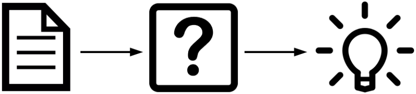
Gavagai

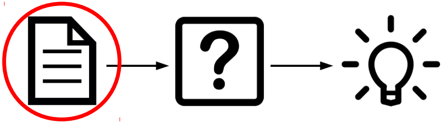
Text Analysis for Big Data

Magnus Sahlgren

Text analysis



Text analysis



Text analysis

Data

Size

Style
(editorial vs social)

Language
(there are other languages than English out there!)

Text analysis

Data

Size

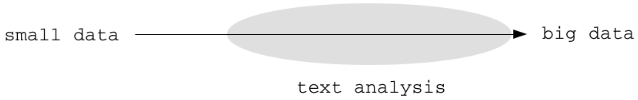
Style
(editorial vs social)

Language
(there are other languages than English out there!)

Text analysis

Data

Size



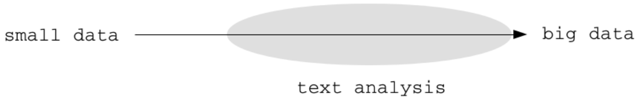
Style
(editorial vs social)

Language
(there are other languages than English out there!)

Text analysis

Data

Size



Style

(editorial vs social)

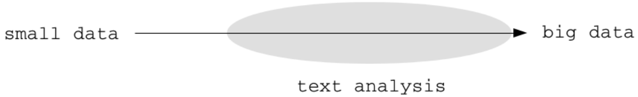
Language

(there are other languages than English out there!)

Text analysis

Data

Size



small data

big data

text analysis

Style

(editorial vs social)

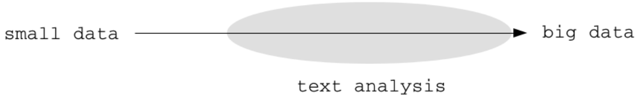
Language

(there are other languages than English out there!)

Text analysis

Data

Size



Style

(editorial vs social)

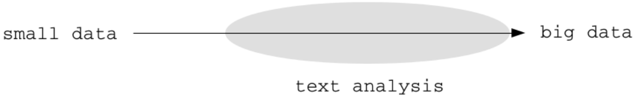
Language

(there are other languages than English out there!)

Text analysis

Data

Size



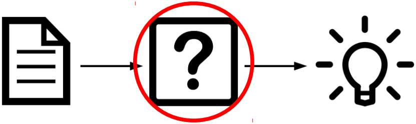
Style

(editorial vs social)

Language

(there are other languages than English out there!)

Text analysis



Gavagai

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

Text analysis

Technologies

Knowledge-based
(use resources like Wikipedia)

Supervised machine learning
(use annotated data)

Unsupervised machine learning
(use unstructured data)

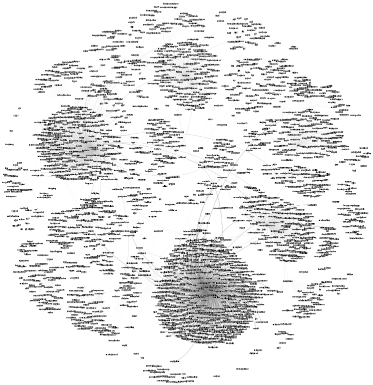
Semantic memories

Semantic memories

(systems that learn language by reading large amounts of text)

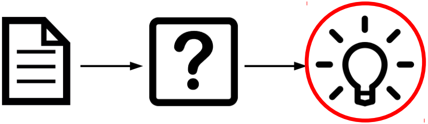
Semantic memories

(systems that learn language by reading large amounts of text)



Gavagai

Text analysis



Gavagai

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Text analysis

Insights

Identify and extract items

(e.g. entities and events)

Find relations

(e.g. synonyms and associations)

Text analysis

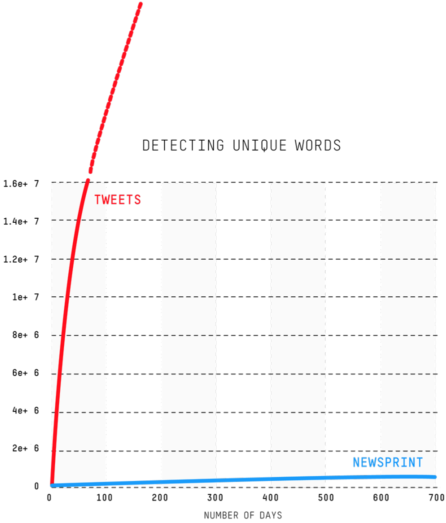
Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

