

Text Analytics for Semantic Search

Dr Diana Maynard
University of Sheffield, UK

Examples: Government Web Archive

- We developed a semantic annotation application to process every crawled page in the archive.
- Entities annotated included; people, companies, locations, government departments, ministerial positions, social documents, dates, money....
- Where possible, annotations were linked to an ontology which
 - was based on DBpedia
 - was extended with UK government-specific concepts
 - included the modelling of the evolution of government
- Annotations were indexed to allow for complex semantic querying of the collection

Examples: Political Futures Tracker

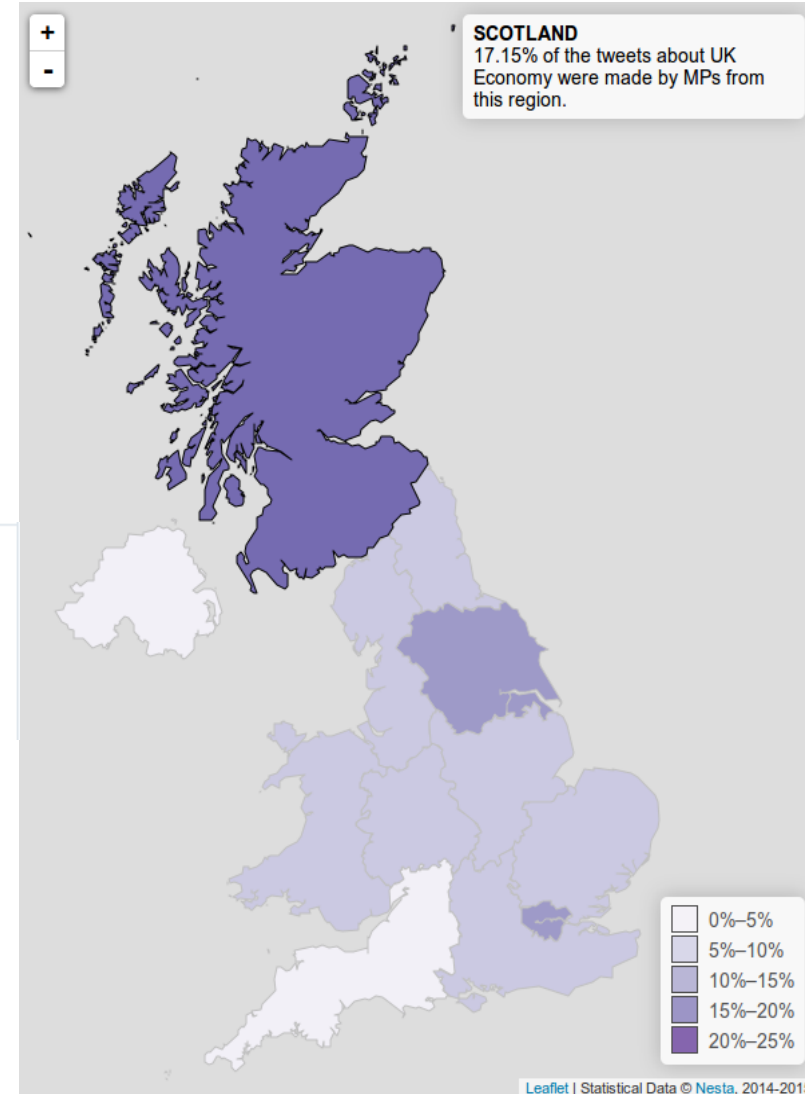
Where in the UK did Conservative MPs tweet more about the economy?



Richard Short @ToryShorty · May 28

I spoke up for the living wage today and a work colleague said he thought I was a Tory. He obviously hasn't heard of [#WorkingClassTories](#)

↩️ ↻ 2 ★ 5 + 👤 ⋮



GATE Mímir

- can be used to index and search over text, annotations, semantic metadata (concepts and instances)
- allows queries that arbitrarily mix full-text, structural, linguistic and semantic annotations
- is open source

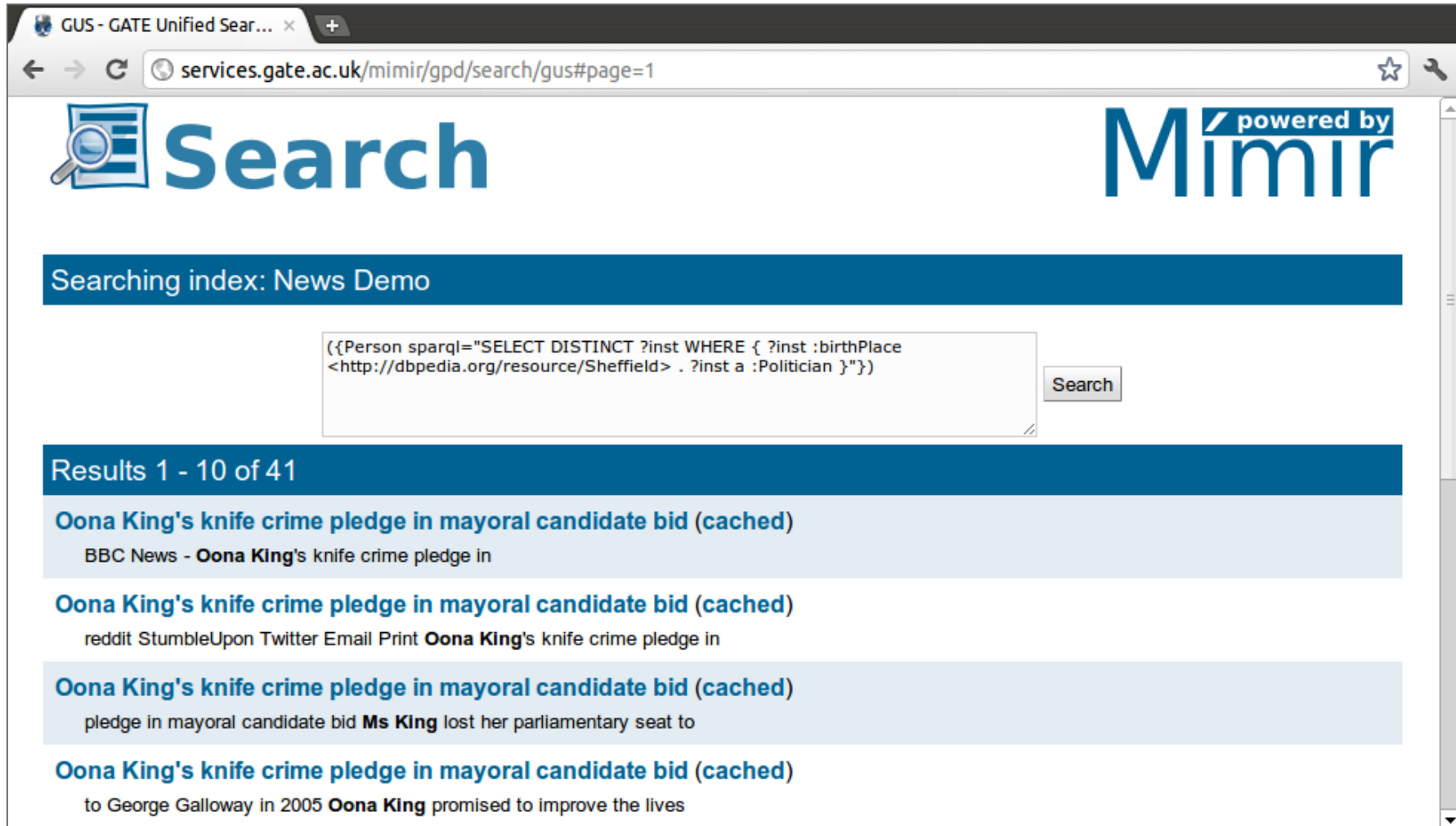


What can GATE Mimir do that Google can't?

Show me:

- all documents mentioning a temperature between 30 and 90 degrees F (expressed in any unit)
- all abstracts written in French on patent documents from the last 6 months which mention any form of the word “transistor” in the English abstract
- the names of the patent inventors of those abstracts
- all documents mentioning steel industries in the UK, along with their location

Search News Articles for Politicians born in Sheffield



The screenshot shows a web browser window with the address bar displaying `services.gate.ac.uk/mimir/gpd/search/gus#page=1`. The page features a large "Search" header with a magnifying glass icon and the text "powered by Mimir". Below this, a blue bar indicates "Searching index: News Demo". A search query is entered in a text box: `{Person sparql="SELECT DISTINCT ?inst WHERE { ?inst :birthPlace <http://dbpedia.org/resource/Sheffield> . ?inst a :Politician }"} }`. A "Search" button is located to the right of the query box. The results section, titled "Results 1 - 10 of 41", displays four search results, each with a title and a snippet of text:

- Oona King's knife crime pledge in mayoral candidate bid (cached)**
BBC News - **Oona King's** knife crime pledge in
- Oona King's knife crime pledge in mayoral candidate bid (cached)**
reddit StumbleUpon Twitter Email Print **Oona King's** knife crime pledge in
- Oona King's knife crime pledge in mayoral candidate bid (cached)**
pledge in mayoral candidate bid **Ms King** lost her parliamentary seat to
- Oona King's knife crime pledge in mayoral candidate bid (cached)**
to George Galloway in 2005 **Oona King** promised to improve the lives

<http://demos.gate.ac.uk/mimir/gpd/search/gus>

Easily Create Your Own Custom GATE Mimir Interfaces

People in the News

demos.gate.ac.uk/pin/?name=&bornIn=Sheffield&famousAs=Politician|OfficeHolder&after=01%2F04%2F2011&before=30%2F04%2F2011

PEOPLE IN THE NEWS

Looking For...

