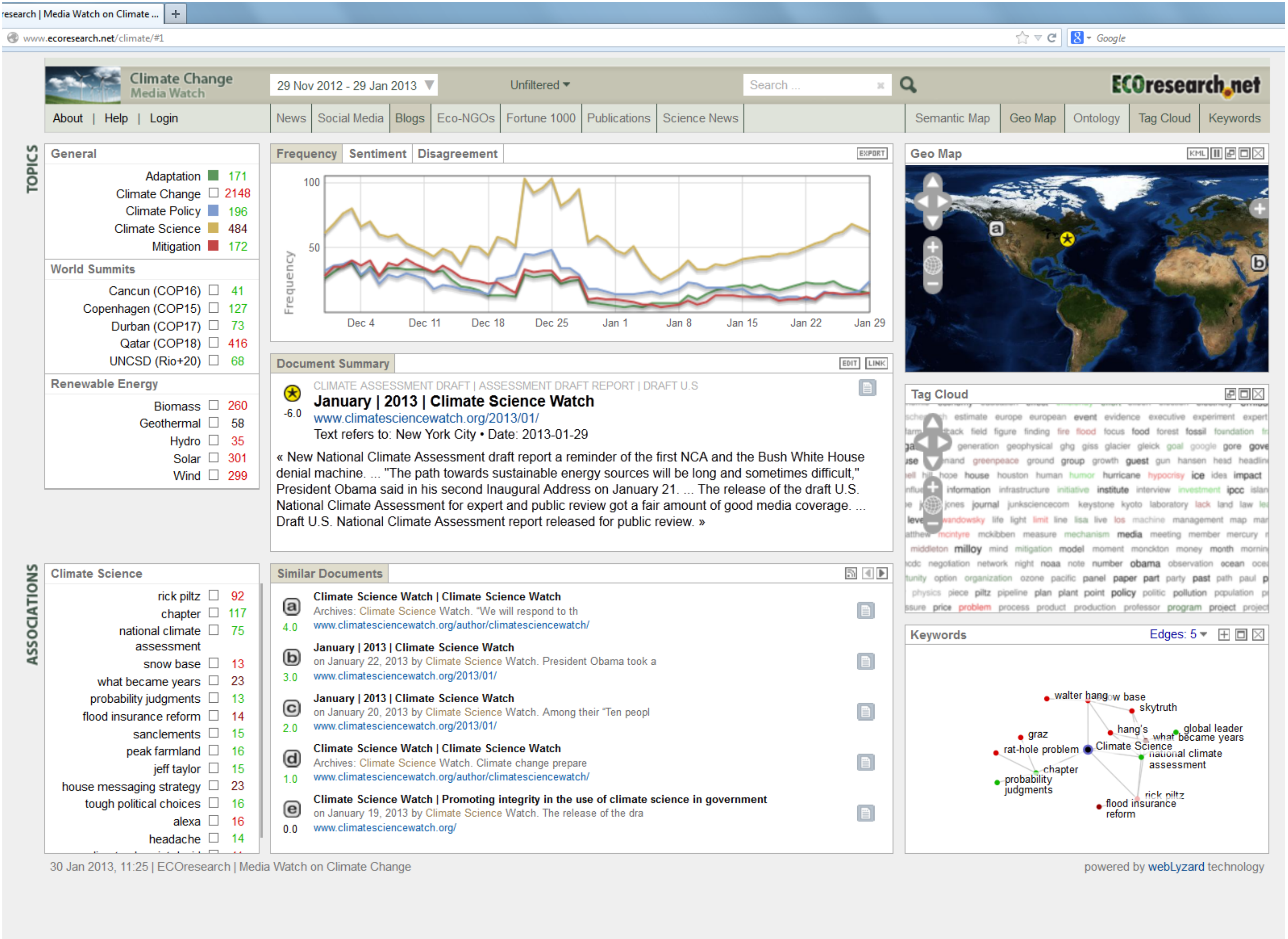
The baffling response to Arctic **climate change** impacts | rabble.ca
rabble.ca/
11 hours ago by David Suzuki
The Arctic is a focal point for some of the most profound impacts of **climate change**. One of the world's top ice experts, Peter Wadhams of Cambridge University, calls the situation a "global disaster," suggesting ice is **...**

Blueprint for **climate change** action launched | Holyrood Magazine
www.holyrood.com/
18 hours ago by Neil Evans
Environment Minister Paul Wheelhouse insisted today that his government could deliver on tough **climate change** targets. He told the Scottish Parliament that a.