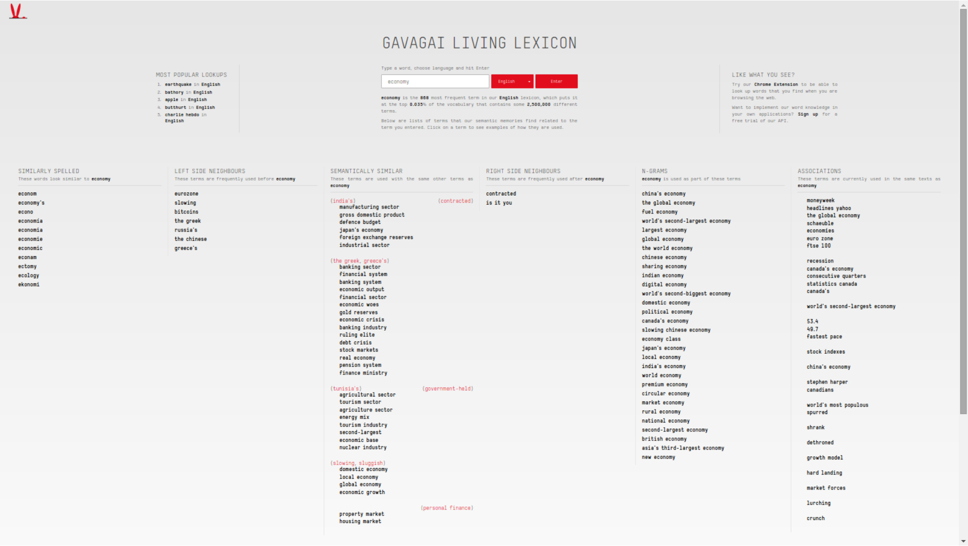
Find relations

Find relations



Gavagai

Find relations (lexicon.gavagai.se)



GAVAGAI LIVING LEXICON

MOST POPULAR LOOKUPS

1. **earthquake** in **English**
2. **bathory** in **English**
3. **apple** in **English**
4. **butthurt** in **English**
5. **charlie hebdo** in **English**

Type a word, choose language and hit Enter

economy | English | Enter

economy is the **868** most frequent term in our **English** lexicon, which puts it at the top **0.035%** of the vocabulary that contains some **2,500,000** different terms.

Below are lists of terms that our semantic memories find related to the term you entered. Click on a term to see examples of how they are used.

LIKE WHAT YOU SEE?

Try our **Chrome Extension** to be able to look up words that you find when you are browsing the web.

Want to implement our word knowledge in your own applications? **Sign up** for a free trial of our API.

SIMILARLY SPELLED

These words look similar to **economy**

econom
economy's
econo
economia
economia
economie
economic
econam
ectomy
ecology
ekonomi

LEFT SIDE NEIGHBOURS

These terms are frequently used before **economy**

eurozone
slowing
bitcoins
the greek
russia's
the chinese
greece's

SEMANTICALLY SIMILAR

These terms are used with the same other terms as **economy**

(india's) (contracted)
manufacturing sector
gross domestic product
defence budget
japan's economy
foreign exchange reserves
industrial sector

(the greek, greece's)
banking sector
financial system
banking system
economic output
financial sector
economic woes
gold reserves
economic crisis
banking industry
ruling elite
debt crisis
stock markets
real economy
pension system
finance ministry

(tunisia's) (government-held)
agricultural sector
tourism sector
agriculture sector
energy mix
tourism industry
second-largest
economic base
nuclear industry

(slowing, sluggish)
domestic economy
local economy
global economy
economic growth

(personal finance)
property market
housing market

RIGHT SIDE NEIGHBOURS

These terms are frequently used after **economy**

contracted
is it you

N-GRAMS

economy is used as part of these terms

china's economy
the global economy
fuel economy
world's second-largest economy
largest economy
global economy
the world economy
chinese economy
sharing economy
indian economy
digital economy
world's second-biggest economy
domestic economy
political economy
canada's economy
slowing chinese economy
economy class
japan's economy
local economy
india's economy
world economy
premium economy
circular economy
market economy
rural economy
national economy
second-largest economy
british economy
asia's third-largest economy
new economy