Name:

Fuzzy Name Matching ☐

Born In:

Famous As:

In Articles...

Published Between and

Classified As:

Ignore Boilerplate Text ☒

Search

Results 1 to 2 of 2

Show Underlying Mimir Query

[Scottish election: Respect Coalition Against Cuts profile](http://www.bbc.co.uk/news/uk-scotland-13048761)
<http://www.bbc.co.uk/news/uk-scotland-13048761> - [Cached](#)
... Bow - whose sitting MP Oona King had voted for the war ...
... success came when Galloway overturned Oona King's 10,000- ...

Powered by GATE Mimir
© The University of Sheffield, 2011

TRY IT!

<http://demos.gate.ac.uk/pin/>

MIMIR: Searching Text Mining Results

- Searching and managing text annotations, semantic information, and full text documents in one search engine
- Queries over annotation graphs
- Regular expressions, Kleene operators
- Designed to be integrated as a web service in custom end-user systems with bespoke interfaces
- **Try it at** <http://demos.gate.ac.uk/mimir/>

Search for string **Harriet Harman**

Harriet Harman

Search

Documents 1 to 20 of 81:

[UK childcare needs to be more affordable - CentreForum](#) (cached)

quality' MP **Harriet Harman** was the architect

[Birth weight among social mobility checks - Nick Clegg](#) (cached)

's deputy leader **Harriet Harman** said Mr Clegg

[Ed Miliband's shadow cabinet and ministerial teams](#) (cached)

Miliband Opposition leader **Harriet Harman** Deputy Leader & in 2011. **HARRIET HARMAN** - DEPUTY LEADER

[PM's response to Skinner Commons question 'shameful'](#) (cached)

. Deputy leader **Harriet Harman** wrote on Twitter

[Daily Politics and Sunday Politics highlights of 2012](#) (cached)

Sunday April 29 **Harriet Harman** on Hunt, Cameron and Clegg **Harriet Harman** struggles with bank on health by **Harriet Harman**
PMQs: Harriet ...

[Leveson Inquiry: Jeremy Hunt 'sought News Corp guidance'](#) (cached)

Shadow Culture Secretary **Harriet Harman** says Jeremy Hunt culture secretary, **Harriet Harman**, told the

[No Rupert Murdoch deal, says Alastair Campbell](#) (cached)

Search for string **Harriet Harman says**

Harriet Harman says

Search

Documents 1 to 20 of 29:

Leveson Inquiry: Jeremy Hunt 'sought News Corp guidance' (cached)

Shadow Culture Secretary **Harriet Harman says** Jeremy Hunt was

Ed Miliband defends Iraq war condemnation (cached)

, says Harman **Harriet Harman says** Labour will be

Ed Miliband tells Labour: We're the optimists now (cached)

, says Harman **Harriet Harman says** Labour will be

Labour must have credible deficit plan, says Darling (cached)

, says Harman **Harriet Harman says** Labour will be

David Miliband says he won't join brother Ed's team (cached)

, says Harman **Harriet Harman says** Labour will be

Balls: Labour must fight cuts 'every inch of the way' (cached)

, says Harman **Harriet Harman says** Labour will be

Search with morphological variants:

Harriet Harman root:say

Harriet Harman root:say

Search

Documents 21 to 18 of 38:

[David Cameron criticised for 'calm down dear' jibe](#) (cached)

former equality minister **Harriet Harman** said Mr Cameron's

[Queen's Speech: Biggest change to voter registration](#) (cached)

, Labour's **Harriet Harman** said the government was

[Harriet Harman struggles with bank bonus and job figures](#) (cached)

in Coventry, **Harriet Harman** said: "I

[PMQs: Harriet Harman and Nick Clegg on unemployment](#) (cached)

Labour, but **Harriet Harman** said unemployment was falling

[Leveson Inquiry: Jeremy Hunt fair on BSkyB, says top civil servant](#) (cached)

Shadow culture secretary **Harriet Harman** said: "David

[Jeremy Hunt: I followed due process over BSkyB](#) (cached)

But Labour's **Harriet Harman** said Mr Hunt had

[Ed Miliband 'will marry' but politics 'got in the way'](#) (cached)

Replace strings with NEs: {Person} root:say

{Person} root:say

Search

Documents 1 to 20 of 3980:

Apple's Sir Jonathan Ive reaffirms desire to stay at company (cached)

Today programme, **Sir Jonathan said** he would stay partner". **Sir Jonathan said** that Apple's

Diamond Jubilee Tube train was faulty (cached)

be happening' **Ms Siggs said**: "It

Warning over deep-ocean stowaways (cached)

embarrassment. But **Dr Voight says** the experience is it," **Dr Voight said**. "We

School building system not fit for purpose, review says (cached)

shadow education secretary **Andy Burnham said** Mr Gove had . General secretary **Chris Keates said** the capital budget BCSE) director **Ty Goddard said** there was "

EU wants Greece to stay in eurozone, says Van Rompuy (cached)

's Europe editor **Gavin Hewitt says** the crisis gives UK Prime Minister **David Cameron said** "there was . German Chancellor **Angela Merkel said** the bonds,

Huhne partner loses privacy case (cached)

000 costs. **Ms Trimingham said** this could become Daily Mail's **Andrew Pierce said** it was a Speaking outside court **Ms Trimingham said** she was disappointed

{Person} AND root:say – 11803 hits

Stone Roses reunion gig hailed by fans (cached)

by fans By Ian Youngs Entertainment reporter, BBC News The gi ... re. "They've never played so well together," said 43-year-old together," said 43-year-old Andrew Rudder, from Ashton 43-year-old Andrew Rudder, from Ashton under Lyne. But opinion ... s voice. "He can't sing but he never could," said Tom Six, ...

000011_ http://www.bbc.co.uk/news/northern_ireland/ (cached)

02:07 Michaela McAreavey trial - the first day Watch 02:51 Minister says she has ' :51 Minister says she has 'heard enough' Watch 01:12 Flat bombs find - man arrested Watch 01:39 Michaela's husband braves

Warning over deep-ocean stowaways (cached)

using the famous Alvin sub say the vehicle picked famous Alvin sub say the vehicle picked up limpets from a depth of ... s had to cope with huge pressure changes as Alvin conducted its dives pressure changes as Alvin conducted its dives. The researchers report ... matter of some embarrassment. But Dr Voight says the experience is ...

In pictures: Royal arts gathering (cached)

Queen. Sir Paul McCartney, who was among the musicians the event, said he was " the event, said he was "a big fan" of the monarch. Artist David Hockney shared a few

School building system not fit for purpose, review says (cached)

purpose, review says Accessibility links Skip to content Skip to I ... ucation & Family Home World UK England N. Ireland Scotland Wales Business World UK England N. Ireland Scotland Wales Business Politics Health ... building system not fit for purpose, review says Some schools awaiting purpose, review says Some schools awaiting rebuilds rely on tempor ... government-commissioned review by Sebastian James of Dixons Group ...

Hewlett-Packard to cut 27,000 jobs by the end of 2014 (cached)

World UK England N. Ireland Scotland Wales Business Politics Health ... cut 27,000 jobs by end of 2014. The company said the cuts - . The company said the cuts - about 8% of its workforce - will r ... riod fell 3% on a year ago to \$30.7bn. Meg Whitman, HP's year. HP said in a statement that the money would be reinve ... riod fell 3% on a year ago to \$30.7bn. Meg Whitman, HP's ...

{Person} [0..5] root: say – 5495 hits

Documents 1 to 20 of 5495:

Apple's Sir Jonathan Ive reaffirms desire to stay at company (cached)

Today programme, **Sir Jonathan said** he would stay partner". **Sir Jonathan said** that Apple's

Diamond Jubilee Tube train was faulty (cached)

be happening' **Ms Siggs said**: "It

Warning over deep-ocean stowaways (cached)

using the famous **Alvin sub say** the vehicle picked embarrassment. But **Dr Voight says** the experience is it," **Dr Voight said**. "We

School building system not fit for purpose, review says (cached)

government-commissioned review by **Sebastian James of Dixons Group said** value for money by Education Secretary **Michael Gove**, **Mr James said**. Schools with shadow education secretary **Andy Burnham said** Mr Gove had ...

EU wants Greece to stay in eurozone, says Van Rompuy (cached)

, European Council **President Herman Van Rompuy has said**. He was European Council President **Herman Van Rompuy has said**. He was 's Europe editor **Gavin Hewitt says** the crisis gives ...

Huhne partner loses privacy case (cached)

000 costs. **Ms Trimingham said** this could become Daily Mail's **Andrew Pierce said** it was a Speaking outside court **Ms Trimingham said** she was disappointed ...

Migration to UK more than double government target (cached)

. Immigration Minister **Damian Green said**: "Our figures. Chairman **Sir Andrew Green said**: "You

No 'inappropriate' government contact, News Corp lobbyist tells Leveson (cached)

adviser. Fred **Michel said** he did not 's team. **Adam Smith stood down after saying** his e-mails with inquiry, Mr **Michel says** he did not ...

Hands-On with patent data

<http://demos.gate.ac.uk/mimir/patents/search/index>

Text. Matches plain text.

Example: **nanomaterial**

Linguistic variations of text

Example: **(root:nanomaterial | root:nanoparticle)**

Annotation. Matches semantic annotations.

Syntax: {Type feature1=value1 feature2=value2...}

Example: **{Abstract lang="DE"}**

Sequence Query. Sequence of other queries.

Syntax: Query1 [n..m] Query2...

Example: **from {Measurement} [1..5] {Measurement}**

Inclusion Queries

IN Query. Hits of one query only if inside another.

Syntax: Query1 IN Query2

Example: (root:nanomaterial | root:nanoparticle) IN {Abstract}

Number of times these words are mentioned in patent abstracts (as well as links to the actual documents)

OVER Query. Hits of a query, only if overlapping hits of another.