Financial Crisis Overshadows **Climate Change** In WEF Agenda ...
blog.mslgroup.com/
2 hours ago by admin
It's one of the most important topics, and yet **climate change** was ignored at WEF, according to our resident expert Karin.

In The Spotlight: 'Zombie theories' and **climate change** | Online Athens
onlineathens.com/do/not/override/panel/.../term/.../2
12 hours ago by Staff
OnlineAthens.com is the daily online edition of The Athens (GA) Banner-Herald.

Is **climate change** putting Toronto's infrastructure at risk? | CTV ...



General

Adaptation	🟩	171
Climate Change	□	2148
Climate Policy	🟦	196
Climate Science	🟨	484
Mitigation	🟥	172

World Summits

Cancun (COP16)	□	41
Copenhagen (COP15)	□	127
Durban (COP17)	□	73
Qatar (COP18)	□	416
UNCSD (Rio+20)	□	68

Renewable Energy

Biomass	□	260
Geothermal	□	58
Hydro	□	35
Solar	□	301
Wind	□	299

Document Summary

CLIMATE ASSESSMENT DRAFT | ASSESSMENT DRAFT REPORT | DRAFT U.S

-6.0

January | 2013 | Climate Science Watch

www.climatesciencewatch.org/2013/01/

Text refers to: New York City • Date: 2013-01-29

« New National Climate Assessment draft report a reminder of the first NCA and the Bush White House denial machine. ... "The path towards sustainable energy sources will be long and sometimes difficult," President Obama said in his second Inaugural Address on January 21. ... The release of the draft U.S. National Climate Assessment for expert and public review got a fair amount of good media coverage. ... Draft U.S. National Climate Assessment report released for public review. »

Tag Cloud
estimate europe european event evidence executive experiment expert
field figure finding fire flood focus food forest fossil foundation
generation geophysical ghg giss glacier gleick goal google gore
greenpeace ground group growth guest gun hansen head
hope house houston human humor hurricane hypocrisy ice idea impact
information infrastructure initiative institute interview investment ipcc
jones journal junksciencecom keystone kyoto laboratory lack land law
life light limit line lisa live los machine management map
mcintyre mckibben measure mechanism media meeting member mercury
middleton milloy mind mitigation model moment monckton money month
negotiation network night noaa note number obama observation ocean
option organization ozone pacific panel paper part party past path paul
physics piece piltz pipeline plan plant point policy politic pollution population
price problem process product production professor program project

Climate Science

rick piltz	□	92
chapter	□	117
national climate assessment	□	75
snow base	□	13
what became years	□	23
probability judgments	□	13
flood insurance reform	□	14
sanclements	□	15
peak farmland	□	16
jeff taylor	□	15
house messaging strategy	□	23
tough political choices	□	16
alexa	□	16
headache	□	14

Similar Documents

(a) 4.0 **Climate Science Watch | Climate Science Watch**
Archives: Climate Science Watch. "We will respond to th
www.climatesciencewatch.org/author/climatesciencewatch/

(b) 3.0 **January | 2013 | Climate Science Watch**
on January 22, 2013 by Climate Science Watch. President Obama took a
www.climatesciencewatch.org/2013/01/