ASSOCIATIONS

These terms are currently used in the same texts as **economy**

moneyweek
headlines yahoo
the global economy
schaeuble
economies
euro zone
ftse 100

recession
canada's economy
consecutive quarters
statistics canada
canada's

world's second-largest economy

53.4
49.7
fastest pace

stock indexes

china's economy

stephen harper
canadians

world's most populous
spurred

shrank

dethroned

growth model

hard landing

market forces

lurching

crunch

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Compress and refine the information
(e.g. summarization and topic detection)

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Compress and refine the information
(e.g. summarization and topic detection)

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

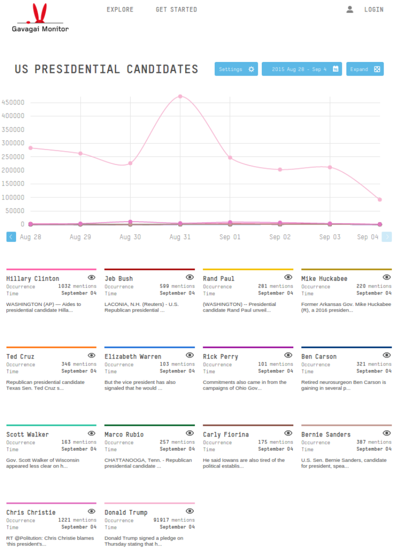
Compress and refine the information
(e.g. summarization and topic detection)

Compress and refine

Summarization and topic detection

Compress and refine (monitor.gavagai.se)

Summarization and topic detection



US PRESIDENTIAL CANDIDATES

Hillary Clinton
Occurrence 1032 mentions
Time September 04
WASHINGTON (AP) — Aides to presidential candidate Hilla...

Jeb Bush
Occurrence 599 mentions
Time September 04
LACONIA, N.H. (Reuters) - U.S. Republican presidential ...

Rand Paul
Occurrence 281 mentions
Time September 04
(WASHINGTON) -- Presidential candidate Rand Paul unveil...

Mike Huckabee
Occurrence 220 mentions
Time September 04
Former Arkansas Gov. Mike Huckabee (R), a 2016 presiden...

Ted Cruz
Occurrence 346 mentions
Time September 04
Republican presidential candidate Texas Sen. Ted Cruz s...

Elizabeth Warren
Occurrence 103 mentions
Time September 04
But the vice president has also signaled that he would ...

Rick Perry
Occurrence 101 mentions
Time September 04
Commitments also came in from the campaigns of Ohio Gov...

Ben Carson
Occurrence 321 mentions
Time September 04
Retired neurosurgeon Ben Carson is gaining in several p...

Scott Walker
Occurrence 163 mentions
Time September 04
Gov. Scott Walker of Wisconsin appeared less clear on h...

Marco Rubio
Occurrence 257 mentions
Time September 04
CHATTANOOGA, Tenn. - Republican presidential candidate ...

Carly Fiorina
Occurrence 175 mentions
Time September 04
He said Iowans are also tired of the political establis...

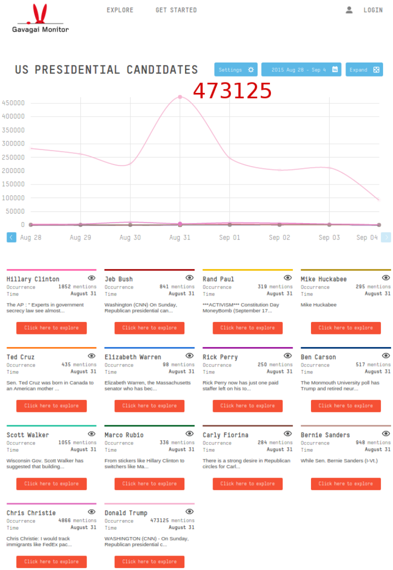
Bernie Sanders
Occurrence 387 mentions
Time September 04
U.S. Sen. Bernie Sanders, candidate for president, spea...

Chris Christie
Occurrence 1221 mentions
Time September 04
RT @Politution: Chris Christie blames 'this president's...

Donald Trump
Occurrence 91917 mentions
Time September 04
Donald Trump signed a pledge on Thursday stating that h...

