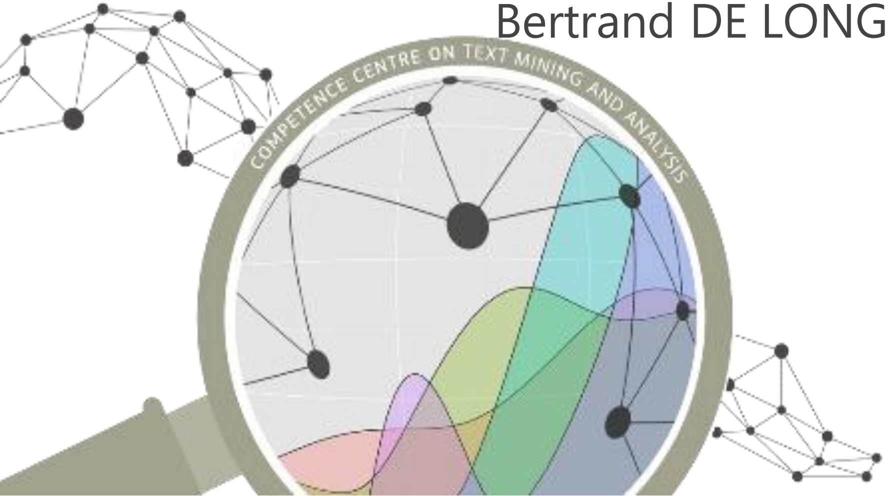
European Commission

JRC’s Text Mining and Analysis Competence Centre

Natural Language Processing for EU policy making

Bertrand DE LONGUEVILLE, DG JRC.I.3



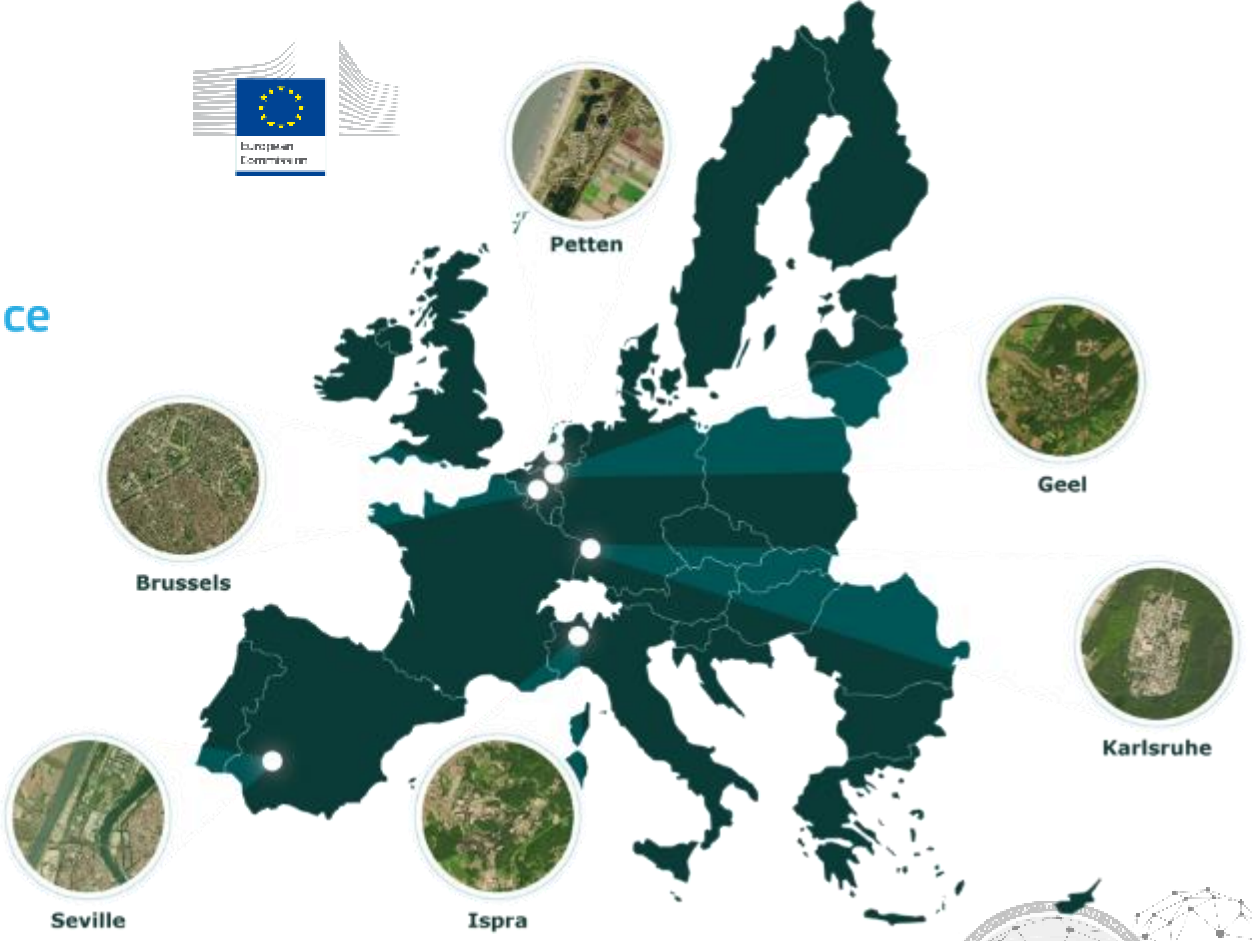
CC0 : Image free of use via pxfuel.com

European
Commission

The European Commission's science and knowledge service

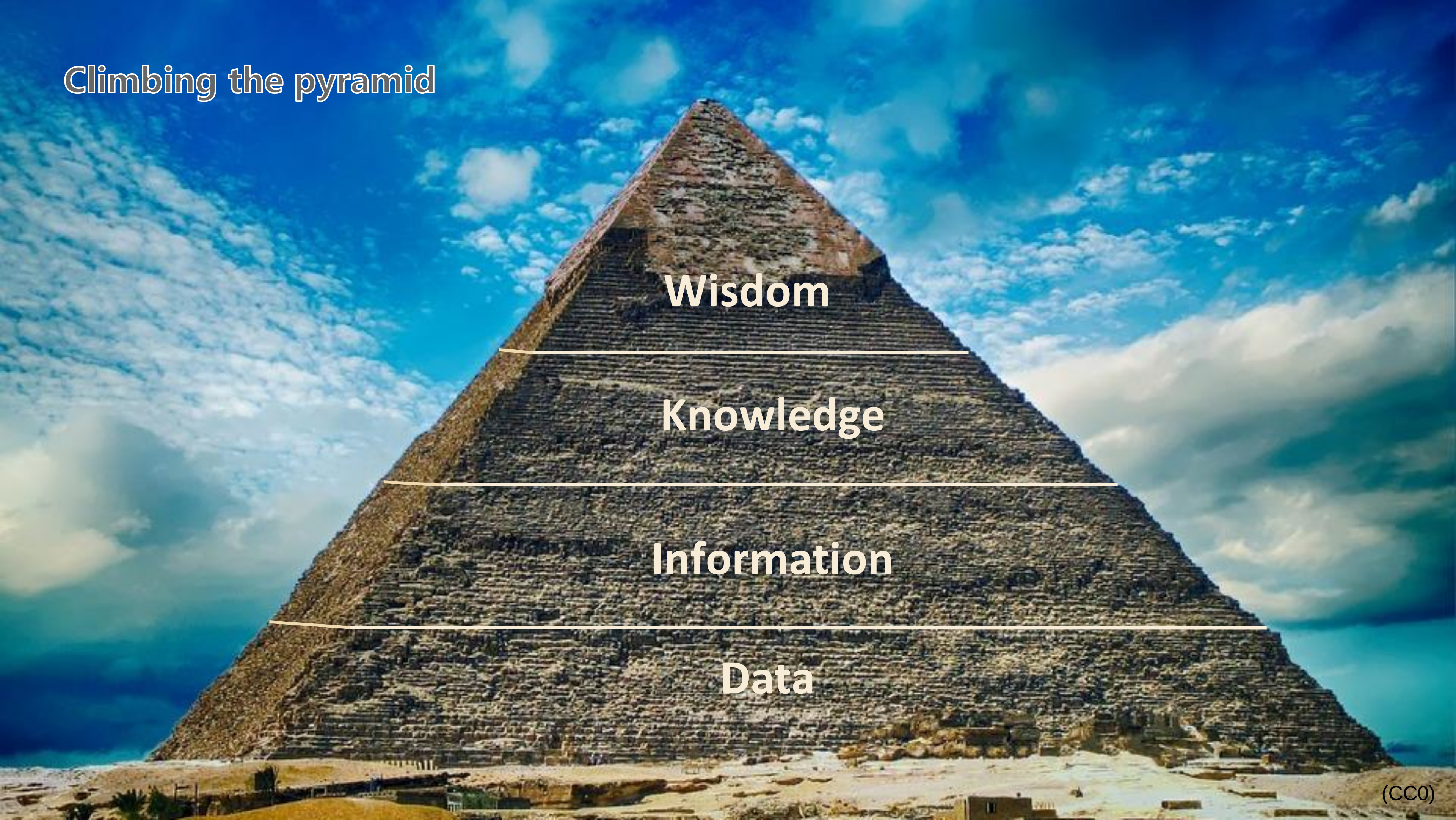
Joint Research Centre

"To play a central role in creating, managing and making sense of collective scientific knowledge for better EU policies"



"Document Deluge"

Climbing the pyramid



Wisdom

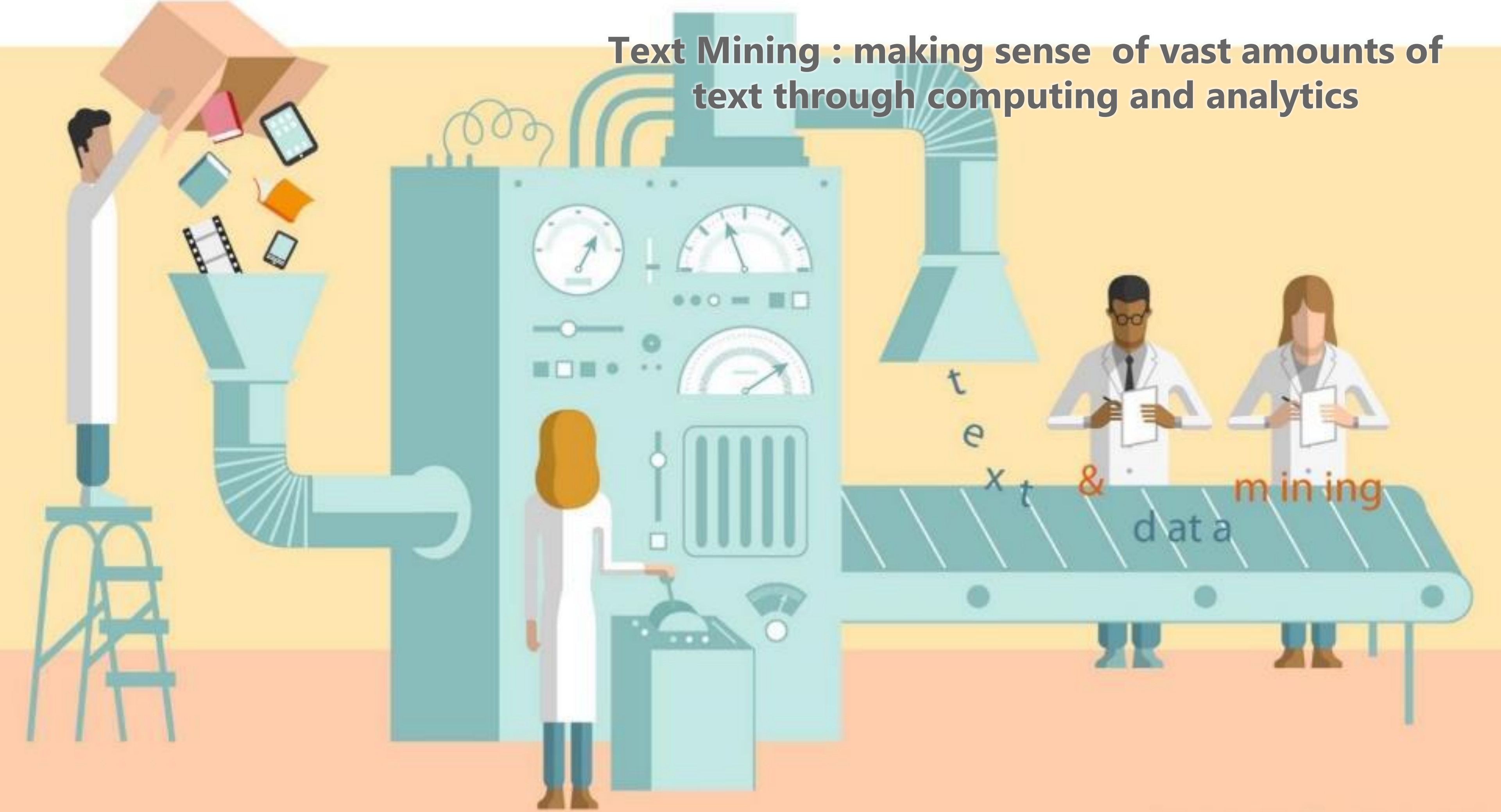
Knowledge

Information

Data

(CC0)