Syntax: Query1 OVER Query2

Example: {Abstract} OVER (root:nanomaterial | root:nanoparticle)

Finds all abstracts that contain nanomaterial(s) or nanoparticle(s)

Date restrictions

```
(  
  {Abstract lang="EN"} OVER  
    (root:nanomaterial | root:nanoparticle )  
)  
IN  
{PatentDocument date > 20050000}
```

YYYYMMDD

Find references to literature or patents in the prior art or background sections, which contain nanomaterial/nanoparticle

{Reference type="Literature"}

|

{Reference type="Patent"}

) IN

{Section type="PriorArt"}

|

{Section type="BackgroundArt"}

)

OVER

(root:nanomaterial | root:nanoparticle)

Queries Using External Knowledge

{Measurement spec="1 to 100 volts"}

- Uses GNU Units (<http://www.gnu.org/software/units/>) to convert measurements and normalise them to SI units

{Measurement spec="1 to 100 kg m² / A s³"}

- **Example hits:** 10 volts, 2V, +20 and -20 volts; ±10V; +/- 100V; +3.3 volts

{Measurement spec="1 to 100 m / s"}

- **Example hits:** 40 km/hr, 60m/min, 100cm/sec, 60 fps; 10 to 2000 cm/sec

Searching LOD with SPARQL

- SQL-like query language for RDF data
- Simple protocol for querying remote databases over HTTP
- Query types
 - *select*: projections of variables and expressions
 - *construct*: create triples (or graphs) based on query results
 - *ask*: whether a query returns results (result is true/false)
 - *describe*: describes resources in the graph

SPARQL Example

Software companies founded in the US

PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX dbpedia: <http://dbpedia.org/resource/>

PREFIX dbp-ont: <http://dbpedia.org/ontology/>

PREFIX geo-ont: <http://www.geonames.org/ontology#>

PREFIX umbel-sc: <http://umbel.org/umbel/sc/>

SELECT DISTINCT ?Company ?Location

WHERE {

 ?Company rdf:type dbp-ont:Company ;

 dbp-ont:industry dbpedia:Computer_software ;

 dbp-ont:foundationPlace ?Location .

 ?Location geo-ont:parentFeature dbpedia:United_States .

}

SPARQL Results



SPARQL Query

Results for your query (81) - [Edit query](#)

View as [Exhibit](#)

| Company | Location |
|---------------------------------|------------------------------|
| dbpedia:Redxpress | dbpedia:Glen_Allen,_Missouri |
| dbpedia:Borland | dbpedia:California |
| dbpedia:Lawson_Software | dbpedia:Minneapolis |
| dbpedia:Exterro,_Inc. | dbpedia:Oregon |
| dbpedia:Tableau_Software | dbpedia:Seattle |
| dbpedia:NeuroDimension | dbpedia:Florida |
| dbpedia:Computer_Usage_Company | dbpedia:New_York_City |
| dbpedia:Macromedia | dbpedia:San_Francisco |
| dbpedia:Core_International,_Inc | dbpedia:Florida |
| dbpedia:Cerulean_Studios | dbpedia:Connecticut |
| dbpedia:Cornerstone_OnDemand | dbpedia:California |

Try at: <http://factforge.net/sparql>

Try these on the BBC News demo:

<http://services.gate.ac.uk/mimir/gpd/search/index>

Gordon Brown [0..3] root:say

{Person} [0..3] root:say

{Person inst ="http://dbpedia.org/resource/Gordon_Brown"} [0..3] root:say

{Person sparql="SELECT ?inst WHERE { ?inst a :Politician }"} [0..3] root:say

{Person sparql = "SELECT ?inst WHERE { ?inst :party
<http://dbpedia.org/resource/Labour_Party_%28UK%29> }" } [0..3] root:say

{Person sparql = "SELECT ?inst WHERE {
 ?inst :party <http://dbpedia.org/resource/Labour_Party_%28UK%29> .
 ?inst :almaMater <http://dbpedia.org/resource/University_of_Edinburgh> }"
} [0..3] root:say

User Interfaces for SPARQL-based Semantic Search

- SPARQL-based semantic searches, tapping into LOD resources are extremely powerful
- However, impossible to write by the vast majority of users
- User interfaces for SPARQL-based semantic search:
 - Faceted searches (see ExoPatent next)
 - Form-based searches (see EnviLOD, PIN)
 - Text-based searches (natural language interfaces for querying ontologies), e.g. FREyA

Faceted Search: ExoPatent Example

- Use semantic information to expose linkages between documents based on the intersecting relationships between various sets of data from
 - FDA Orange Book (23,000 patented drugs)
 - Unified Medical Language System (UMLS) – database of medical terms (370,000)
 - Patent bibliographic information
- Search for diseases, drug names, body parts, references to literature and other patents, numeric values, ranges
- Demo uses a small set of patents (40,000)

ExoPatent: Faceted Search

ExoPatent | [PATTERNS](#) / [FACETS](#) / [BOOLEAN](#) / [MIMIC SEARCH](#)

Facets >

Selected Items

☐ AMOXICILLIN

☐ GENTAMICIN

Recent Items

☐ Human herpesvirus 1

☐ MERCK & CO INC

Terms from FDA Orange Book

FDA Drug Name

25 of 1456 shown below.

ALBUTEROL SULFATE
CARBAMAZEPINE
CLOTRIMAZOLE
ETHOSUXIMIDE
GENTAMICIN
GUANFACINE HYDROCHLORIDE
INDOCIN
MERCAPTOPURINE
MISOPROSTOL
NALBUPHINE
NEOSAR
NIZATIDINE
OCTREOTIDE ACETATE
PERPHENAZINE
PODOFILOX
POTASSIUM CITRATE

Active Ingredients

25 of 1460 shown below.

ALBUTEROL SULFATE
AMOXICILLIN
AMPHETAMINE SULFATE
CEFTAZIDIME
CETRORELIX
CLAVULANATE POTASSIUM
DEXTROAMPHETAMINE SULFAT...
DOXAZOSIN MESYLATE
ETHOSUXIMIDE
GENTAMICIN SULFATE
HEPARIN SODIUM
HYDRALAZINE HYDROCHLORID...
HYDROCHLOROTHIAZIDE
METHOTREXATE SODIUM
METRIZOIC ACID
NALBUPHINE HYDROCHLORIDE

Applicant

25 of 273 shown below.

ABBOTT GMBH & CO KG
ALLERGAN INC
ALZA CORP
ALZA CORPORATION
BOARD OF REGENTS, THE UN...
BOEHRINGER INGELHEIM PHA...
FUISZ TECHNOLOGIES LTD
FUISZ TECHNOLOGIES LTD.
HOFFMANN LA ROCHE
HOFFMANN-LA ROCHE INC.
MERCK & CO INC
MERCK & CO., INC.
NOVEN PHARMA
NOVEN PHARMACEUTICALS, I...
PROCTER & GAMBLE
SCHERING CORP

UMLS Concept

25 of 4130 shown below.

Ankylosing spondyl...
Autoimmune hemol...
Brucella melitensis
Bullous pemphigoi...
Cytomegalovirus
Enterobacter aerog...
Enterovirus
Genus: Coronaviru...
Heart failure
Human herpesvirus...
Human Herpesviru...
Myotonic Dystrophy...
Postpericardiotomy...
Respiratory syncyti...
Rhinovirus
Scleroderma

Document Keyword Filter

Matching documents: **362**

Patent Documents Containing FDA-related Terms

1-10 of 362 documents matching the search criteria.

| Publication Date | Patent Number | Assignee(s) | Title |
|------------------|-----------------------------------|-----------------------------------|---|
| 10-11-2005 | US-20050250705-A1 | BOEHRINGER INGELHEIM PHARMA GM... | Spray-dried powder comprising at least one 1,... ... pefloxacin, amifloxacin, fleroxacin, tosufloxacin, pruliflo... irioxacin, pazufloxacin, clinafloxacin and sitafloxacin; amin... such as, for example, gentamicin , netilmicin, paramecin, t... |

Mitochondria is an Examples of Metathesaurus concepts: Golgi Apparatus; Microsomes; Organelles , Trusted^{tip!}

"Semiautonomous, self-reproducing organelles that occur in the cytoplasm of all cells of most, but not all, eukaryotes. Each mitochondrion is surrounded by a double limiting membrane. The inner membrane is highly invaginated, and its projections are called cristae. Mitochondria are the sites of the reactions of oxidative phosphorylation, which result in the formation of ATP. They contain distinctive RIBOSOMES, transfer RNAs (RNA TRANSFER); AMINO ACYL T RNA SYNTHETASES; and several other enzymes. Mitochondria depend upon genes that are located in the nucleus of the cells in which they reside for many essential functions, including the synthesis of messenger RNAs (RNA, MESSENGER). Mitochondria are thought to have arisen from aerobic bacteria that established a symbiotic relationship with primitive protoeukaryotes. (King & Stanfield, Dictionary of Genetics, 4th ed)"

comment "Semiautonomous, self-reproducing organelles that occur in the cytoplasm of all cells of most, but not all, eukaryotes. Each mitochondrion is surrounded by a double limiting membrane. The inner membrane is highly invaginated, and the projections are called cristae. Mitochondria are the sites of reactions of oxidative phosphorylation, which result in the formation of ATP. They contain distinctive RIBOSOMES, transfer RNAs (RNA, TRANSFER); AMINO ACYL T RNA synthetases, and elongation and termination factors. Mitochondria posses their own DNA upon genes within the nucleus of the cell. They replicate independently for many essential messenger RNAs (RNA, MESSENGER). Mitochondria are believed to have arisen from aerobic bacteria that established a symbiotic relationship with primitive protoeukaryotes. (King et al., A Dictionary of Genetics, 4th ed)"

has Broader Organelle / dy as a whole

Allowed qualifier Procession, genetic aspects, physiological aspects

| | |
|---|--------------------------------|
| Has parent Organelle , fractionation technique | Similar structure, isolated by |
|---|--------------------------------|