Compress and refine (`monitor.gavagai.se`)

Summarization and topic detection



Hillary Clinton
Occurrence 1852 mentions
Time August 31
The AP : " Experts in government secrecy law see almost...
Click here to explore

Jeb Bush
Occurrence 841 mentions
Time August 31
Washington (CNN) On Sunday, Republican presidential can...
Click here to explore

Rand Paul
Occurrence 319 mentions
Time August 31
ACTIVISM Constitution Day MoneyBomb (September 17...
Click here to explore

Mike Huckabee
Occurrence 295 mentions
Time August 31
Mike Huckabee
Click here to explore

Ted Cruz
Occurrence 435 mentions
Time August 31
Sen. Ted Cruz was born in Canada to an American mother ...
Click here to explore

Elizabeth Warren
Occurrence 98 mentions
Time August 31
Elizabeth Warren, the Massachusetts senator who has bec...
Click here to explore

Rick Perry
Occurrence 250 mentions
Time August 31
Rick Perry now has just one paid staffer left on his lo...
Click here to explore

Ben Carson
Occurrence 517 mentions
Time August 31
The Monmouth University poll has Trump and retired neur...
Click here to explore

Scott Walker
Occurrence 1055 mentions
Time August 31
Wisconsin Gov. Scott Walker has suggested that building...
Click here to explore

Marco Rubio
Occurrence 336 mentions
Time August 31
From stickers like Hillary Clinton to switchers like Ma...
Click here to explore

Carly Fiorina
Occurrence 284 mentions
Time August 31
There is a strong desire in Republican circles for Carl...
Click here to explore

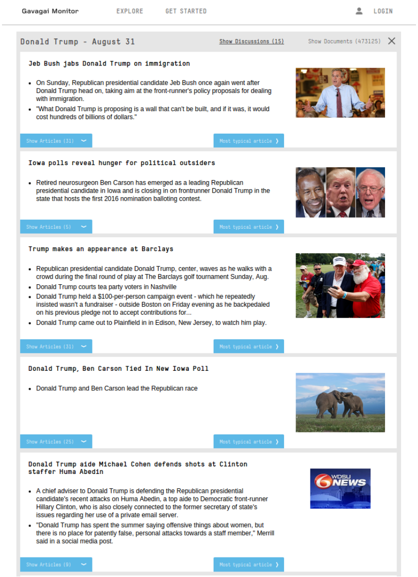
Bernie Sanders
Occurrence 948 mentions
Time August 31
While Sen. Bernie Sanders (I-Vt.)
Click here to explore

Chris Christie
Occurrence 4866 mentions
Time August 31
Chris Christie: I would track immigrants like FedEx pac...
Click here to explore

Donald Trump
Occurrence 473125 mentions
Time August 31
WASHINGTON (CNN) - On Sunday, Republican presidential c...
Click here to explore

Compress and refine (`monitor.gavagai.se`)

Summarization and topic detection



Donald Trump - August 31

Jeb Bush jabs Donald Trump on immigration

- On Sunday, Republican presidential candidate Jeb Bush once again went after Donald Trump head on, taking aim at the front-runner's policy proposals for dealing with immigration.
- "What Donald Trump is proposing is a wall that can't be built, and if it was, it would cost hundreds of billions of dollars."

Iowa polls reveal hunger for political outsiders

- Retired neurosurgeon Ben Carson has emerged as a leading Republican presidential candidate in Iowa and is closing in on frontrunner Donald Trump in the state that hosts the first 2016 nomination balloting contest.

Trump makes an appearance at Barclays

- Republican presidential candidate Donald Trump, center, waves as he walks with a crowd during the final round of play at The Barclays golf tournament Sunday, Aug.
- Donald Trump courts tea party voters in Nashville
- Donald Trump held a $100-per-person campaign event - which he repeatedly insisted wasn't a fundraiser - outside Boston on Friday evening as he backpedaled on his previous pledge not to accept contributions for...
- Donald Trump came out to Plainfield in in Edison, New Jersey, to watch him play.

Donald Trump, Ben Carson Tied In New Iowa Poll

- Donald Trump and Ben Carson lead the Republican race

Donald Trump aide Michael Cohen defends shots at Clinton staffer Huma Abedin

- A chief adviser to Donald Trump is defending the Republican presidential candidate's recent attacks on Huma Abedin, a top aide to Democratic front-runner Hillary Clinton, who is also closely connected to the former secretary of state's issues regarding her use of a private email server.
- "Donald Trump has spent the summer saying offensive things about women, but there is no place for patently false, personal attacks towards a staff member," Merrill said in a social media post.