Text Mining : making sense of vast amounts of text through computing and analytics



Original illustration by Davide Bonazzi for CopyrightUser.org

Text Mining & Analysis Comptence Centre

*The Text Mining & Analysis Competence Centre is an in-house **consultancy** and **innovation** service supporting EU Institution's **Policy Makers, Investigators and Analysts** in their **knowledge-intensive tasks** by providing **advice** and advanced analytical **tools** in the field of text mining.*

copyright © fotomek - Stock.Adobe.com

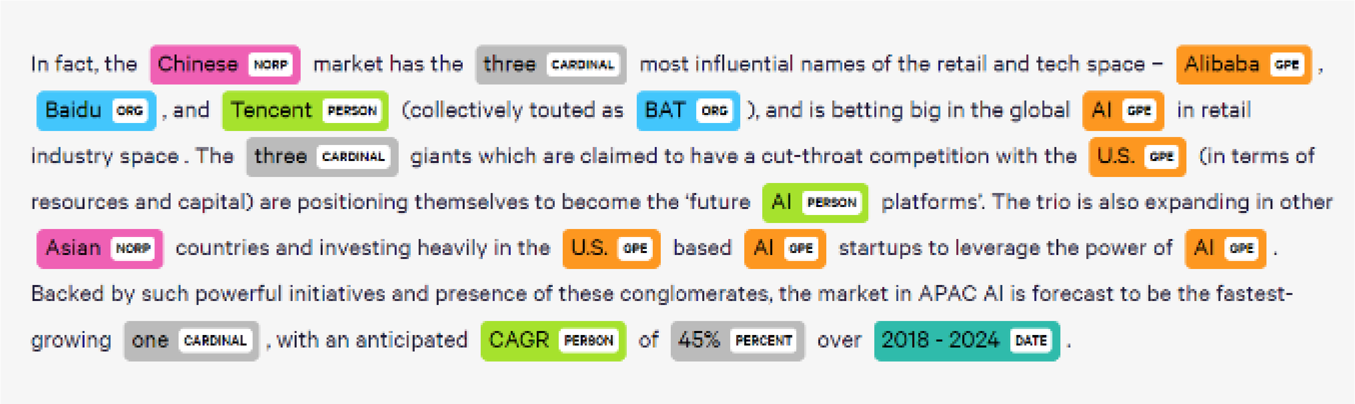
Some typical Text Mining use cases

copyright © Aleksandar Kosev - Stock.Adobe.com

Sorting (aka “classifiers”)

copyright © Rixie - Stock.Adobe.com

Named Entity Recognition (NER)



In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space – Alibaba GPE , Baidu ORG , and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space . The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the 'future AI PERSON platforms'. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE . Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL , with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE .

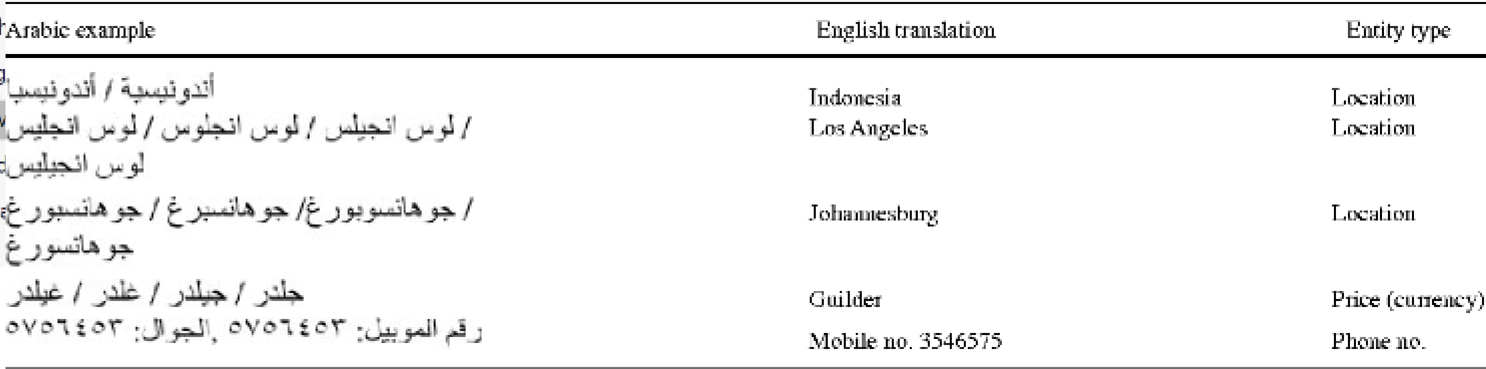
To further elaborate on the geograph

in 2017 DATE and has been leading

credit in the regional trends with ov

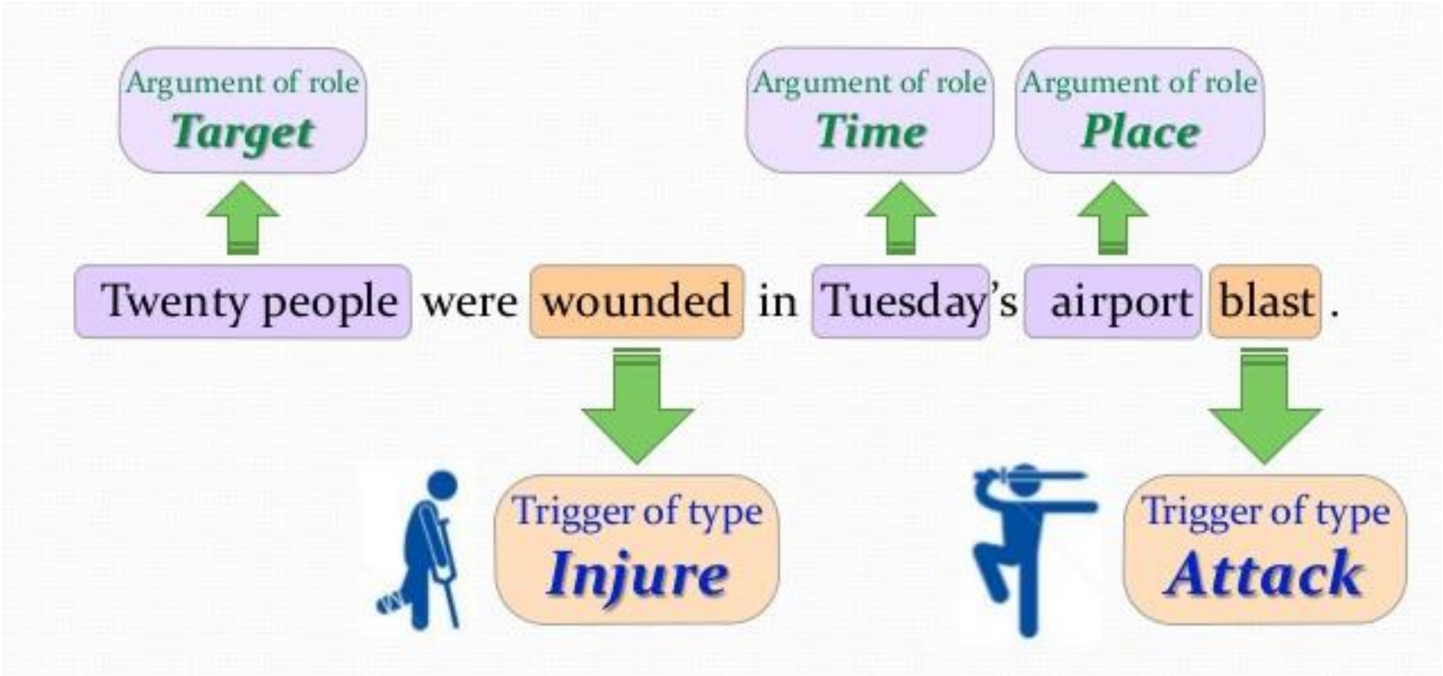
artificial intelligence technology. Add

such as Google ORG , IBM ORG ,

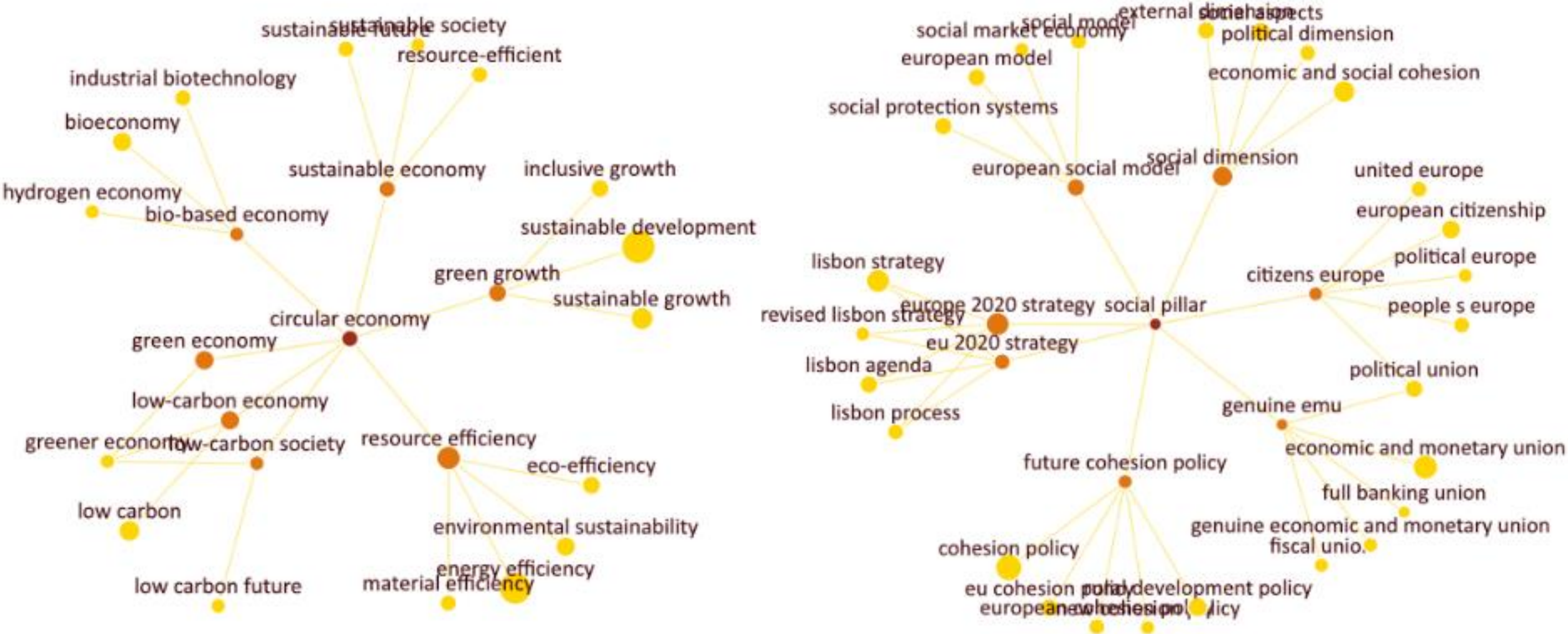


Arabic example	English translation	Entity type
أندونيسية / أندونيسيا	Indonesia	Location
لوس انجيلس / لوس انجلوس / لوس انجليس / لوس انجيليس	Los Angeles	Location
جوهانسوبورغ/ جوهانسبرغ / جوهانسبورغ / جوهانسورغ	Johannesburg	Location
جلدر / جيلدر / غلدر / غيلدر	Guilder	Price (currency)
رقم الموبيل: ٥٧٥٦٤٥٣, الجوال: ٥٧٥٦٤٥٣	Mobile no. 3546575	Phone no.