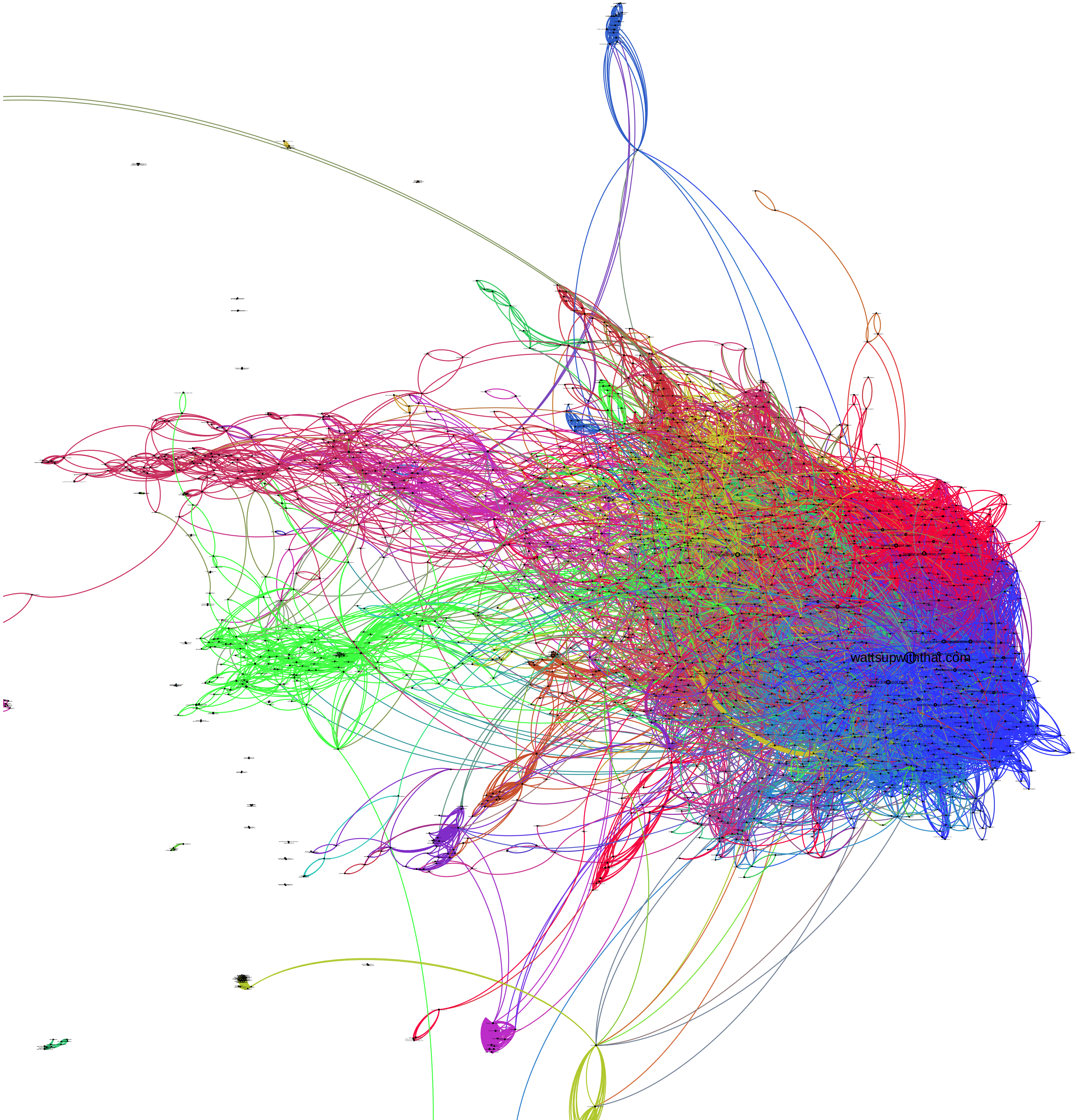
(c) 2.0 **January | 2013 | Climate Science Watch**
on January 20, 2013 by Climate Science Watch. Among their "Ten peopl
www.climatesciencewatch.org/2013/01/

(d) 1.0 **Climate Science Watch | Climate Science Watch**
Archives: Climate Science Watch. Climate change prepare
www.climatesciencewatch.org/author/climatesciencewatch/

(e) 0.0 **Climate Science Watch | Promoting integrity in the use of climate science in government**
on January 19, 2013 by Climate Science Watch. The release of the dra
www.climatesciencewatch.org/

30 Jan 2013, 11:25 | ECOresearch | Media Watch on Climate Change

powered by webLyzard technology



itsgettinghotinhere.org
rabett.blogspot.com
tamino.wordpress.com
bravenewclimate.com
wattsupwiththat.com
motls.blogspot.com
junkscience.com
tomnelson.blogspot.com

NTAP: Networks of Texts and People

NTAP is developing methods to analyse and visualize the social and epistemological contexts of information contained in blogs.