Has narrower [Mitochondria](#), [Heart](#), [Mitochondria](#), [Liver](#), [Mitochondria](#), [Muscle](#), [Submitochondrial Particles](#), [Inner mitochondrial membrane](#) ... (15 more)

Has sibling [Nucleus](#), [Cell-Free System](#), [Ergastoplasm](#), [Golgi apparatus](#), [Intracellular Membranes](#) ... (7 more)

Mitochondria, Heart is an Examples of Metathesaurus concepts: Golgi Apparatus; Microsomes; Organelles , Trusted^{tip!}

"The mitochondria of the myocardium."

comment "The mitochondria of the myocardium."

has Broader Mitochondrion

Allowed qualifier Process of secretion, genetic aspects, physiological aspects

Has parent Mitochondria, Muscle

Related Entities

Mitochondrion Has narrower Mitochondria,
Heart

Mitochondria, Muscle Has child Mitochondria, Heart

Process of secretion Qualified by Mitochondria, Heart

genetic aspects Qualified by Mitochondria,
Heart

physiological aspects Qualified by
Mitochondria, Heart

Copyright © 2006-2010 Ontotext AD

Find all applicants who filed patents related to mitochondria, as well as drug names and active ingredients

Selected Items

Mitochondria

Recent Items

(No recent items)

Terms from FDA Orange Book

FDA Drug Name

25 of 1451 shown below.

ALBUTEROL SULFATE
CARBAMAZEPINE
CLOTRIMAZOLE
ETHOSUXIMIDE
GENOTROPIN
GENTAMICIN
HYZAAR
MERCAPTOPYRINE
MISOPROSTOL
NALBUPHINE
NEOSAR
NIZATIDINE
OCTREOTIDE ACETATE
PERPHENAZINE
PODOFILOX
POTASSIUM CITRATE

Active Ingredients

25 of 1141 shown below.

ALBUTEROL SULFATE
AMOXICILLIN
AMPHETAMINE SULFATE
CEFTAZIDIME
CETRORELIX
CINOXACIN
CLAVULANATE POTASSIUM
DEXTROAMPHETAMINE SULFAT...
DOXAZOSIN MESYLATE
ETHOSUXIMIDE
ETOPOSIDE
GENTAMICIN SULFATE
HEPARIN SODIUM
HYDROCHLOROTHIAZIDE
METHOTREXATE SODIUM
NALBUPHINE HYDROCHLORIDE

Applicant

25 of 828 shown below.

AJINOMOTO CO., INC.
ALLERGAN INC
BOARD OF REGENTS, THE UN...
BOEHRINGER MANNHEIM GMBH
BRITISH TECH GROUP
CHIRON CORP
CHIRON CORPORATION
FUJIREBIO KABUSHIKI KAIS...
FUJIREBIO KK
HARVARD COLLEGE
HESKA CORPORATION
MANDEL ARKADY
MEDICAL DISCOVERIES INC
MERCK & CO INC
MERCK & CO., INC.
MINEMURA TSUYOSHI

UMLS Concept

25 of 7998 shown below.

Acetylation
Adenoviruses
Aedes
Alphavirus
Animals
Anopheles Genus
Ants
Anus
Bacteria
Bacteriophage la
Bacteriophages
Blood
Blood capillaries
Cell Wall
Cells
Chagas Disease

We don't just have to look at politicians saying and measuring things

- If we first process the text with other NLP tools such as sentiment analysis, we can also search for positive or negative documents
- Or positive/negative comments about certain people/topics/things/companies
- In the Decarbonet project, we looked at people's opinions about topics relating to climate change, e.g. fracking
- We could index on the sentiment annotations too
- Other people are using the combination of opinion mining and MIMIR to look at e.g. customer feedback about products / companies on a huge corpus

Explicitly Choosing The Search Classes

GATE Prospector

Diseases

Pathogens

Pathogenesis

Vaccine

Animals and Models

Custom Mimir Query

URI:

Disease

Bacterial_disease

Viral_disease

Acquired_immunity

Acute_hepatitis

Argentinian_ha

Document Metadata

Dates: to

source:

Documents

Terms

PubMed By PMIDAbstract1000.txt

pertussis and Haemophilus influenzae.

PubMed By PMIDAbstract10033.txt

syncytial virus (RSV). The injection of purified RSV in Freund's inoculation of purified RSV with Bordetella pertussis ...

PubMed By PMIDAbstract10058.txt

cell infiltrates. Hepatitis and splenitis with

PubMed By PMIDAbstract10085.txt

coli, Haemophilus influenzae and Proteus mirabilis

PubMed By PMIDAbstract1012.txt

and one-half received hepatitis A vaccine (control

PubMed By PMIDAbstract1013.txt

features of fatal influenza virus infection in national surveillance for influenza-associated deaths

1-20 of 10,256

Environmental signals implicated in Dr fimbriae release by pathogenic Escherichia coli. Afa/Dr diffusely adhering Escherichia coli have been shown to cause urinary tract infections and enteric infections. Virulence of Dr-positive IH11128 bacteria is associated with the presence of Dr fimbriae. In this report, we show for the first time that the Dr fimbriae are released in the extracellular medium in response to multiple environmental signals. Production and secretion of Dr fimbriae are clearly thermoregulated. A comparison of the amounts of secreted fimbriae showed that the secretion is drastically increased during anaerobic growth in minimal medium. The effect of anaerobiosis on secretion seemed to depend on both the growth phase and the culture medium. The secretion was maximal during the logarithmic-phase growth and corresponded to 27 and 57% of total Dr fimbriae produced by bacteria grown in mineral medium+glucose and LB broth, respectively. Thus, the anaerobic environment of the colon would favour the secretion of Dr fimbriae during bacterial multiplication. The controlled release of the Dr fimbriae, which is carried out in the absence of cellular lysis, appears independent of the action of proteases or a process of maturation. The mechanism employed in the liberation of Dr fimbriae thus seems different from that described for the adhesins EHA and Hap of Bordetella pertussis.

Choosing A Specific Instance

GATE Prospector

Search ▼

DiseasesPathogensPathogenesisVaccineAnimals and ModelsCustom Mimir Query

URI:

Disease ▶

Bacterial_disease

Viral_disease ▶

Burkitts_lymphoma

Cervical_cancer

Chandipura_encephalitis

Document Metadata

Dates: to

source:

Search

DocumentsTerms

[PubMed_By_PMIDAbstract1306.txt](#)
the development of **cervical cancer**. The HPV

[PubMed_By_PMIDAbstract3999.txt](#)
volume regulation of **cervical cancer** cells. On
of RVD in **cervical cancer** cells, while

[PubMed_By_PMIDAbstract664.txt](#)
, and **cervical cancer**. Except for

[PubMed_By_PMIDAbstract10702.txt](#)
all women like **cervical cancer**, which might

[PubMed_By_PMIDAbstract11522.txt](#)
menstruating women with **cervical cancer** in situ
showing

[PubMed_By_PMIDAbstract11796.txt](#)
management of invasive **cervical cancer** becomes
even more

Eradication of established tumors by vaccination with recombinant Bordetella pertussis adenylate cyclase carrying the human papillomavirus 16 E7 oncoprotein. High-risk human papillomaviruses (HPV) such as HPV16 are associated with the development of **cervical cancer**. The HPV16-E6 and HPV16-E7 oncoproteins are expressed throughout the replicative cycle of the virus and are necessary for the onset and maintenance of malignant transformation. Both these tumor-specific antigens are considered as potential targets for specific CTL-mediated immunotherapy. The adenylate cyclase (CyaA) of Bordetella pertussis is able to target dendritic cells through specific interaction with the alpha(M)beta(2) integrin. It has been previously shown that this bacterial protein could be used to deliver CD4(+) and CD8(+) T cell epitopes to the MHC class II and class I presentation pathways to trigger specific Th and CTL responses in vivo, providing protection against subsequent viral or tumoral challenge. Here, we constructed recombinant CyaA containing either the full sequence or various subfragments from the HPV16-E7 protein. We show that, when injected to C57BL/6 mice in absence of any adjuvant, these HPV16-recombinant CyaAs are able to induce specific Th1 and CTL responses. Furthermore, when injected into mice grafted with HPV16-E7-expressing

1-20 of 77

GATE Prospector

Search

Diseases

Pathogens

Pathogenesis

Vaccine

Animals and Models

Custom Mimir Query

URI:

Disease

Bacterial_disease

Viral_disease

Burkitts_lymphoma

Cervical_cancer

Chandipura_encephalitis

Document Metadata

Dates: to

source:

Search

Documents

Terms

Select

top

40

terms of type {Disease}

from the top <All>

retrieved documents.

| Term | Count |
|------------------------|-------|
| Influenza | 6,683 |
| Tick_borne_encephaliti | 5,501 |
| Japanese_encephalitis | 3,743 |
| Hepatitis_B | 3,201 |
| Dengue_haemorrhagic_ | 1,944 |
| Pertussis | 1,852 |
| Hepatitis | 1,588 |
| Yellow_fever | 1,580 |
| Measles | 1,428 |
| Tetanus | 1,344 |

Save this as a term set named {Disease} (40)

Saved term sets {Disease} (40)

What Pathogens?