Event Extraction



Semantic Vector Space Analysis

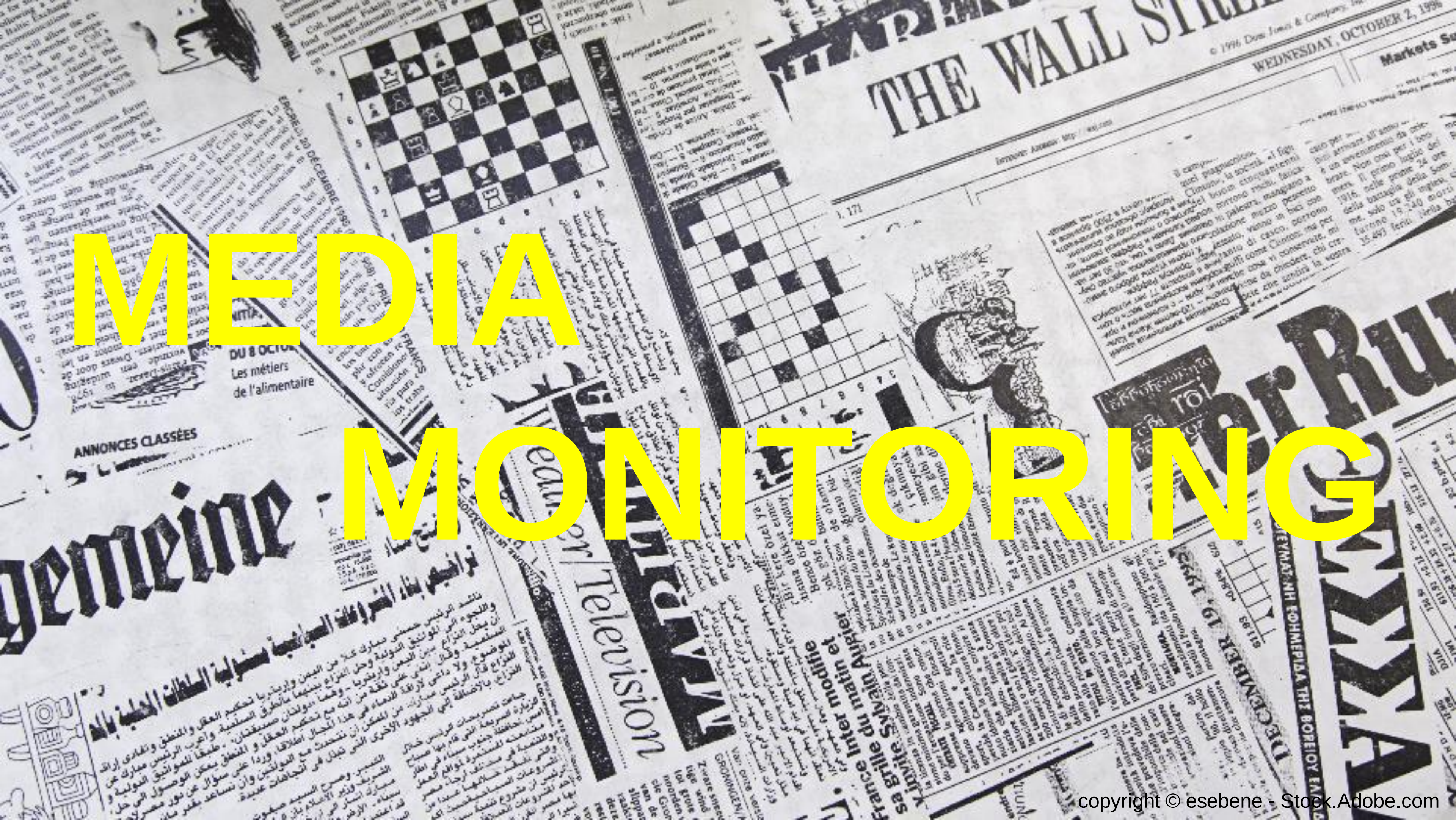


Tonality / Sentiment / Emotion detection

Emotional Criteria	Example topic sentences
Trust	“Forbes Article Predicts Bitcoin Value will "Explode"” / “Good news for the Bitcoin” / “Don't panic, China is NOT banning bitcoin”
Fear	“Mining cartel attack” / “OMG! What has Satoshi created? He has opened Pandora's box” / “We are victims of our own success”
Surprise	“Whatever happened to the Bitcoin Police?” / “I think the rapture happened.....?” / “Blockchain.info "firstbits" changing/disappearing?!”

Flagship projects

Public Domain image – Willem van de Velde via WikiMedia Commons



MEDIA MONITORING

Europe Media Monitor

European Commission

A tool to monitor, analyse and aggregate online and other electronically available sources

Wide Coverage

- Local, Regional, National and International
- Multi-Lingual

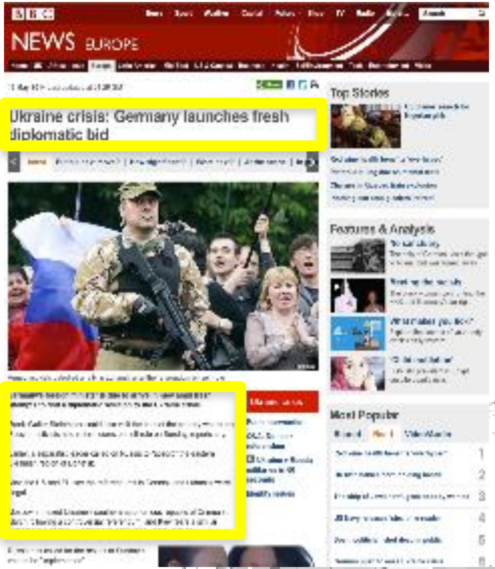
Rapid Updates

- High Frequency (5min)

Sources Covered

- Online media
- News wires
- Blogs
- BBC Monitoring
- Oxford Analytica
- Etc…

Full Text Analysis


NEWS EUROPE

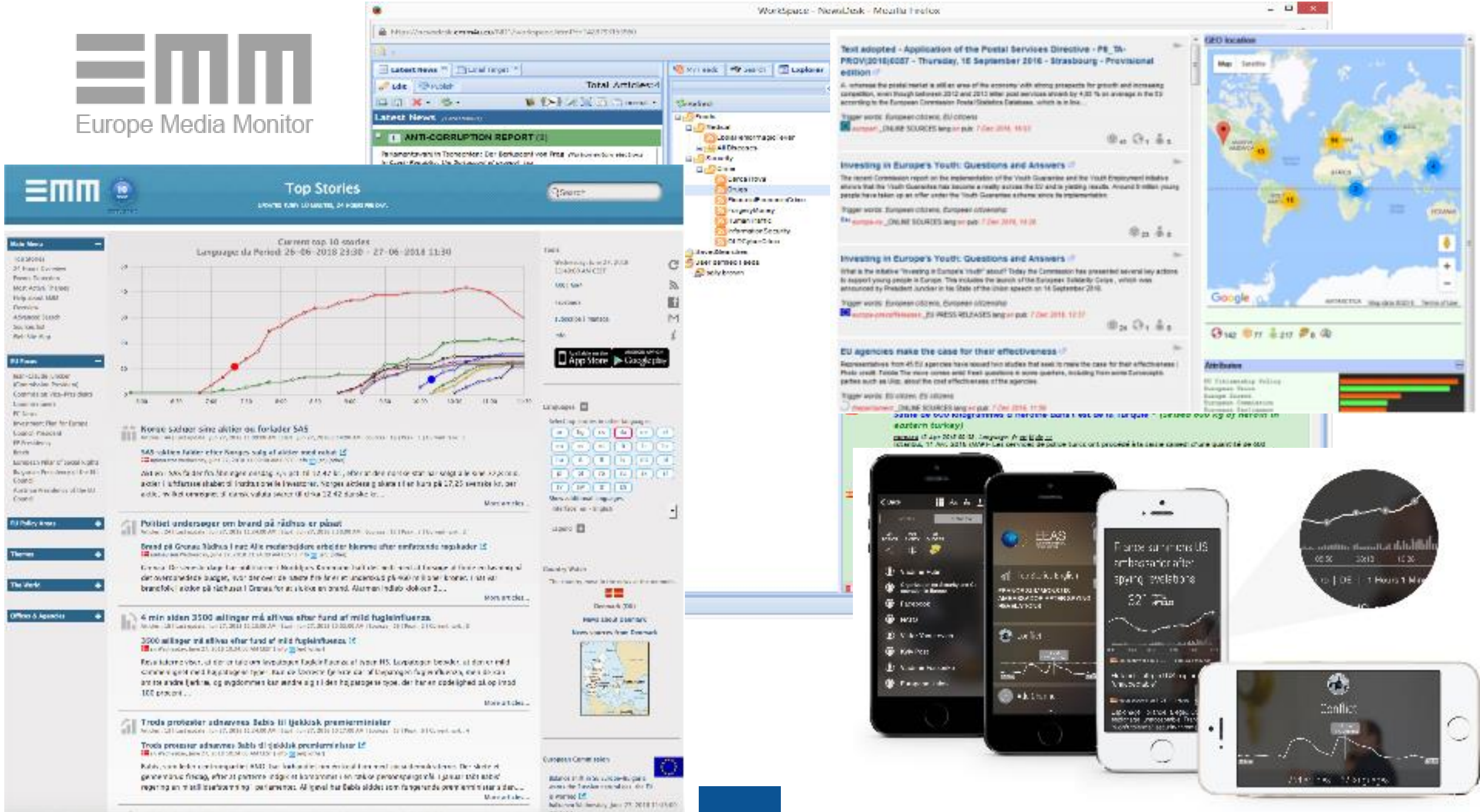
Ukraine crisis: Germany launches fresh diplomatic bid

Top Stories

Features & Analysis

Most Popular

Europe Media Monitor



WorkSpace - NewsDesk - Mozilla Firefox

Latest News

ANTI-CORRUPTION REPORT (3)

Top Stories

Current top 10 stories

Language: da Period: 26-06-2018 23:30 - 27-06-2018 11:30

Norske sælger sine aktier og forlader SAS

Politiet undersøger om brand på rådhus er påsat

Trods protester udnævnes Babis til tjekkisk premierminister

Text adopted - Application of the Postal Services Directive - P8_TA-PROV(2016)0387 - Thursday, 15 September 2016 - Strasbourg - Provisional edition

Investing in Europe's Youth: Questions and Answers

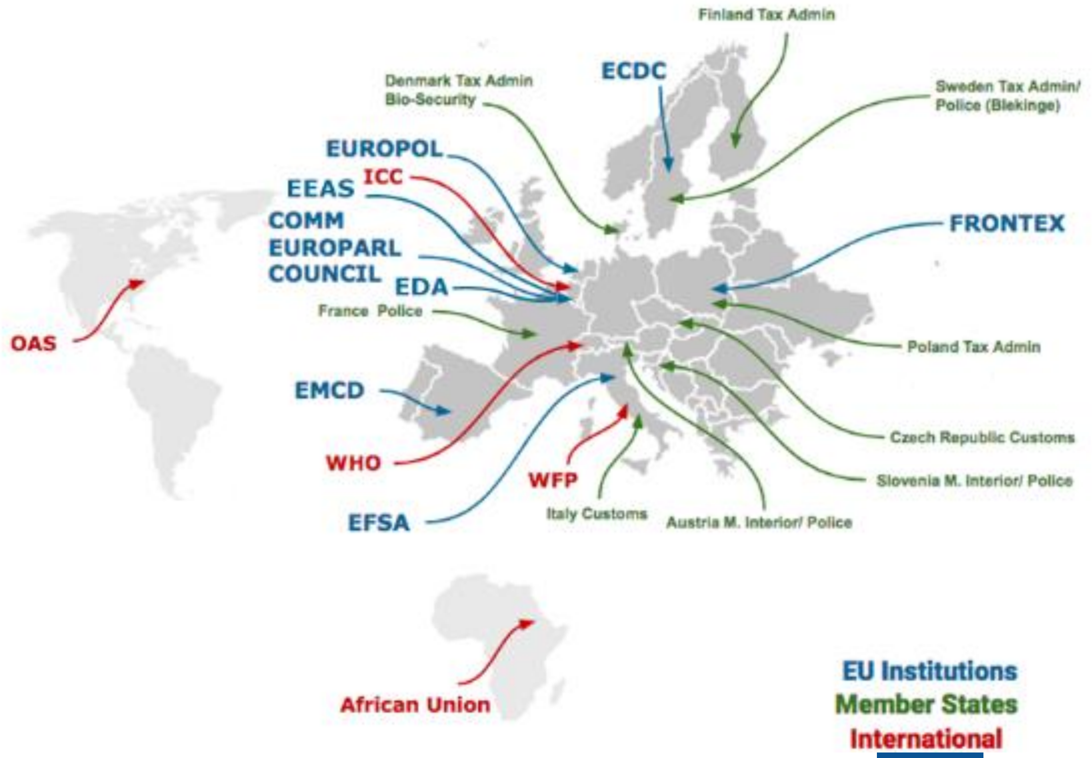
Investing in Europe's Youth: Questions and Answers