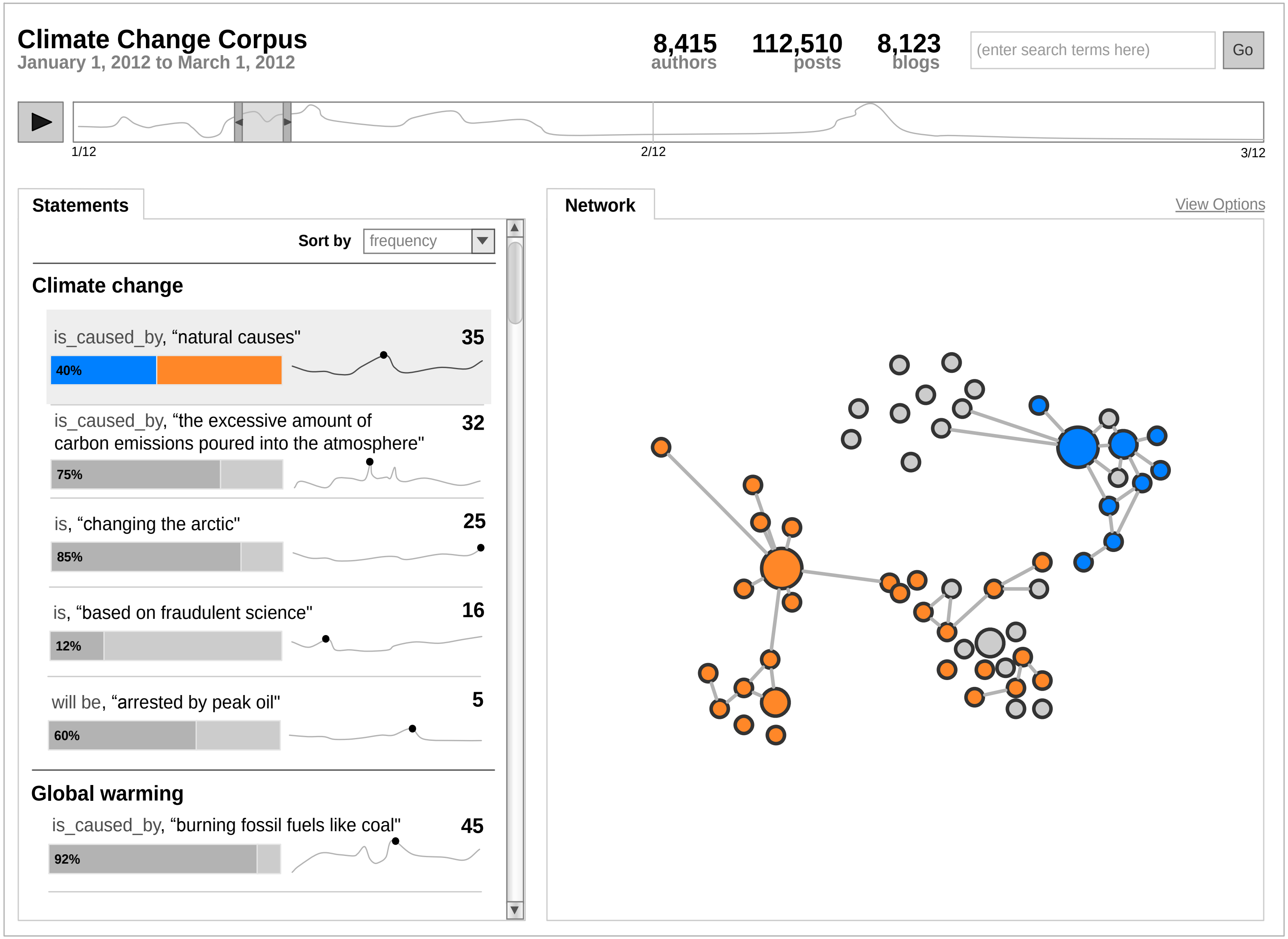
- Multidisciplinary:
  - information and media science
  - natural language processing
  - network analysis
  - information visualization
- Collaboration between University of Bergen and Uni Computing
- Funded by Research Council of Norway, 2012-2015
- Project team: Nick Diakopoulos, Dag Elgesem, Knut Hofland, Andrew Salway, Lubos Steskal, Samia Touileb
- Associated partners: University of Sheffield, OCAD University, Retriever

www.ntap.no

Our Approach

- Focus on helping users to understand the phenomena of **information diffusion** and **polarisation**
- Represent blog content as **key statements**
- Use visual analytic tools to **integrate data** about when and where key statements occur with data about the structure of blog networks