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Compress and refine the information
(e.g. summarization and topic detection)

Measure things
(e.g. attitudes and opinions)

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Compress and refine the information
(e.g. summarization and topic detection)

Measure things
(e.g. attitudes and opinions)

Text analysis

Insights

Identify and extract items
(e.g. entities and events)

Find relations
(e.g. synonyms and associations)

Compress and refine the information
(e.g. summarization and topic detection)

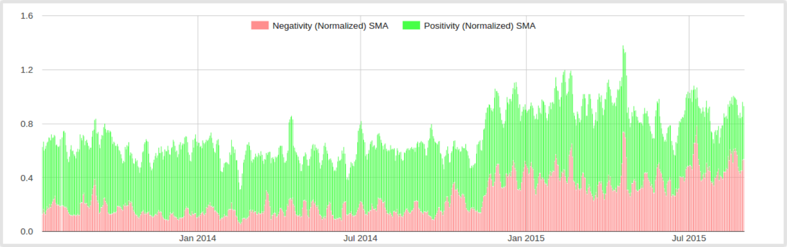
Measure things
(e.g. attitudes and opinions)

Measure

Sentiment analysis

Measure

Sentiment analysis

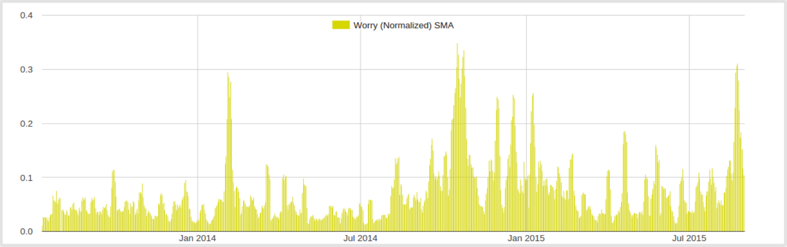


Positivity vs negativity wrt the global economy in English online media

Gavagai

Measure

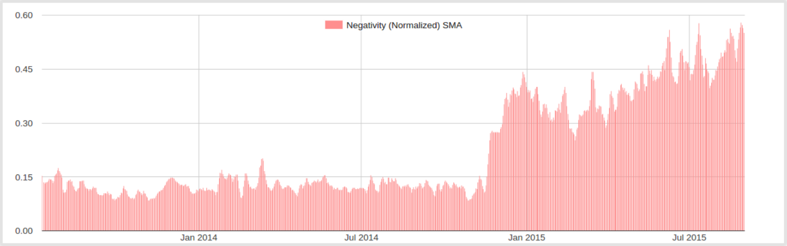
Sentiment analysis



Worry wrt the global economy in English online media

Measure

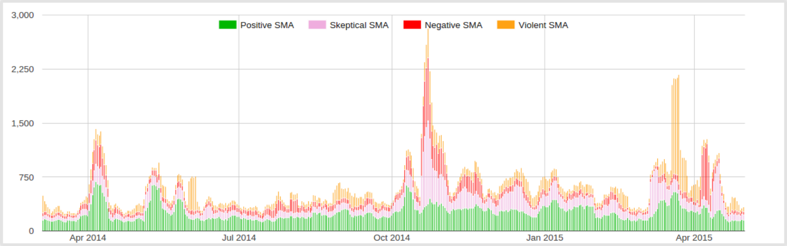
Sentiment analysis



Negativity towards China in English online media

Measure

Sentiment analysis



Attitude towards Sweden in Russian online media

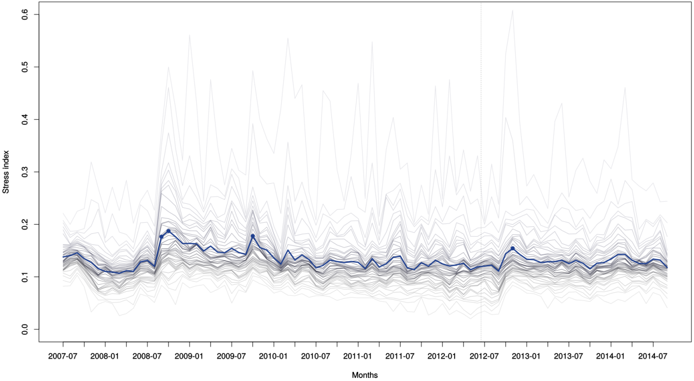
Gavagai

Measure

Predict

Measure

Predict



Rönnqvist & Sarlin (2015): Detect & Describe: deep learning of bank stress in the news