EU agencies make the case for their effectiveness

Europe Media Monitor

European Commission

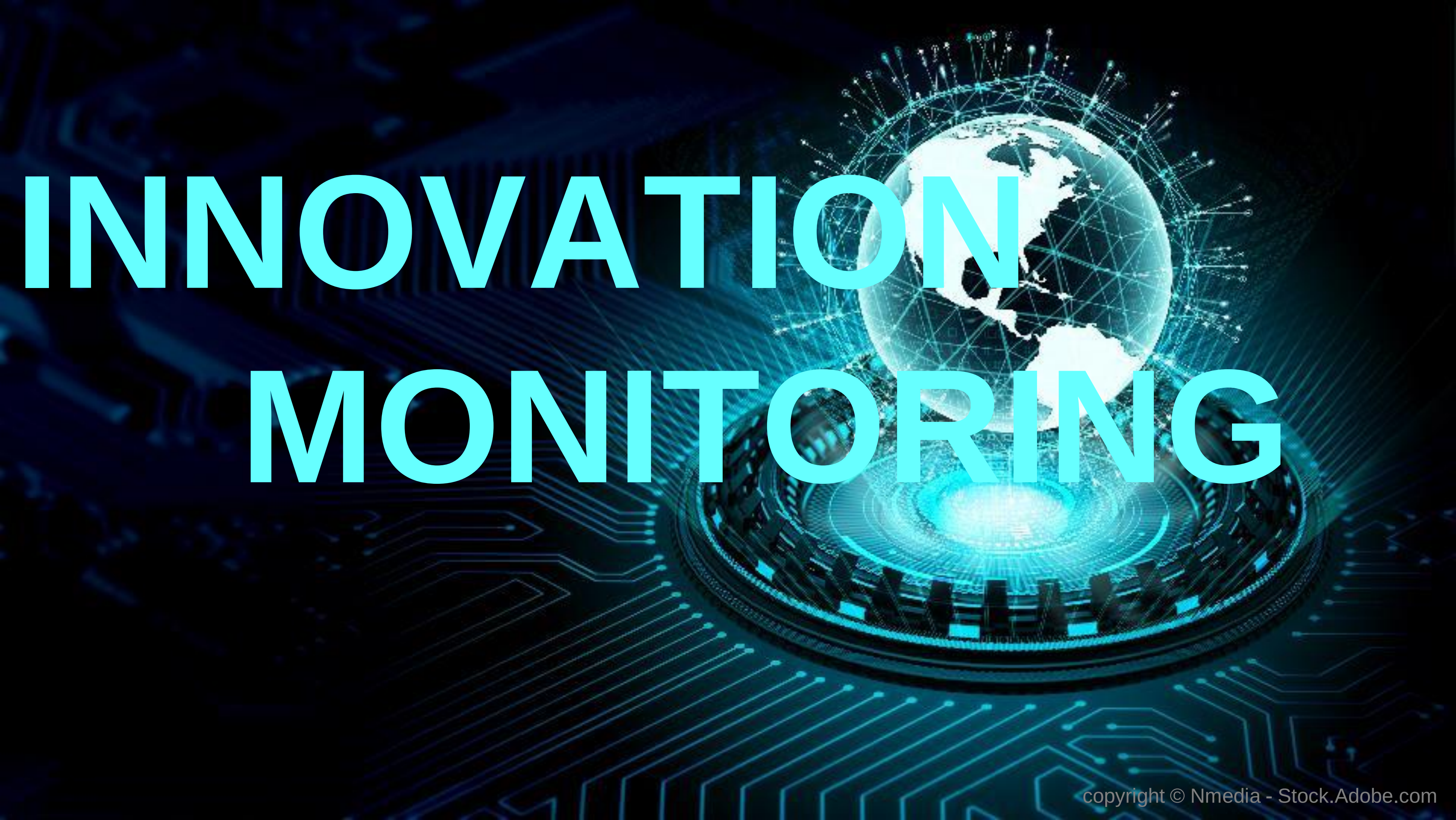
Many, many users worldwide



Related projects

MEDISYS

EMM Open Source Intelligence Suite



INNOVATION MONITORING

copyright © Nmedia - Stock.Adobe.com

TIM Tools for Innovation Monitoring

Data intelligence tools

European Commission

~ 65 M documents

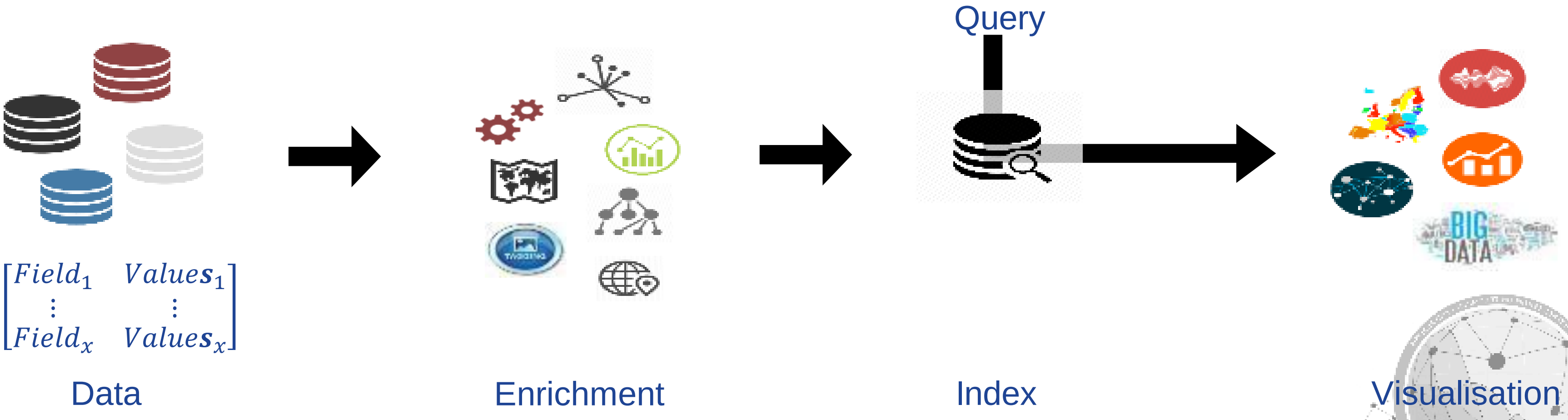
Scopus

The largest abstract and citation database of peer-reviewed literature from more than 5,000 publishers

Europäisches Patentamt
European Patent Office
Office européen des brevets

European Commission CORDIS

EUREKA
innovation across borders



$$\begin{bmatrix} Field_1 & Values_1 \\ \vdots & \vdots \\ Field_x & Values_x \end{bmatrix}$$