GATE Prospector

Search ▼

Diseases

Pathogens

Pathogenesis

Vaccine

Animals and Models

Custom Mimir Query

URI:

Pathogen

Bacteria

Virus

Document Metadata

Dates: to

source:

Search

Documents

Terms

Select

top

40

terms of type {Pathogen}

from the top

<All>

retrieved documents.

| Term | Count |
|--------------------------|--------|
| Dengue_virus | 17,867 |
| West_Nile_virus | 11,726 |
| Bordetella_pertussis | 9,909 |
| Japanese_encephalitis_vi | 5,093 |
| Human_immunodeficien | 4,939 |
| Tick_borne_encephalitis | 3,899 |
| Haemophilus_influenzae | 2,709 |
| Escherichia_coli | 2,342 |
| Hepatitis_B_virus | 2,043 |
| Hepatitis_C_virus | 1,567 |



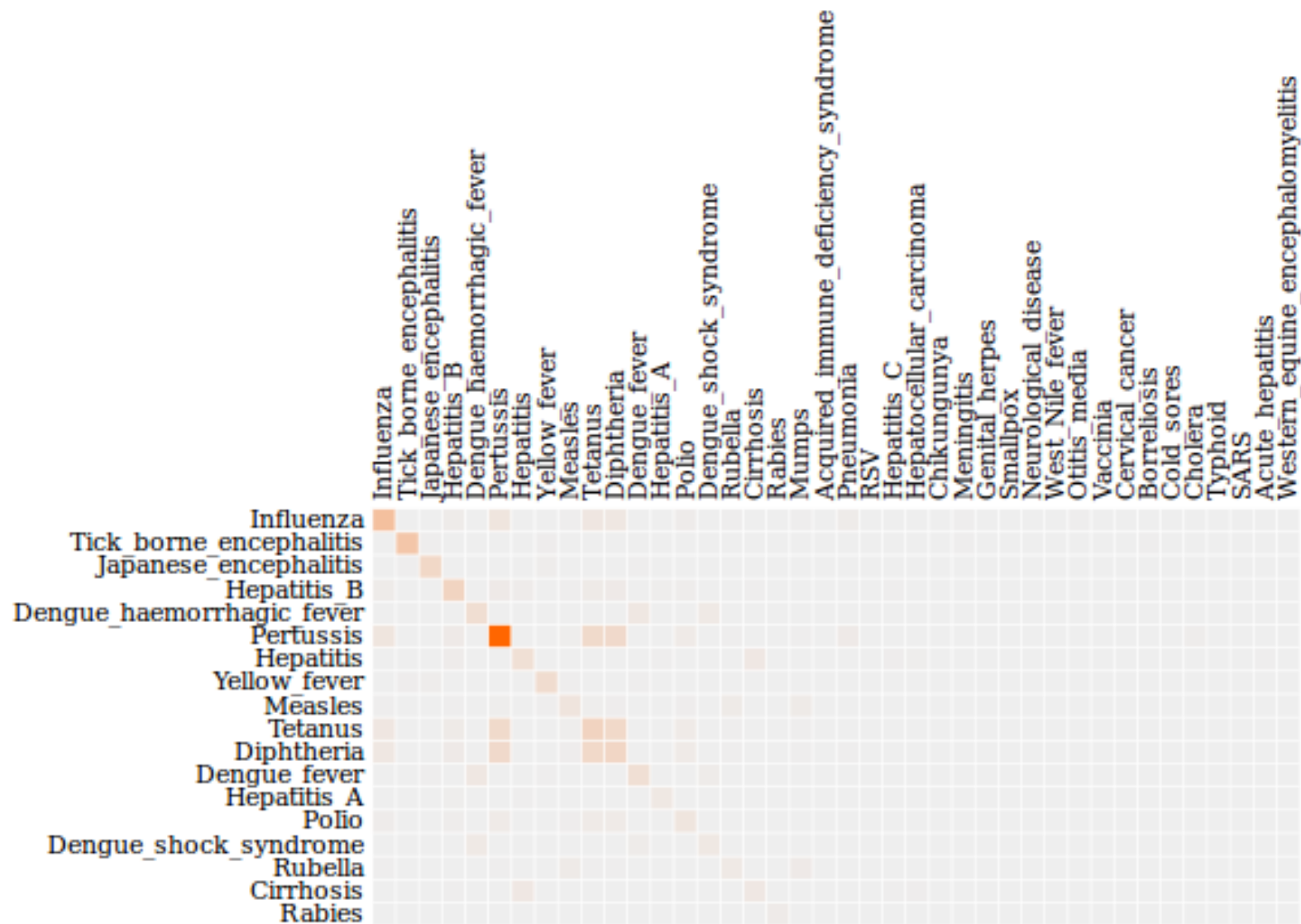
A word cloud visualization of various pathogens. The most prominent terms are 'Dengue_virus', 'West_Nile_virus', 'Bordetella_pertussis', 'Haemophilus_influenzae', 'Hepatitis_B_virus', and 'Hepatitis_C_virus'. Other visible terms include 'Poliovirus', 'Neisseria_meningitidis', 'Rubella_virus', 'Influenza_A_virus', 'Varicella_zoster_virus', 'Adenovirus', 'Mycobacterium_bovis', 'Hepatitis_C_virus', 'cytomegalovirus', 'Legionella_pneumophila', 'Staphylococcus_aureus', 'Escherichia_coli', 'Yellow_fever_virus', 'H1N1', 'Ross_River_virus', 'Streplococcus_pneumoniae', 'Mycobacterium_tuberculosis', 'Hepatitis_A_virus', 'Human_respi...yrial_virus', 'Human_papilloma_virus', 'Staphylococcus_aureus', 'Stigella_basent', 'Helicobacter_pylori', 'Salmonella_typhi', 'Vaccinia_virus', 'H5N1', and 'H3N2'.

Save this as a term set named {Pathogen} (40)

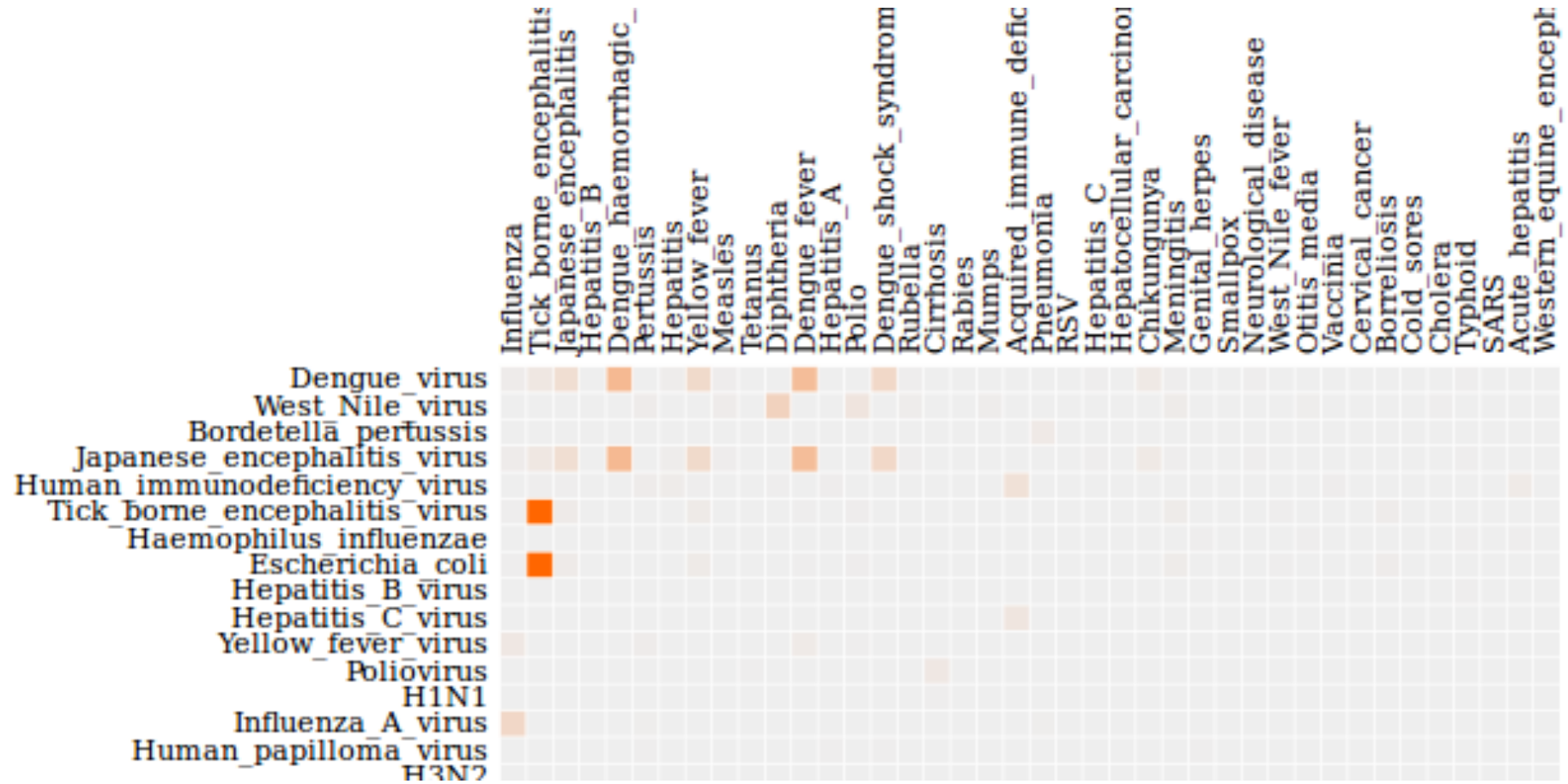
Saved term sets

{Disease} (40) ✖

Disease vs Disease Co-occurrences



Diseases vs Pathogens



Try Prospector!