(climate change, is_caused_by, “burning fossil fuels like coal”)



Climate Change Corpus

January 1, 2012 to March 1, 2012

8,415 authors

112,510 posts

8,123 blogs

Statements

Network

Sort by frequency

Climate change

Statement	Share	Count
is_caused_by, “natural causes"	40%	35
is_caused_by, “the excessive amount of carbon emissions poured into the atmosphere"	75%	32
is, “changing the arctic"	85%	25
is, “based on fraudulent science"	12%	16
will be, “arrested by peak oil"	60%	5

Global warming

Statement	Share	Count
is_caused_by, “burning fossil fuels like coal"	92%	45

Information Extraction from the Web

- Learn relation templates from small set of exemplar tuples, e.g. (England, London), (France, Paris), etc. One million facts extracted from the web at levels of precision between 75-98% (Pasca et al., 2006)
- Use deeper syntactic parsing to discover both relations and facts which are somewhat richer than simple tuples (Banko and Etzioni, 2008; Etzioni et al., 2008).

Fact Extraction to Enhance Image Captions
(Salway et al 2010)

Using GPS and other camera data, and geographic databases →
“A view of the Eiffel Tower”

Add fact extraction from the web →
“A view of the Eiffel Tower, which was built in 1889 for an international exhibition in Paris.”

1: Get Snippets from Search Engine

- Form queries around the given entity with a set of cues, e.g.

  “Eiffel Tower is famous for”

  “Eiffel Tower was built”

  “Eiffel Tower is popular with”

  ...

- Get up to 50 snippets per query (Yahoo BOSS API); same cue can get very different kinds of information

  “Paris tourism guide from Yahoo! Travel UK. Recommended Paris tours and Paris ... While the **Eiffel Tower is famous for** its views of the city, the Arc de Triomphe ...”

2: Chunk Snippets

- Use a single regular expression to chunk...

‘`BOUNDARY ENTITY CUE` ***`TEXT-FRAGMENT`*** `BOUNDARY`’

✓ “...in London. Big Ben was named after Sir Benjamin Hall.”

→

`(Big Ben, was named, after Sir Benjamin Hall)`

✘ “The square next to Big Ben was named in 1848...”

3: Filter Candidate Facts

- Subjective words: remove candidate facts containing words such as ‘me’, ‘my’, ‘amazing’, etc.
- Invalid end words: removes candidate facts ending with ‘to’, ‘from’, ‘by’, etc.
- Minimum number of words
- Words all in capitals

4(i): Score and Rank Facts

For each Entity-Cue pair, count frequencies of words in text fragments (remove stop words).

```
(Eiffel Tower, was built)

      Paris          7
      1889      5
      exhibition4
      ...
```

4(ii): Score and Rank Facts

For each fact, sum the frequencies of each word it contains. (Optionally, divide by number of words in fact).