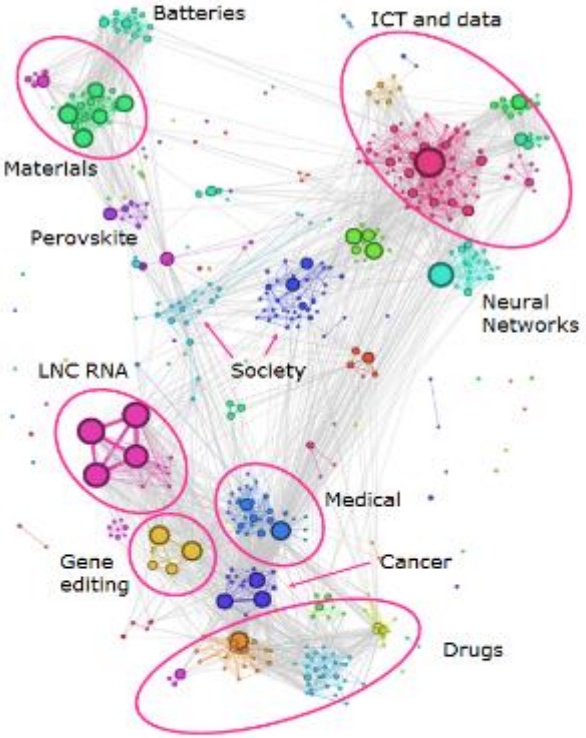
Data

Enrichment

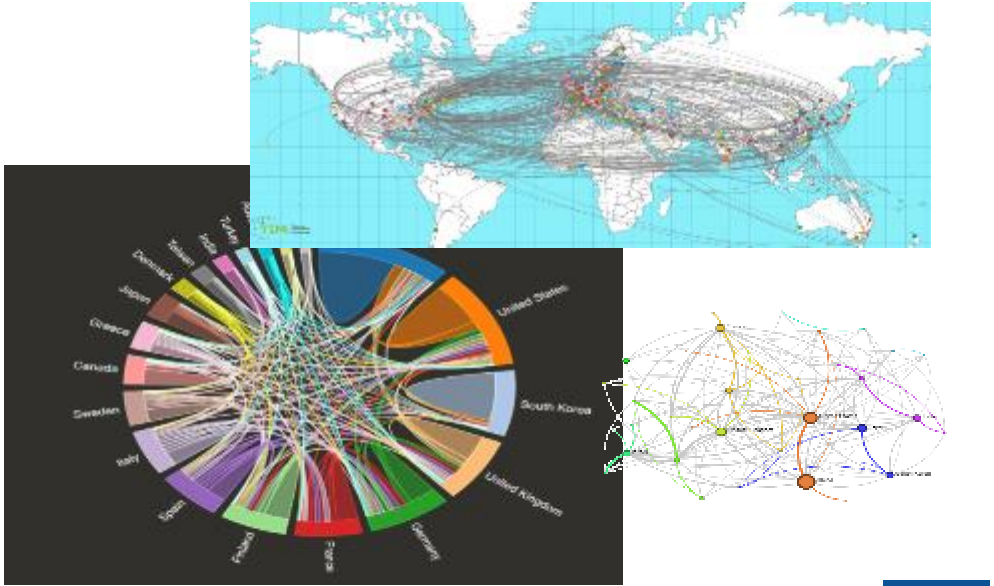
Index

Visualisation

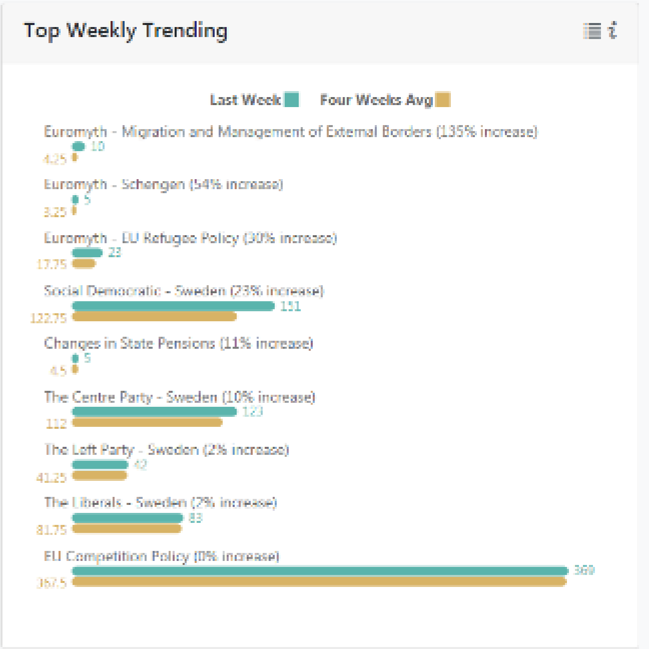
Pre-emerging Technologies detection



Visualisation

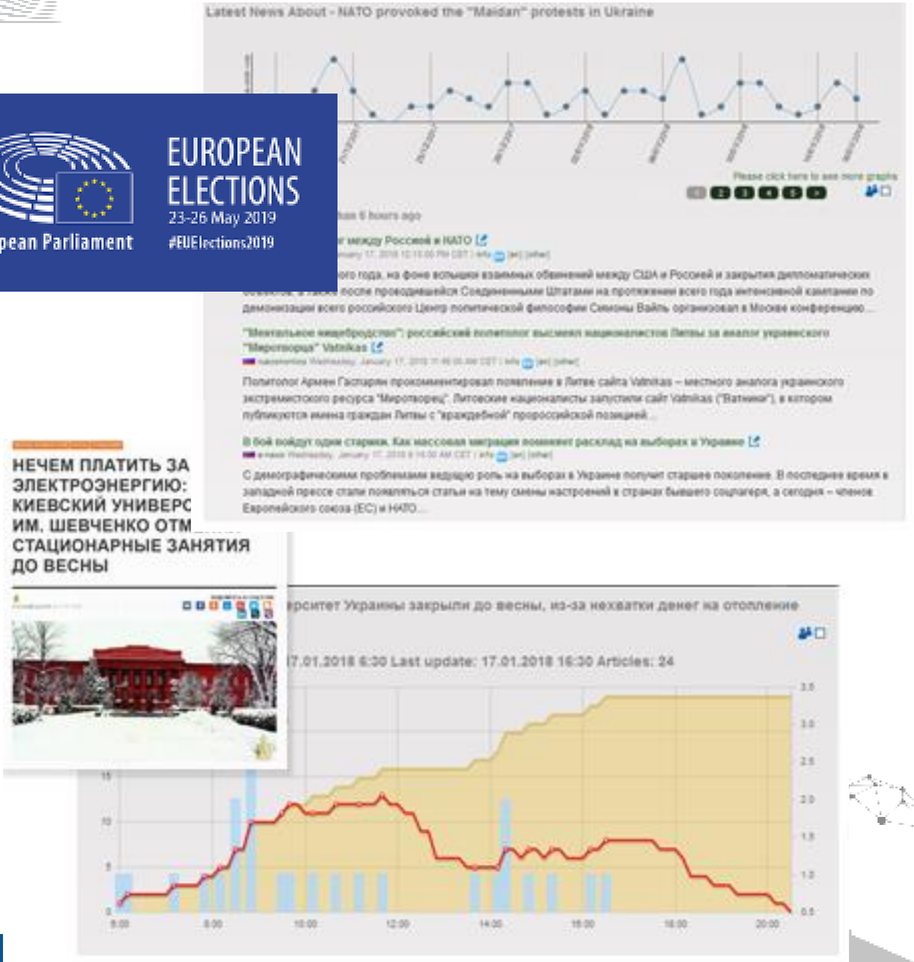


disinformation



European Parliament

EUROPEAN ELECTIONS
23-26 May 2019
#EUElections2019



Latest News About - NATO provoked the "Maidan" protests in Ukraine

...than 6 hours ago

...г между Россией и НАТО

...January 17, 2018 12:15:00 PM CET | info [en] [other]

...ого года, на фоне вспышки взаимных обвинений между США и Россией и закрытия дипломатических объектов, а также после проводившейся Соединенными Штатами на протяжении всего года интенсивной кампании по демонизации всего российского Центр политической философии Симоны Вайль организовал в Москве конференцию ...

"Ментальное [illegible]": российский политолог высмеял националистов Литвы за аналог украинского "Миротворца" Vatnikas

[illegible] Wednesday, January 17, 2018 11:46:00 AM CET | info [en] [other]

Политолог Армен Гаспарян прокомментировал появление в Литве сайта Vatnikas – местного аналога украинского экстремистского ресурса "Миротворец". Литовские националисты запустили сайт Vatnikas ("Ватники"), в котором публикуются имена граждан Литвы с "враждебной" пророссийской позицией...

В бой пойдут одни старики. Как массовая миграция изменит расклад на выборах в Украине

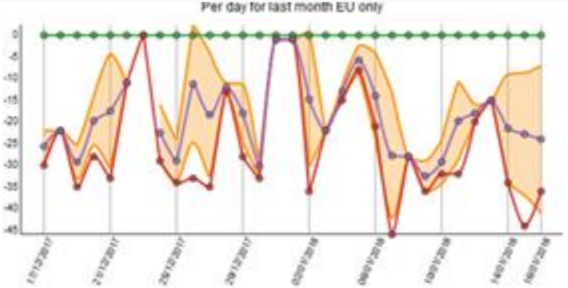
[illegible] Wednesday, January 17, 2018 9:16:00 AM CET | info [en] [other]

С демографическими проблемами ведущую роль на выборах в Украине получит старшее поколение. В последнее время в западной прессе стали появляться статьи на тему смены настроений в странах бывшего соцлагеря, а сегодня – членов Европейского союза (ЕС) и НАТО...

НЕЧЕМ ПЛАТИТЬ ЗА ЭЛЕКТРОЭНЕРГИЮ: КИЕВСКИЙ УНИВЕРС... ИМ. ШЕВЧЕНКО ОТМ... СТАЦИОНАРНЫЕ ЗАНЯТИЯ ДО ВЕСНЫ

...ерситет Украины закрыли до весны, из-за нехватки денег на отопление

...17.01.2018 6:30 Last update: 17.01.2018 16:30 Articles: 24



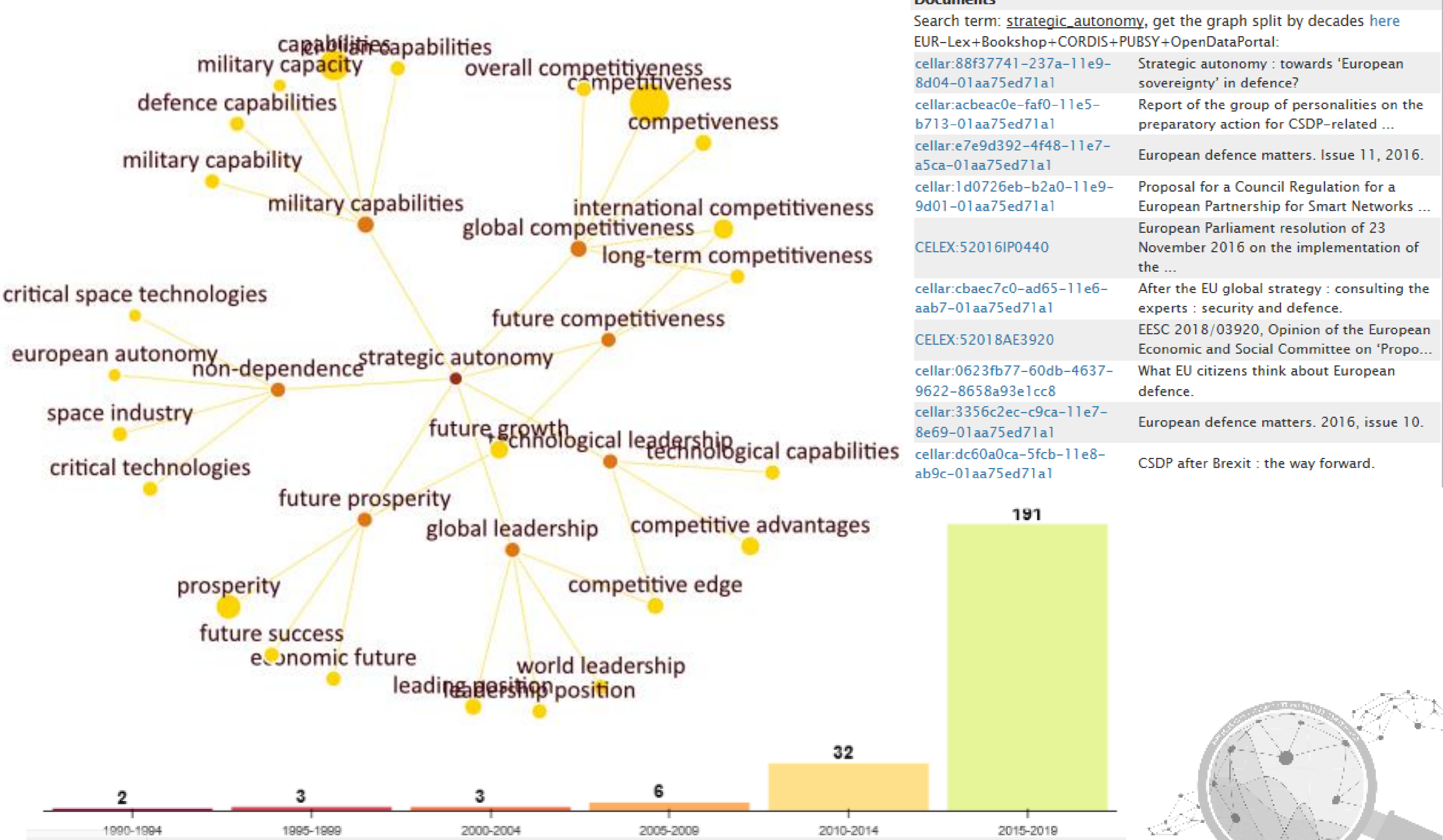
Narrative – "NATO behind Maldan protests in Ukraine"
Unbalanced reporting, not neutral
Intention to polarise

European Commission

Directorate-General for Communication

Semantic Analysis of Legal Texts

copyright © Urupong - Stock.Adobe.com



Documents

Search term: strategic_autonomy, get the graph split by decades here

EUR-Lex+Bookshop+CORDIS+PUBSY+OpenDataPortal:

cellar:88f37741-237a-11e9-8d04-01aa75ed71a1	Strategic autonomy : towards 'European sovereignty' in defence?
cellar:acbeac0e-faf0-11e5-b713-01aa75ed71a1	Report of the group of personalities on the preparatory action for CSDP-related ...
cellar:e7e9d392-4f48-11e7-a5ca-01aa75ed71a1	European defence matters. Issue 11, 2016.
cellar:1d0726eb-b2a0-11e9-9d01-01aa75ed71a1	Proposal for a Council Regulation for a European Partnership for Smart Networks ...
CELEX:52016IP0440	European Parliament resolution of 23 November 2016 on the implementation of the ...
cellar:cbaec7c0-ad65-11e6-aab7-01aa75ed71a1	After the EU global strategy : consulting the experts : security and defence.
CELEX:52018AE3920	EESC 2018/03920, Opinion of the European Economic and Social Committee on 'Propo...
cellar:0623fb77-60db-4637-9622-8658a93e1cc8	What EU citizens think about European defence.
cellar:3356c2ec-c9ca-11e7-8e69-01aa75ed71a1	European defence matters. 2016, issue 10.
cellar:dc60a0ca-5fcb-11e8-ab9c-01aa75ed71a1	CSDP after Brexit : the way forward.

copyright © metamorworks - Stock.Adobe.com

European Commission

The Power of Emotions on the (social) Web

-

when data scientists aim for mindfulness in politics

Matthew KING
Alexandra BALAHUR
Guillaume JACQUET
Bertrand DE LONGUEVILLE

Image credits : WikiMedia Commons

Image credits : WikiMedia Commons

Image credits : © Mountri via Adobe Stock

(social) Web Mining

+

Policy Making

=

?!

emotions politics

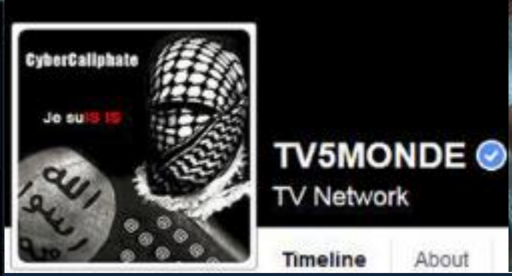
Image credits : © European Parliament/Michel CHRISTEN

Image credits : JRC Report "Understanding our Political Nature"

Image credits : © AFP

Image credits : WikiMedia Commons

WE HAVE A LAW SINCE 1952
THAT ACTUALLY BANS SHARIA LAW ON US SOIL
DO YOU WANT TRUMP TO ENFORCE IT?