- Fastvac demo <http://demos.gate.ac.uk/mimir/fastvac-slr1/index>
- <http://gateservice5.dcs.shef.ac.uk:8080/mimir/admin>

Play with more tools on the GATE Cloud!

<http://cloud.gate.ac.uk>

[Home](#) [Services](#)

Try our services

Type some text to annotate

Or select a text file:

Choose File

 No file chosen

Results

Nothing to show

I want to...

☒ find English named entities ([details](#))

☐ analyse English Tweets ([details](#))
















☐ find mentions of political topics ([details](#))

Go

There are lots to choose from!

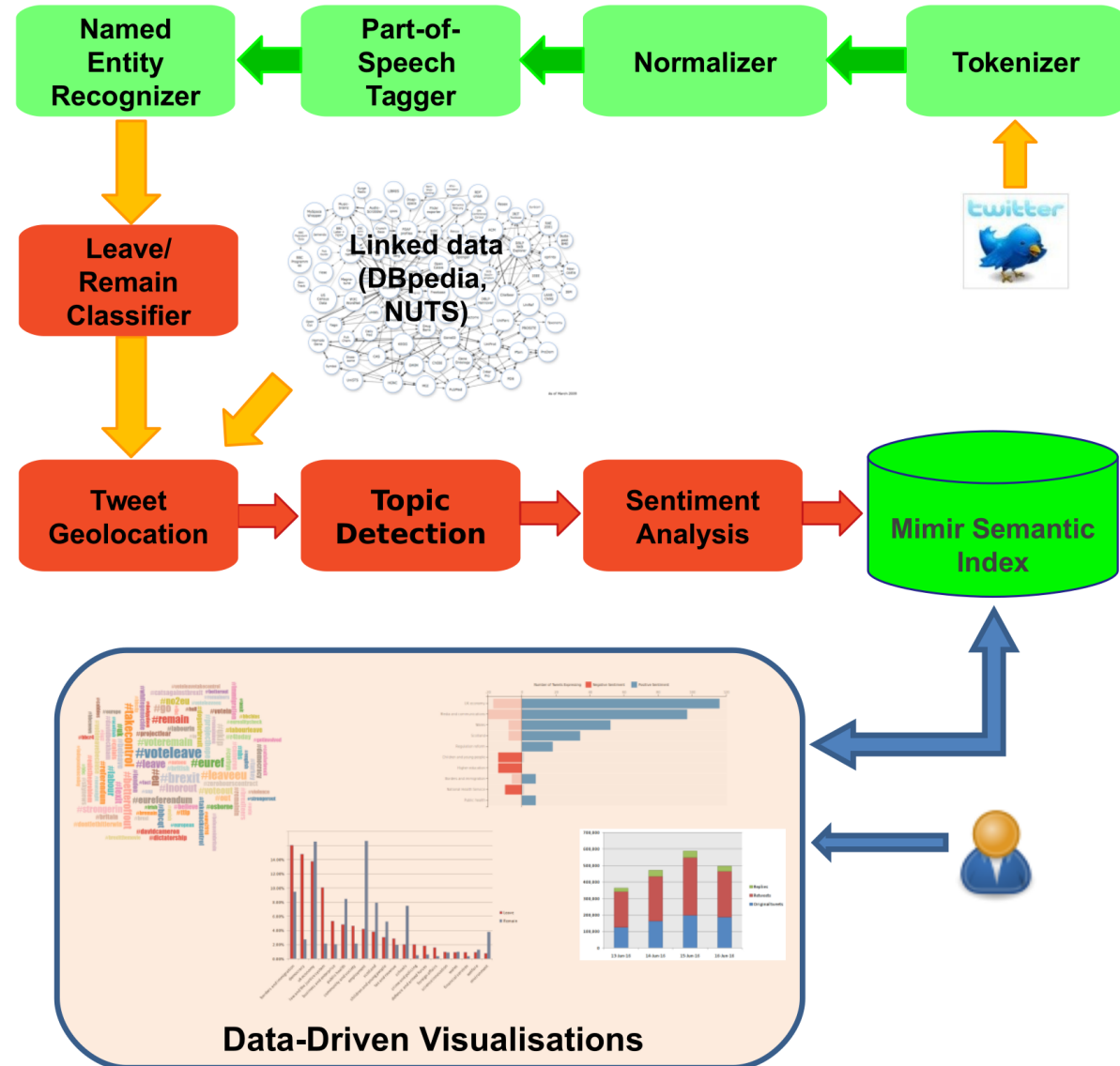
These are just a few of them!

Show only items tagged: Basque (1) Biomed (1) Bulgarian (1) Catalan (1) Chunker (1) Croatian (1) Custom (2) Czech (1) Danish (1) Dutch (2) English (15) Environment (2) Estonian (1) Finnish (1) French (4) German (5) Greek (1) Indonesian (1) Latvian (1) Measurements (2) Morphology (1) Named Entity (23) OpenNLP (3) Opinion Mining (4) Part-of-Speech (22) Polish (1) Politics (1) Portuguese (1) Romanian (2) Russian (3) Server (2) Slovak (1) Slovenian (1) SoBigData (1) Spanish (3) Summarization (2) Swedish (1) Term Recognition (2) Twitter (11) Welsh (1)

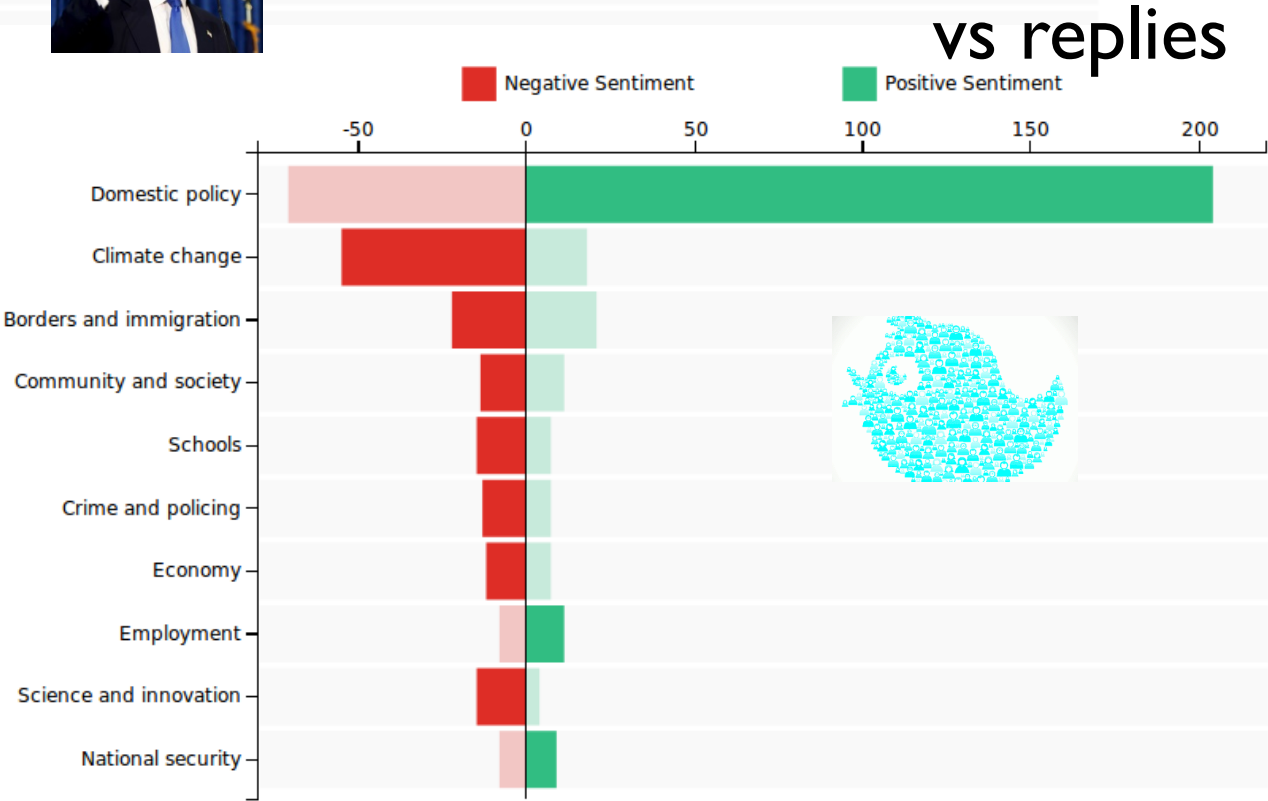
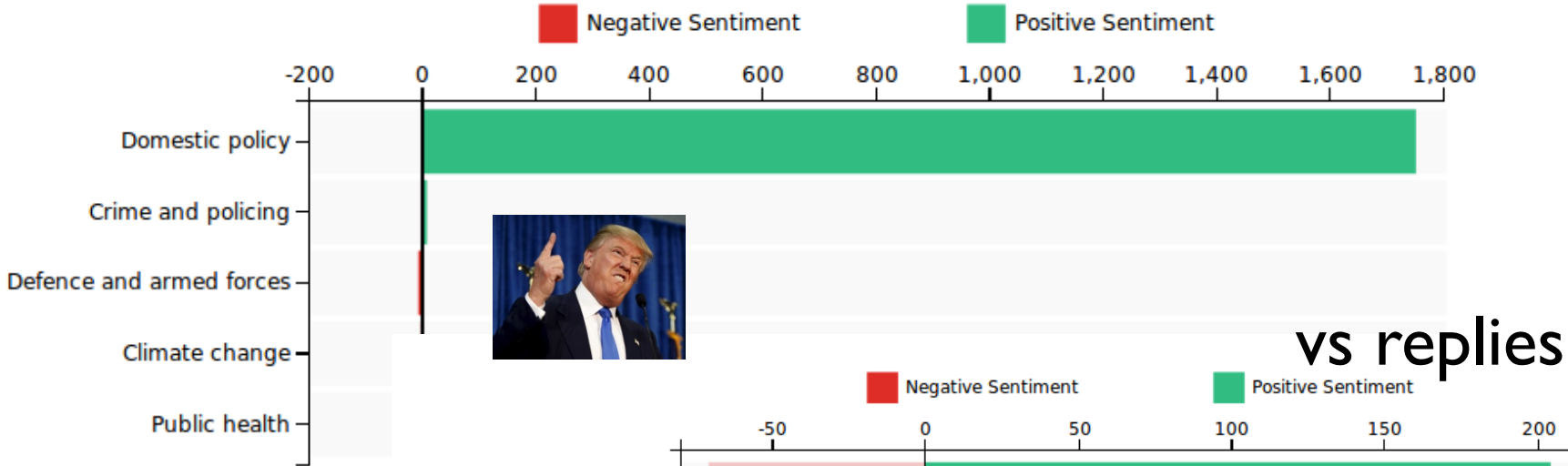
| | | |
|---|---|---|
|  <h3>English Named Entity Recognizer</h3> <p>Identify names of <i>persons, locations, organizations</i>, as well as <i>money amounts, time and date expressions</i> in English texts automatically.</p> <p>1,200 free requests / day Larger batches £0.80 / CPU hour</p> |  <h3>English Named Entity Recognizer for Tweets</h3> <p>Analyse tweets for names of <i>persons, locations, organizations</i> and other entities. Also performs normalization of abbreviations and common shorthands ("brb", "gr8", "2day", etc.).</p> <p>1,200 free requests / day Larger batches £0.80 / CPU hour</p> |  <h3>Twitter Collector</h3> <p>Collect tweets, view tweet statistics, and store results in your dashboard for further analysis.</p> <p>£0.05 / CPU hour</p> |
|  <h3>BioYODIE Named Entity Disambiguation</h3> <p>Identifies biomedical named entities of various types, and disambiguates them against UMLS.</p> <p>800 free requests / day Batch processing not available</p> |  <h3>YODIE Named Entity Disambiguation (English)</h3> <p>Identifies named entities of various types in English documents, and disambiguates them against DBpedia.</p> <p>800 free requests / day Batch processing not available</p> |  <h3>YODIE Named Entity Disambiguation (German)</h3> <p>Identifies named entities of various types in German documents, and disambiguates them against DBpedia.</p> <p>800 free requests / day Batch processing not available</p> |
|  <h3>YODIE Named Entity Disambiguation (French)</h3> <p>Identifies named entities of various types in French documents, and disambiguates them against DBpedia.</p> <p>800 free requests / day Batch processing not available</p> |  <h3>YODIE Named Entity Disambiguation (Spanish)</h3> <p>Identifies named entities of various types in Spanish documents, and disambiguates them against DBpedia.</p> <p>800 free requests / day Batch processing not available</p> |  <h3>German Named Entity Recognizer</h3> <p>A named entity recognition service for documents in German. Based on ANNIE, it identifies names of <i>persons, locations</i>, and <i>organizations</i>.</p> <p>1,200 free requests / day Larger batches £0.80 / CPU hour</p> |
|  <h3>Language Identification for Tweets</h3> <p>Service to identify the languages of tweets.</p> <p>1,200 free requests / day Larger batches £0.80 / CPU hour</p> |  <h3>English Tweet Tokenizer</h3> <p>Identifies words and punctuation in tweets</p> <p>1,200 free requests / day Larger batches £0.80 / CPU hour</p> |  <h3>French Named Entity Recognizer</h3> <p>A named entity recognition service for documents in French. Based on ANNIE, it identifies names of <i>persons, locations</i>, and <i>organizations</i>.</p> <p>1,200 free requests / day Larger batches £0.80 / CPU hour</p> |
|  <h3>Part-of-Speech Tagger for Tweets</h3> <p>A service that tags tweets with part-of-speech information, e.g. nouns, verbs.</p> |  <h3>Romanian Named Entity Recognizer</h3> <p>A named entity recognition service for documents in the Romanian language. Based on ANNIE, it identifies names of <i>persons, locations, organizations</i>, as well as <i>money amounts, time</i></p> |  <h3>Russian Named Entity Recognizer (basic)</h3> <p>A named entity recognition service for documents in the Russian language. Based on ANNIE, it identifies names of <i>persons, locations, organizations</i>, as well as <i>money amounts, time and</i></p> |