`(Eiffel Tower, was built, in Paris in 1889)`

→ 7 + 5 = 12

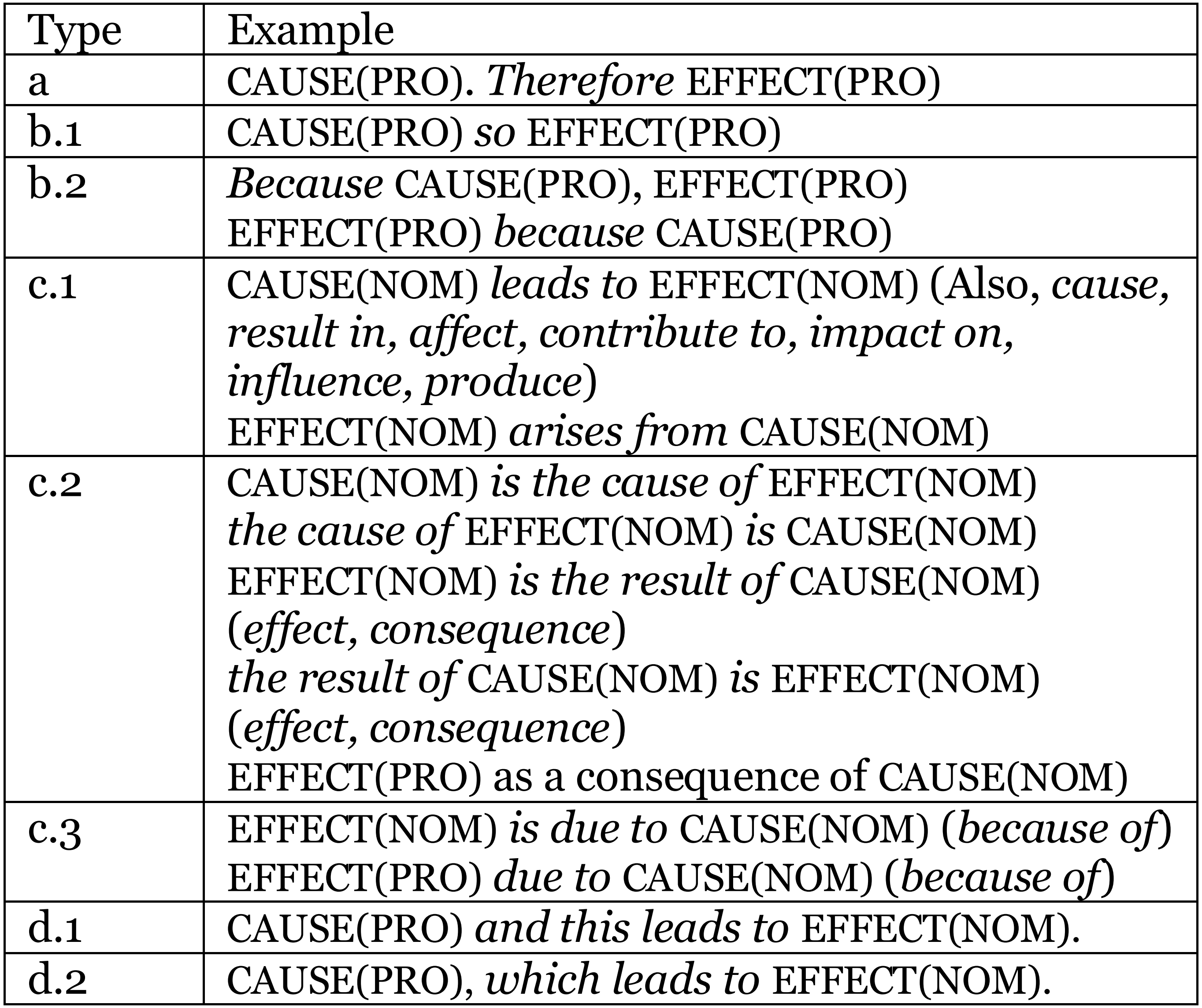
or, (7+5)/4 = 3

```
(Eiffel Tower, was built, in 1889 for an international
  exhibition in Paris)
(Eiffel Tower, was named, after an ingenious engineer
  whose design of the tower turned it into a reality and
  pride of the French nation)
(Eiffel Tower, is, an iron tower built during 1887-1889 on
  the Champ de Mars beside the Seine River in Paris)
(Eiffel Tower, was one of, the first tall structures in
  the world to contain passenger elevators)
(Eiffel Tower, was one of, the landmarks visited by Luigi
  when he came to save Paris from invading Koopa Troopas)


...


(Eiffel Tower, is made, from 18)
(Eiffel Tower, is made, of 3 platforms)
(Eiffel Tower, is made, with 2)
(Eiffel Tower, is famous, throughout the world)
(Eiffel Tower, is famous, for a reason)
```

```
is_the_result_of, “natural causes”
is_the_result_of, “market failure”
is_the_result_of, “something called the greenhouse effect”
is_the_result_of, “natural fluctuations”
is_the_result_of, “man-made activities”
is_caused_by, “climate-control and other energy-intensive practices”
is_caused_by, “mankind’s carbon and other gases output”
is_caused_by, “overpopulation”
is_caused_by, “long-range planetary trends”
is_caused_by, “geological factors”
is_caused_by, “burning of fossil fuels like coal”
is_caused_by, “the excessive amount of carbon emissions poured into the atmosphere”
is_caused_by, “sunspots and cows burping”
is_caused_by, “the cyclical element of nature itself”
will_be, “most severe in Africa and South Asia”
will_be, “one of major reasons for migration across Asia in the years to come”
```