RUSSIAN SPIES RUSSIAN SPIES RUSSIAN SPIES
RUSSIAN SPIES RUSSIAN SPIES RUSSIAN SPIES
RUSSIAN SPIES RUSSIAN SPIES RUSSIAN SPIES

LIKE IF YOU AGREE
OBAMA SHOULD BE PROSECUTED
FOR ENABLING VOTER FRAUD!

IF BREXIT
WHY NOT TEXIT?

In North Korea, people are forced to listen to propaganda.
IN THE USA PEOPLE DO IT WILLINGLY.
CNN

warning : troll factory contents

LEFTY CAN'T ENJOY DRINKING MILK
BECAUSE IT PROMOTES WHITE SUPREMACY

NATIONAL SECURITY ADVISER PROVIDES
TRUMP WITH FALSE INFORMATION,
WANTS 150,000 US SOLDIERS IN SYRIA

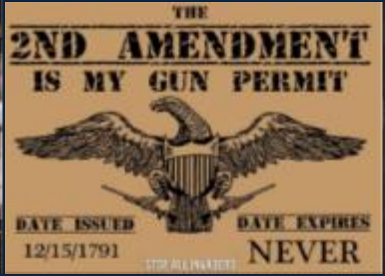
LIKE & SHARE
IF YOU THINK OUR VETERANS
MUST GET BENEFITS BEFORE REFUGEES

Who is going to find out? These women are trash. Nobody's going to believe them.
— Hillary Clinton referring to her husband's sexual misconduct victims.
THE WORDS
OF A FEMINIST?
I DON'T THINK SO.

Source : social media contents as presented in New Knowledge report for US Senate Intelligence Committee "Tactics and Tropes of the IRA"

ANYTIME YOU SEE A NEW DISEASE
JUST POP UP OUT OF NO WHERE
JUST KNOW THEY DIDN'T
DISCOVER IT, THEY INVENTED IT!!!

HILLARY WHEN WIKILEAKS WRITES ABOUT
REPUBLICANS: THEY ARE FREEDOM FIGHTERS!
HILLARY WHEN WIKILEAKS WRITES ABOUT
DEMOCRATS: WE MUST DRONE ASSANGE!



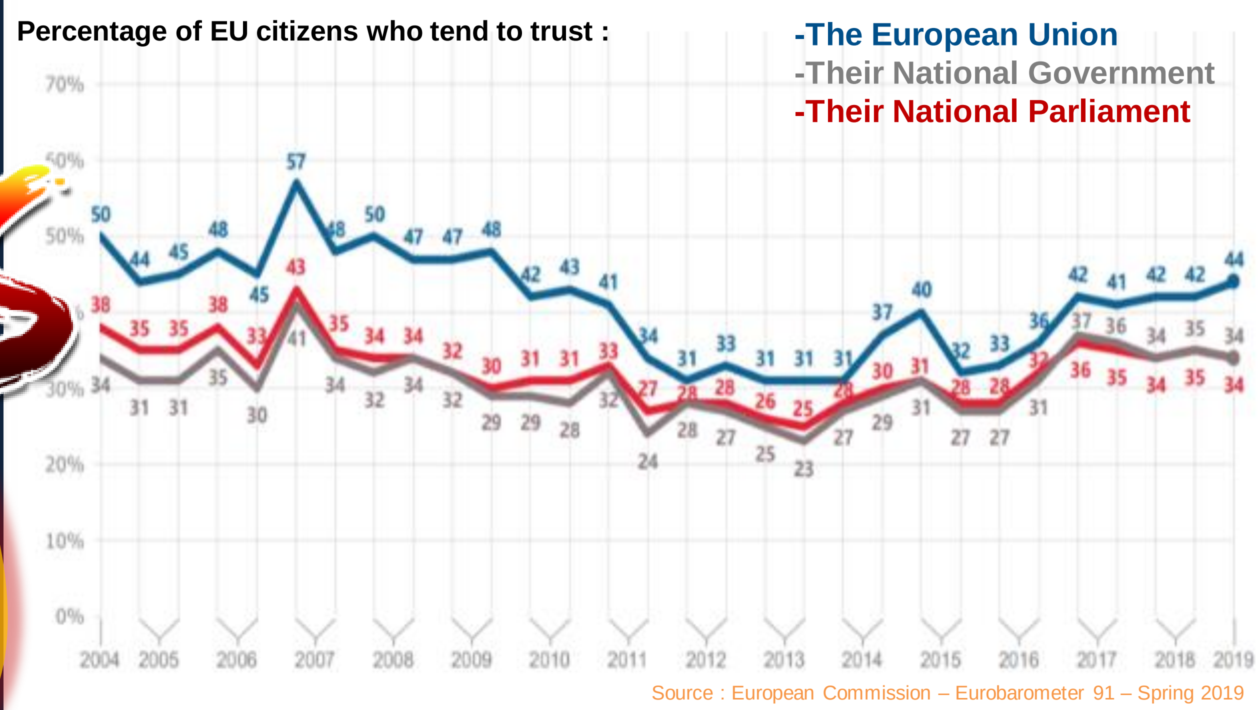
BlackMattersUS.com
DIDN'T BELIEVE THE MEDIA
SO I BECAME ONE

FACEBOOK'S COMMUNITY
STANDARDS DEPARTMENT
FASCIST CENSORSHIP &
DOMESTIC TERRORISM
IN ONE LIBERAL PACKAGE
LIKE IF YOU AGREE

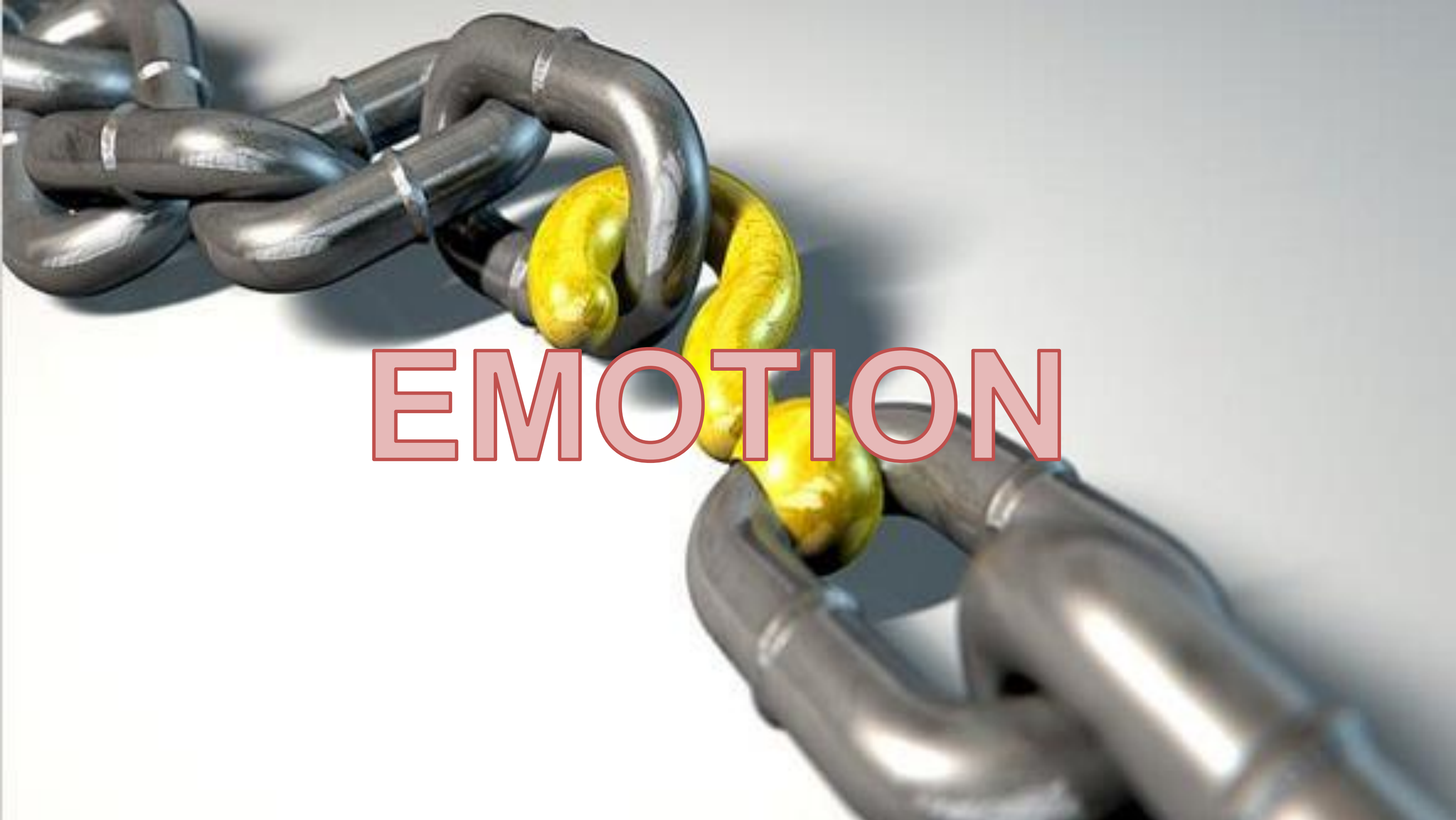
LIKE FOR JESUS TEAM
IGNORE FOR SATAN TEAM

online vs real world ?!

Image credits : unknown authors via Twitter

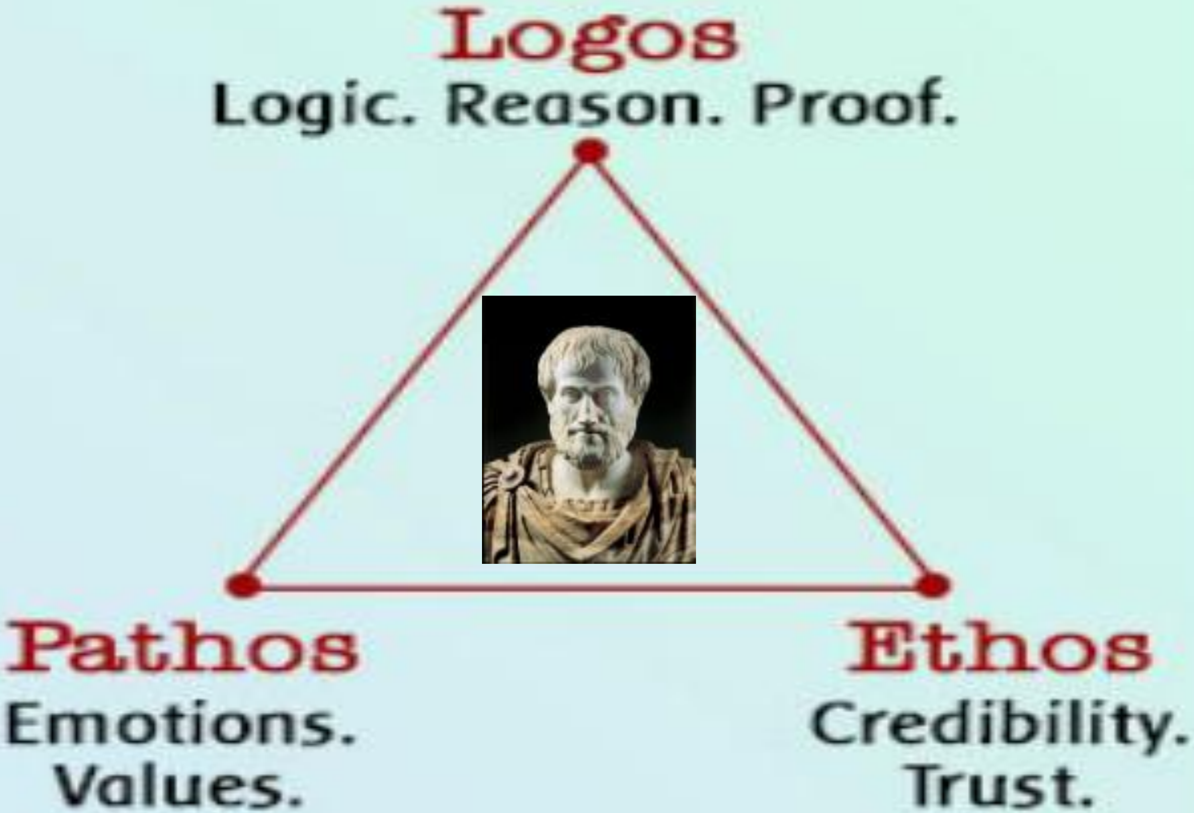

Percentage of EU citizens who tend to trust :