Some cool applications with GATE

Try some more search demos...



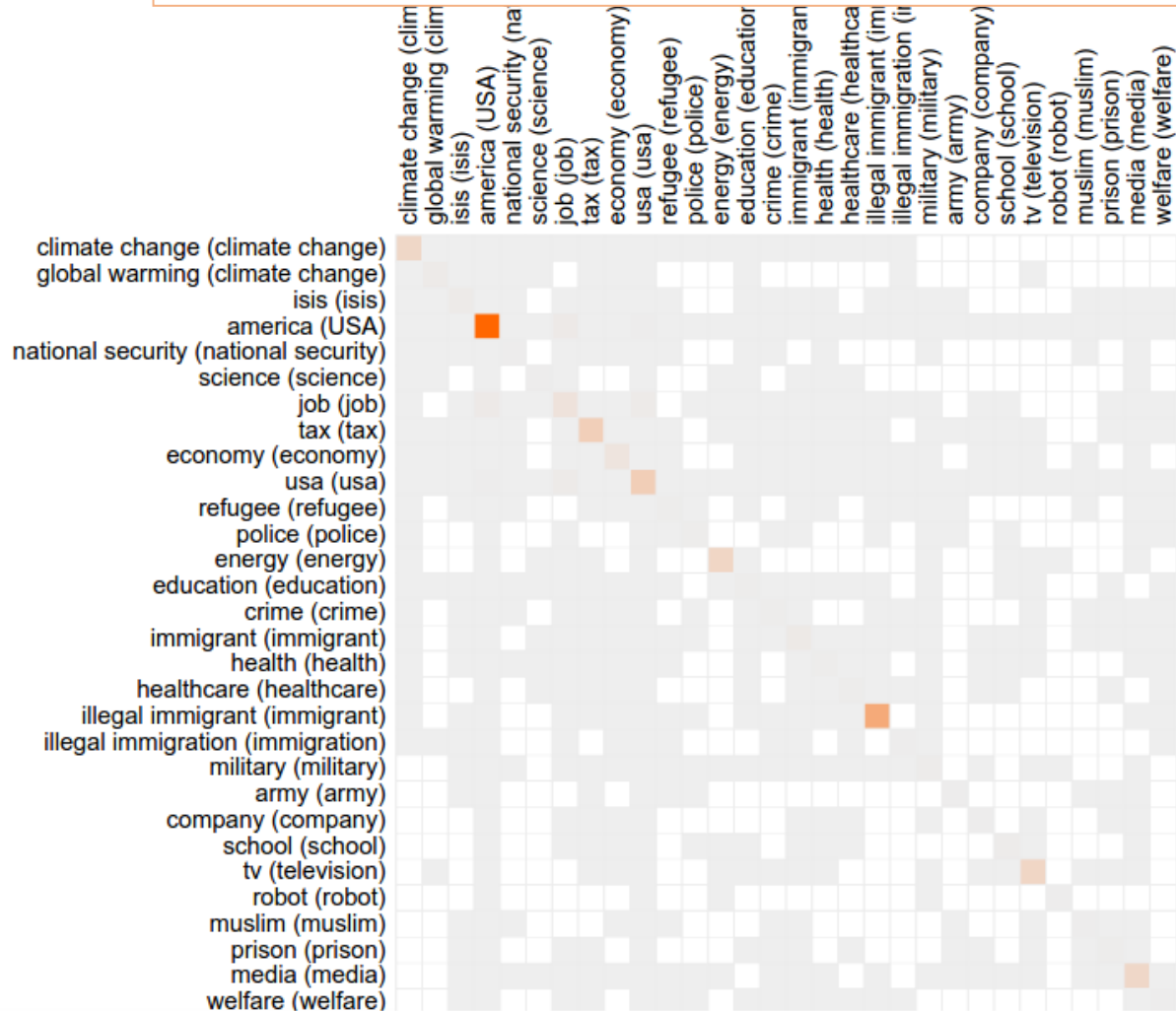
Real-time Opinion Monitoring



Climate change, ISIS and Trump

@realDonaldTrump Someone needs to tell Putin Isis and China to beware, the global warming is coming. That will stop them. Not.

@realDonaldTrump WHY IS EVERYONE IN THIS DEBATE BLAMING GLOBAL WARMING!?!? WHAT DOES THAT HAVE TO DO WITH ISIS?!?!?