Type	Example
a	CAUSE(PRO). *Therefore* EFFECT(PRO)
b.1	CAUSE(PRO) *so* EFFECT(PRO)
b.2	*Because* CAUSE(PRO), EFFECT(PRO) EFFECT(PRO) *because* CAUSE(PRO)
c.1	CAUSE(NOM) *leads to* EFFECT(NOM) (Also, *cause, result in, affect, contribute to, impact on, influence, produce*) EFFECT(NOM) *arises from* CAUSE(NOM)
c.2	CAUSE(NOM) *is the cause of* EFFECT(NOM) *the cause of* EFFECT(NOM) *is* CAUSE(NOM) EFFECT(NOM) *is the result of* CAUSE(NOM) (*effect, consequence*) *the result of* CAUSE(NOM) *is* EFFECT(NOM) (*effect, consequence*) EFFECT(PRO) as a consequence of CAUSE(NOM)

c.3	EFFECT(NOM) *is due to* CAUSE(NOM) (*because of*)
	EFFECT(PRO) *due to* CAUSE(NOM) (*because of*)

d.1	CAUSE(PRO) *and this leads to* EFFECT(NOM).
d.2	CAUSE(PRO), *which leads to* EFFECT(NOM).

Derived from Halliday and Matthiessen's (2006) semantic analysis of the sequence and their grammatical analysis of the clause and the clause complex (Halliday and Matthiessen 2004).

```
:: TOPIC PROBABILITY VERB FILLER

PROBABILITY==NULL, certainly, probably, definitely

VERB==causes, leads to, results in, affects,
contributes to, impacts on, influences, produces
```

- Generated 2391 realizations of causality.
    - Most were with TOPIC expressed as a nominal phrase.
    - Most were of our type c, i.e. verbs in active and passive forms, and in present, present continuous, present perfect and future tenses.
- These were combined with different ways to refer to the phenomenon of climate change, giving 14,393 queries.

Nominals	"climate change", "global warming", "the changing climate", "the earth warming", "the fact that the climate is changing", "the way in which the climate is changing"
Processes	"climate change is getting worse", "climate change is happening", "global warming is happening", "the climate is changing", "the earth is getting warmer", "the earth is warmer", "the rate of climate change is increasing"

Query	Hits	Query	Hits
cause(s) climate change	1000	contribute(s) to global warming	1000
contribute(s) to climate change	1000	is/are causing global warming	1000
impact(s) on climate change	472	cause(s) global warming	999
is/are causing climate change	342	global warming is caused by	695
climate change is caused by	321	impact(s) on global warming	341
affect(s) climate change	268	is/are contributing to global warming	295
is/are contributing to climate change	165	lead(s) to global warming	260
lead(s) to climate change	116	affect(s) global warming	121
influence(s) climate change	109	global warming is due to	117
climate change is due to	74	the cause of global warming is	103
TOTAL	3867	TOTAL	4931

For the four combinations of “climate change” and “global warming”, as CAUSE and as EFFECT, the top 10 queries account for more than 50% of all hits; the top 30 queries account for more than 90% of all hits.

cause, is causing, (which) is caused by, is being caused by, has caused, (which) has been caused by, will cause

affect, is affecting, is affected by, has affected, will affect

lead to, has led to, is leading to, will lead to

produce, has produced, is producing, will produce

result in, has resulted in, is resulting in, will result in

contribute to, has contributed to, is contributing to

impact on, is impacted by

influence, is influencing

arise from

as a consequence of, because of, due to

is|are due to

is|are the main cause(s) of

... **that the increasing concentration of carbon dioxide in the atmosphere** leads to climate change” Martin Rees, president of Britain’s Royal Society, said in a press release...]

... I believe climate change is due to **a combination of factors, including natural cycles, sun spots, and human activity.** But scientists still don't know for certain how much each of these factors contributes to the overall climate ...