Source : European Commission – Eurobarometer 91 – Spring 2019



EMOTION

The Rhetorical Triangle



Logos

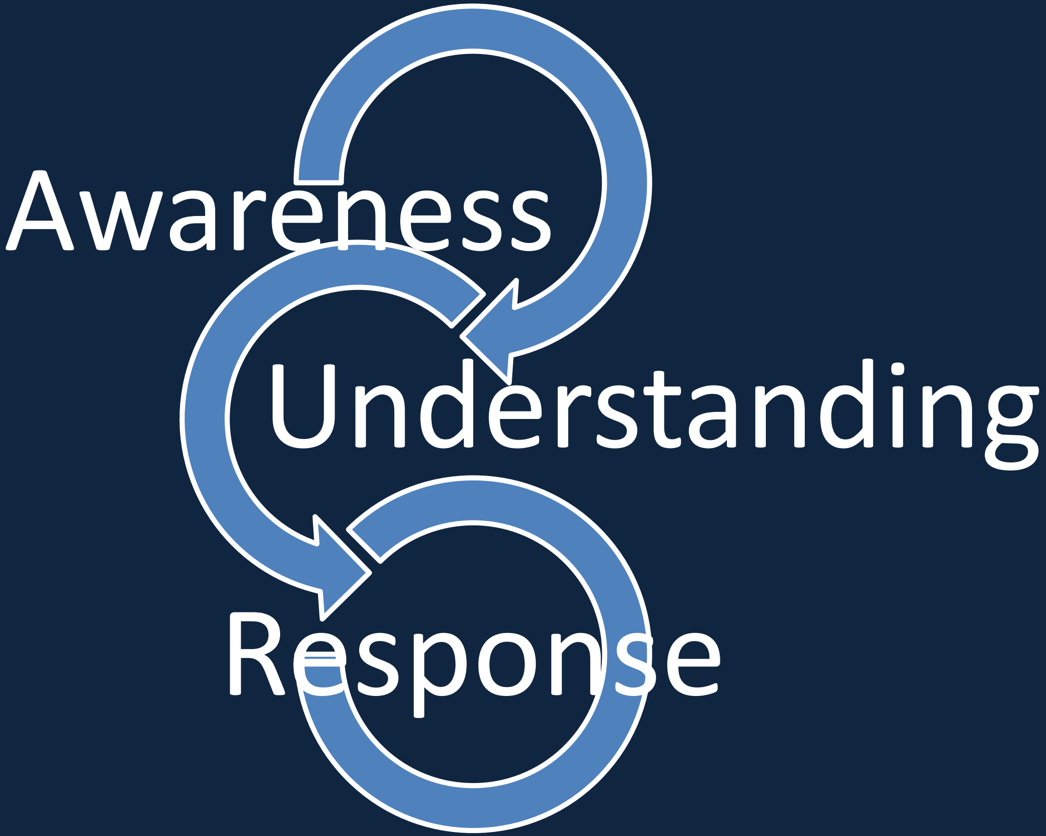
Logic. Reason. Proof.

Pathos

Emotions.
Values.

Ethos

Credibility.
Trust.



Awareness

Understanding

Response

what are we doing about this?

From **Media Monitoring** to Public Opinion Analysis

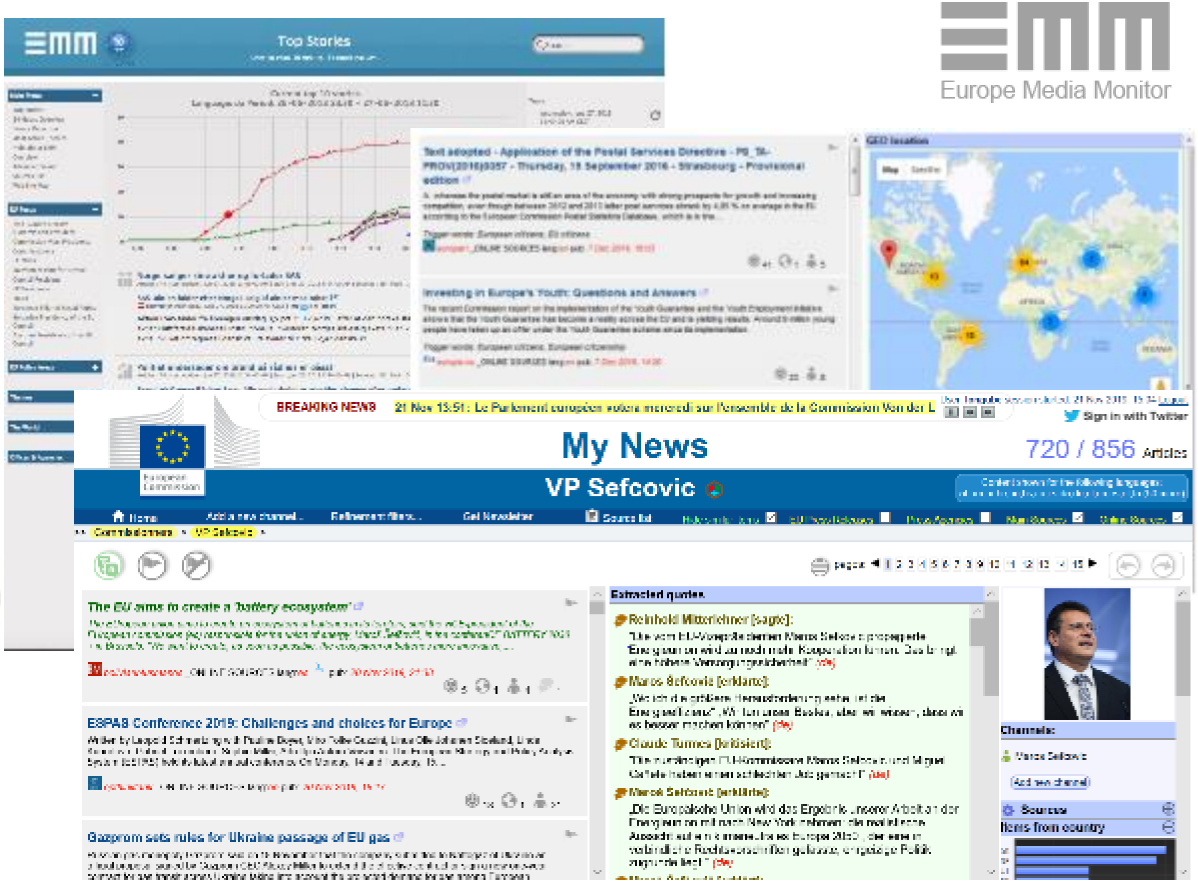
Europe Media Monitor

Large Scale collection :

- 11,000 News Sites
- 325,000 articles per day
- 70 Languages
- 24/7 updates

Advanced inforation extraction

- Named Entity Recognition
- Quotes Extraction
- Events Extraction
- **Sentiment Analysis**



Text adopted - Application of the Postal Services Directive - P8_TA-PROV(2016)0357 - Thursday, 15 September 2016 - Strasbourg - Provisional edition

Investing in Europe's Youth: Questions and Answers

21 Nov 15:51 : Le Parlement européen votera mercredi sur l'ensemble de la Commission Von der L

My News

720 / 856 Articles

VP Sefcovic

The EU aims to create a 'battery ecosystem'

ESPAS Conference 2019: Challenges and choices for Europe

Gazprom sets rules for Ukraine passage of EU gas

Reinhold Mitterlehner [sagte]:

Maros Šefčovič [erklärte]:

Claude Turmes [kritisiert]:

Marcel Šefčovič [erklärte]:

Available Right Now

From Media Monitoring to Public Opinion Analysis

EMM Large Scale News Corpus :

- 10+ years history
- About 1 Billion articles available for analysis

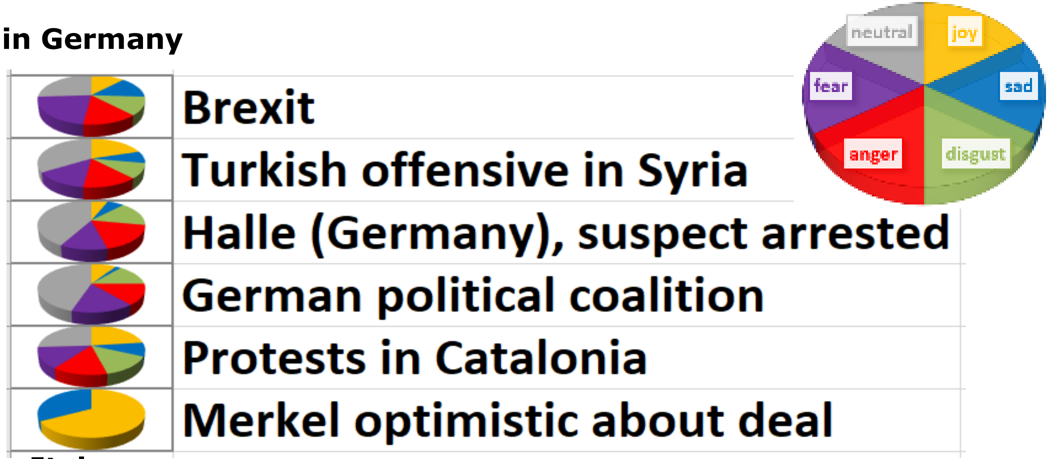
Media Analytics Capability:

- Sentiment per news story
- Sentiment per Person, organisation, topic, …
- Sentiment country/lanaguage
- **Combination of all of the above!**