Querying election data with MIMIR

- Dataset: every tweet by MP / Candidate / Party, plus all replies/retweets
- Find all tweets where a Conservative MP talked about the economy

Searching Index "2015-03-09"

```
{DocumentAuthor author_party="Conservative Party"} OVER  
{Topic theme="uk_economy"}
```

Search

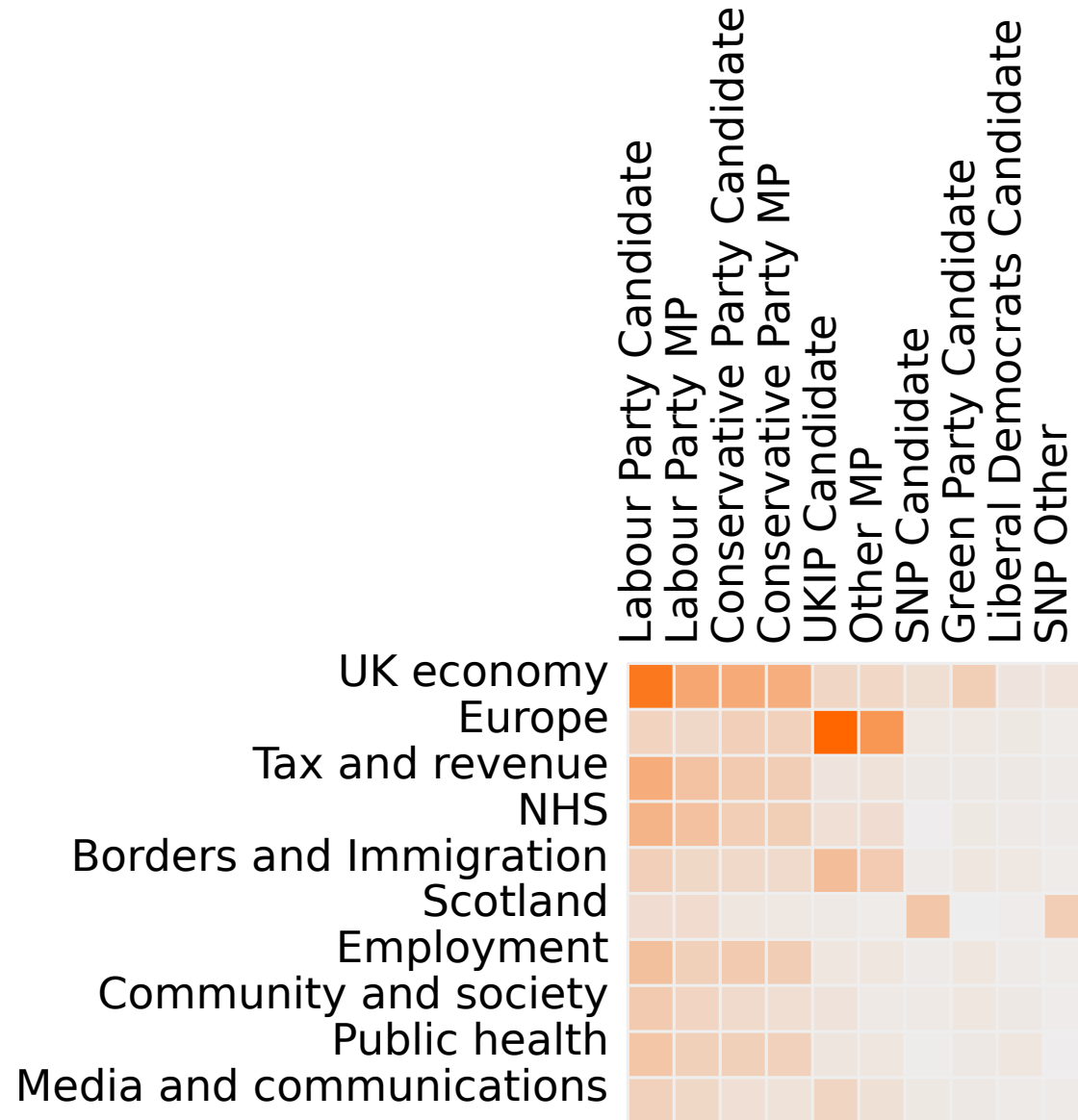


Richard Short
@ToryShorty



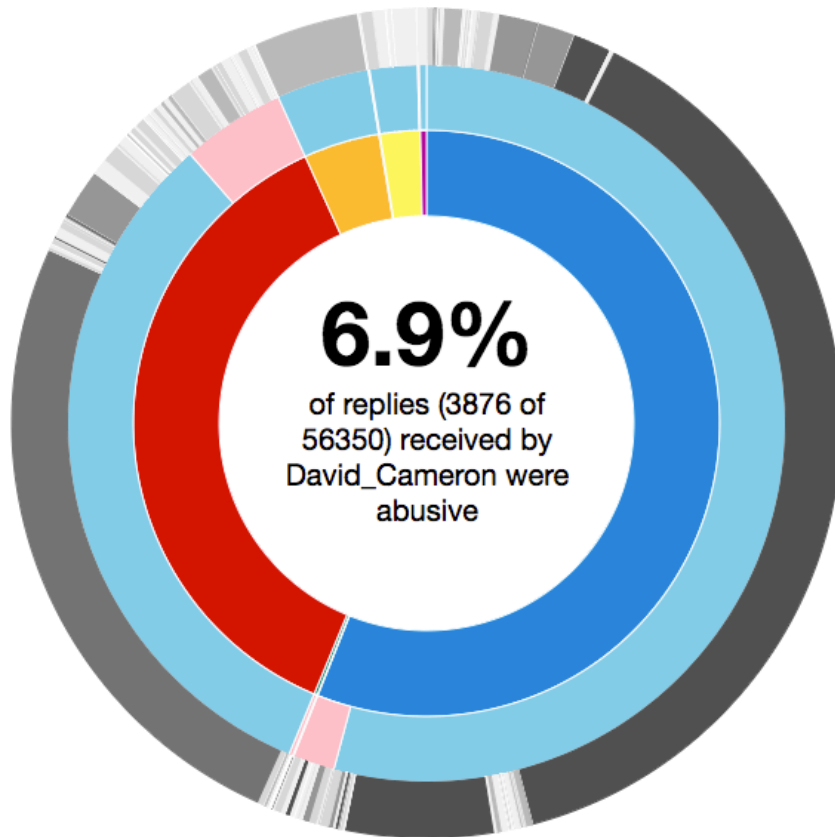
With so many protected from Labour's pension raid are they sure it will even generate £2.7bn [#bbcsp](#)

Parties / themes co-occurrence

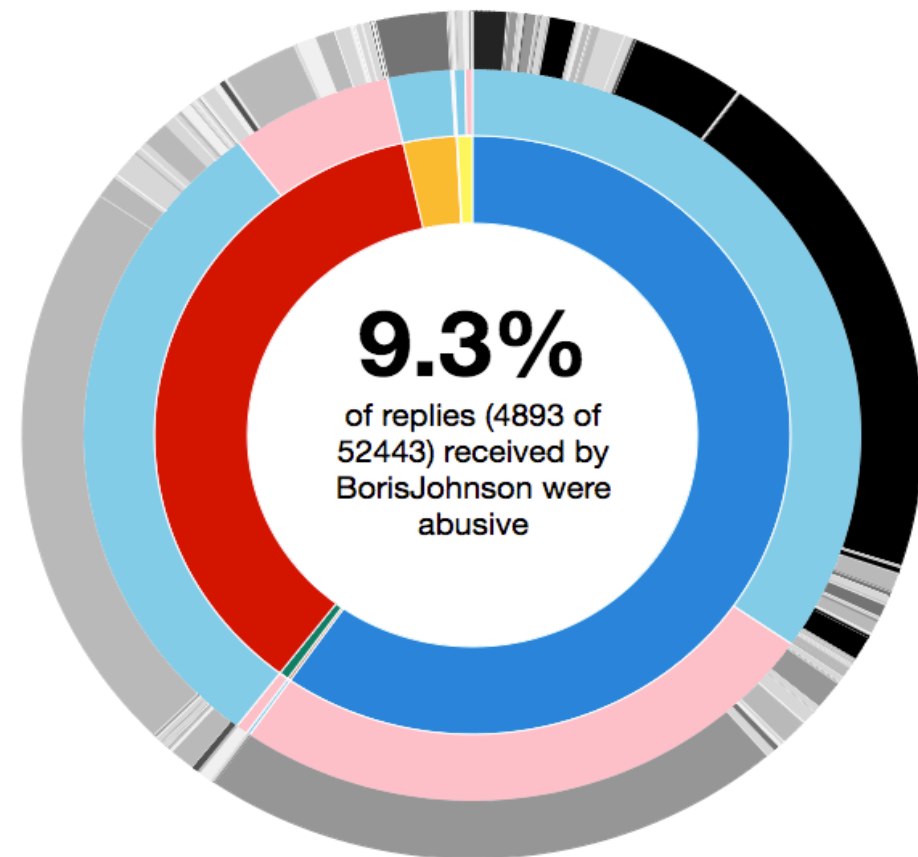


Hate speech towards MPs on Twitter

2015



2017



<http://greenwoodma.servehttp.com/data/buzzfeed/sunburst.html>

Key Publications

- Semantic Search over Documents and Ontologies (2014) K Bontcheva, V Tablan, H Cunningham. Bridging Between Information Retrieval and Databases, 31-53
- D. Maynard, I. Roberts, M. A. Greenwood, D. Rout and K. Bontcheva. A Framework for Real-time Semantic Social Media Analysis. Web Semantics: Science, Services and Agents on the World Wide Web, 2017
- V. Tablan, I. Roberts, H. Cunningham, and K. Bontcheva. GATECloud.net: a Platform for Large-Scale, Open-Source Text Processing on the Cloud. [Philosophical Transactions of the Royal Society A](#), 371(1983), 2013
- More papers on the GATE website: <http://gate.ac.uk/gate/doc/papers.html>

Some useful links

- GATE: <http://gate.ac.uk>
- GateCloud: <https://cloud.gate.ac.uk>
- Annual GATE training course in June in Sheffield:
<https://gate.ac.uk/family/training.html>
- Download GATE: <http://gate.ac.uk/download>
- GATE blog posts on social media analysis: <http://gate4ugc.blogspot.co.uk/>
- UK elections monitor <http://gate.ac.uk/projects/pft>
- [Blog post on abuse of MPs:](#)
- COMRADES project on disasters: <http://gate.ac.uk/projects/comrades>
- KNOWMAK project and demos: <http://gate.ac.uk/projects/knowmak>
- SoBigData project: <http://sobigdata.eu>