	“climate change”	“global warming”
The causes of...	2564 fragments 21,727 tokens	6713 fragments 28,593 tokens
The effects of...	4072 fragments 33,248 tokens	3816 fragments 30,101 tokens

Processing of text fragments

1) Get top 50 keywords from each file (log-likelihood w.r.t. BNC)
2) Get n-grams (2<=n<=6, frequency > 20)
3) Merge results for “climate change” and “global warming”
4) From keyword lists, remove stop words (“that”, “don’t”, …), words that appear in n-grams (“emissions”, “extreme”, …), and words that appear in the topics (“global”, “climate”)
5) From n-grams, remove any containing stop words and any sub-strings, e.g. “weather events” where “extreme weather events” is also present

Extracted causes of cc/gw

(humans, mankind, man, human activity, human activities, human beings), (pollution, pollutants), (carbon emissions, CO2 emissions, carbon dioxide emissions), (GHG, greenhouse gases, greenhouse gas emissions), (fossil fuels, oil, coal), methane, deforestation

actions (human actions), behavior (human behavior), burning (burning fossil fuels), heat (heat trapping gases), natural (natural processes, natural cycles), solar (solar variation, solar activity)

reduce, reducing, energy, atmosphere, scientists, believe, Academy, Sun, evidence, percent, scientific consensus, man made

Extracted effects of cc/gw

(extreme weather events, disasters), (flooding, floods), (storms, hurricanes, Hurricane Katrina), Blackouts, (drought, droughts), rising sea levels, extinction

temperatures (rising temperatures), food (food shortages), biodiversity (biodiversity loss), winters (cold winters), glaciers (glaciers melting)

species, Arctic, rise, oceans, ecosystems, rainfall, frequent, snow, warmer, world, world's, future, planet, increase, people, frequent, extinct, Gore, melt, water resources, sea ice, polar bears, ice caps, weather patterns

Ongoing work

- Processing 3,000 crawled blogs, to extract date, author, text content and links
- A semi-automatic tool to create groups of related statements
- The induction of local grammars from n-grams around “climate change”, as a basis for statement extraction templates

For discussion…

- Sufficient regularity / repetition in how statements about climate change are expressed for portable key statement extraction?
  - What about statements about mitigation policy, etc.?
- What to do with key statements about the causes and effects of climate change?
  - Classify blog posts and bloggers
  - Analyse patterns of information diffusion and polarisation
- How to define and automatically identify “relatedness” of statements?
  - Exploit semi-structure of key statements?
  - Paraphrase / entailment detection?
  - Evidence of relatedness from social network structures?

Androutsopoulos, I. and Malakasiotis, P. (2010). A Survey of Paraphrasing and Textual Entailment Methods. *Journal of Artificial Intelligence Research* 38: 135-187.

Banko, M. and Etzioni, O. (2008). The Tradeoffs Between Open and Traditional Relation Extraction. *Procs. ACL 2008*: 28–36.

Etzioni, O. et al. (2008). Open Information Extraction from the Web. *Communications of the ACM,* 51(12): 68–74.

Halliday, M. A. K. and Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar* (3rd edition). Hodder Education, London.

Halliday, M.A.K. and Matthiessen, C. M. I. M. (2006). *Construing Experience Through Meaning: A Language-Based Approach to Cognition* (study edition). Continuum, London.

Mesquita, F. and Barbosa, D. (2011). Extracting Meta Statements from the Blogosphere. In Procs. International Conference on Weblogs and Social Media 2011.

Pasca, M. et al. (2006). Organizing and Searching the World Wide Web of Facts - Step One: the One-Million Fact Extraction Challenge. *Procs. AAAI*: 1400–1405.

Salway, A., Kelly, L., Skadiņa, I. and Jones, G. (2010). Portable Extraction of Partially Structured Facts from the Web. In H. Loftsson, E. Rögnvaldsson and S. Helgadóttir (eds.). *Lecture Notes in Artificial Intelligence 6233*. Springer, Heidelberg.

Solan, Z., Horn, D., Ruppin, E. and Edelman, S. (2005). Unsupervised learning of natural languages. *Procs. of the National Academy of Sciences* 102(33): 11629-11634.

van Zaanen, M. (2000). ABL: Alignment-Based Learning. *Procs. 18th COLING*: 961–967.