Work in Progress

Top News Stories with Sentiment in end October 2019 …

… in Germany



Brexit

Turkish offensive in Syria

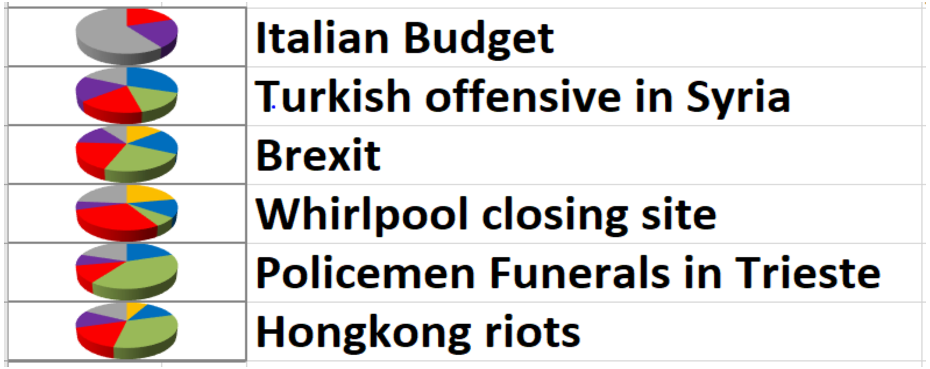
Halle (Germany), suspect arrested

German political coalition

Protests in Catalonia

Merkel optimistic about deal

… in Italy



Italian Budget

Turkish offensive in Syria

Brexit

Whirlpool closing site

Policemen Funerals in Trieste

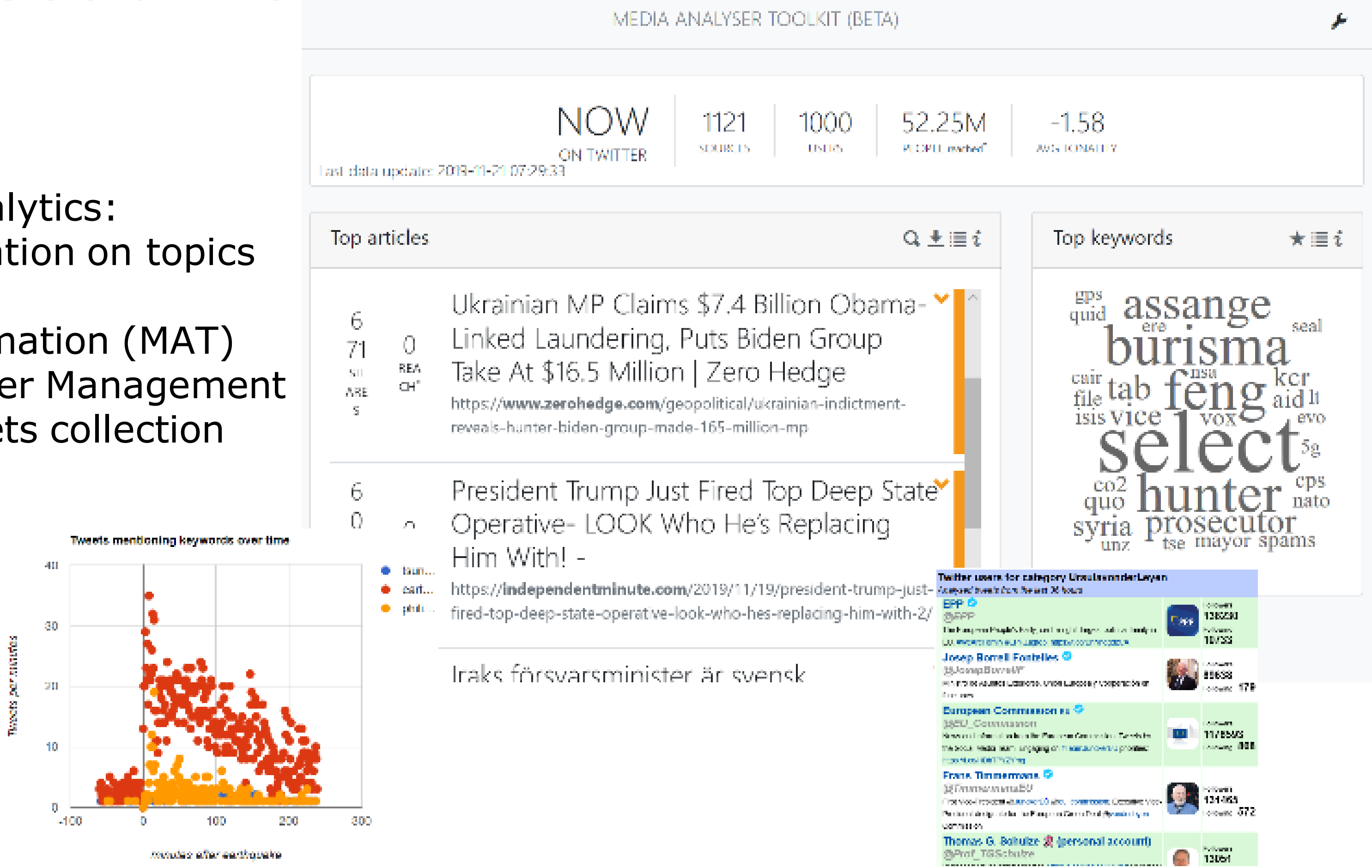
Hongkong riots

From **Social Media Analytics** to Argument Mining

Social Media Analytics:

- Richer information on topics (MyNews)
- Fight Disinformation (MAT)
- Natural Disaster Management via rapid Tweets collection

✓ **Available Right Now**



MEDIA ANALYSER TOOLKIT (BETA)

NOW ON TWITTER | 1121 SOURCES | 1000 USERS | 52.25M PEOPLE reached | -1.58 AVG TONALITY

Last data update: 2019-11-21 07:29:33

Top articles

6 71 SHARES | 0 REACH | Ukrainian MP Claims $7.4 Billion Obama-Linked Laundering, Puts Biden Group Take At $16.5 Million | Zero Hedge https://www.zerohedge.com/geopolitical/ukrainian-indictment-reveals-hunter-biden-group-made-165-million-mp

6 0 | 0 | President Trump Just Fired Top Deep State Operative- LOOK Who He's Replacing Him With! - https://independentminute.com/2019/11/19/president-trump-just-fired-top-deep-state-operative-look-who-hes-replacing-him-with-2/

Iraks försvarsminister är svensk

Top keywords

Twitter users for category UrsulavonderLeyen

Analysed tweets from the last 36 hours

EPP @EPP — Followers 136230, Following 10733

Josep Borrell Fontelles @JosepBorrellF — Followers 85638, Following 178

European Commission @EU_Commission — Followers 1178563, Following 888

Frans Timmermans @TimmermansEU — Followers 134463, Following 572

Thomas G. Schulze (personal account) @Prof_TGSchulze — Followers 13061

From Social Media Analysis to **Argument Mining**

The vision :

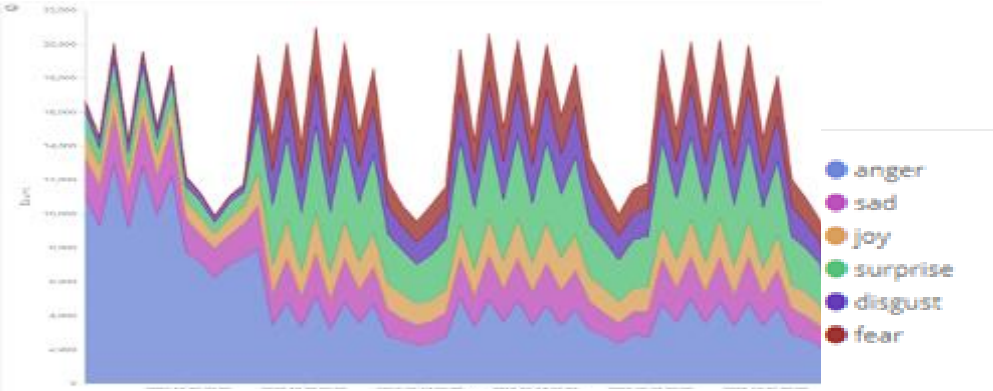
Fine-grained understanding of public debates, through strengthened JRC capabilities on Sentiment Analysis and Argument Mining

Way forward :

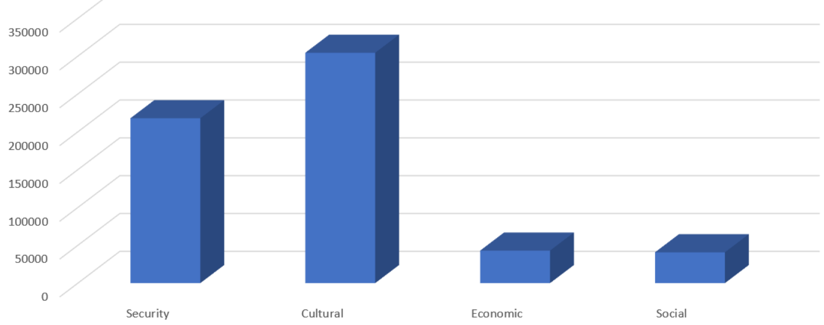
- Generalise successful ‘one shot’ experiments into a stable system
- Include more soucres of public debate : official communication of political actors, eurobarometer, public consultations (e.g. Future of Europe)

Work in Progress

Volume of Tweets related to an EU policy and expressing a given sentiment, over time



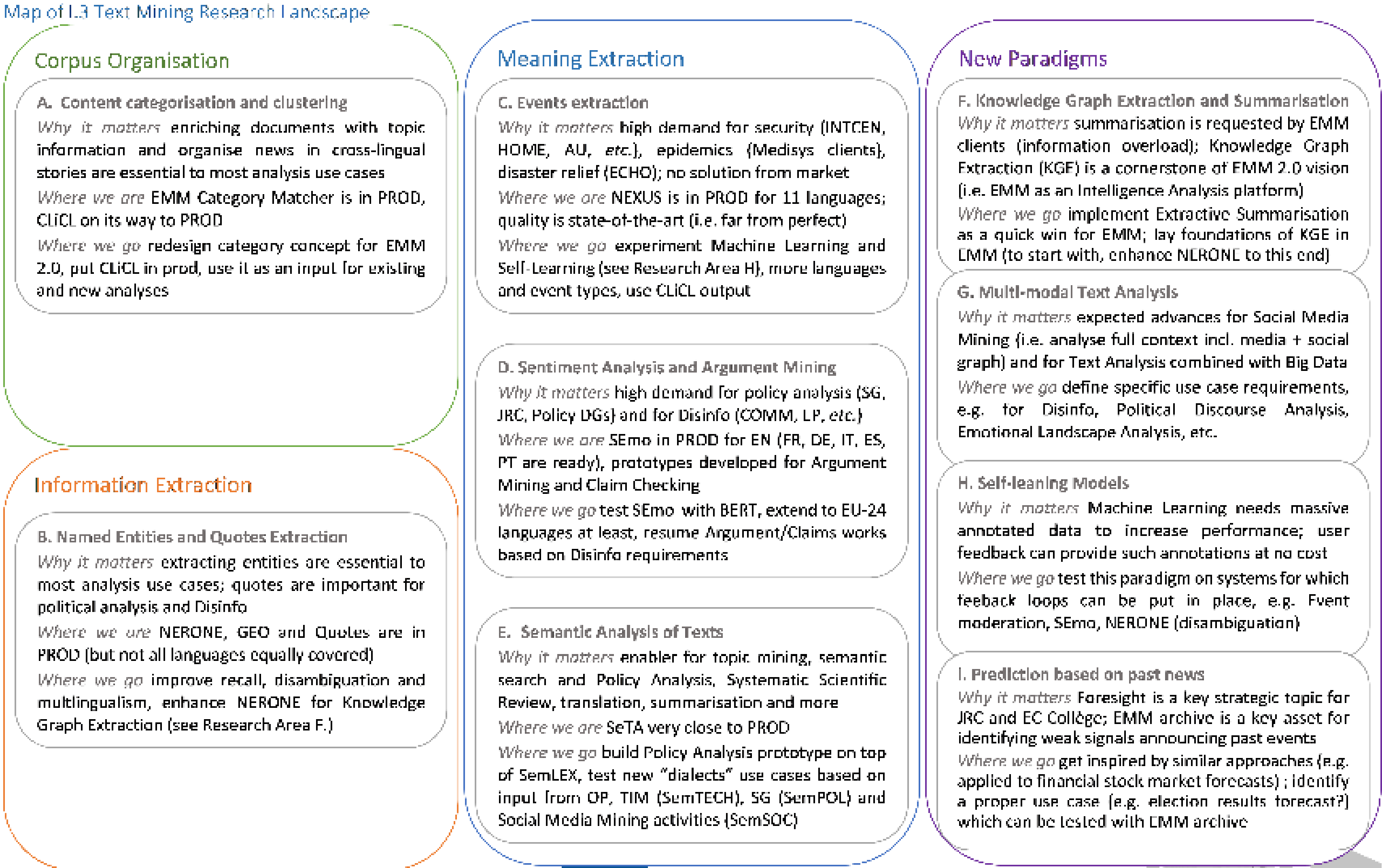
Volume of Tweets arguing against EU migration policy, by argument type, over a given period



Text Mining & Analysis
Comptence Centre

Way forward : our research agenda

CC0 : Image free of use via pxfuel.com



Map of I.3 Text Mining Research Landscape

Corpus Organisation

A. Content categorisation and clustering

Why it matters enriching documents with topic information and organise news in cross-lingual stories are essential to most analysis use cases

Where we are EMM Category Matcher is in PROD, CLiCL on its way to PROD

Where we go redesign category concept for EMM 2.0, put CLiCL in prod, use it as an input for existing and new analyses

Information Extraction

B. Named Entities and Quotes Extraction

Why it matters extracting entities are essential to most analysis use cases; quotes are important for political analysis and Disinfo

Where we are NERONE, GEO and Quotes are in PROD (but not all languages equally covered)

Where we go improve recall, disambiguation and multilingualism, enhance NERONE for Knowledge Graph Extraction (see Research Area F.)

Meaning Extraction

C. Events extraction

Why it matters high demand for security (INTCEN, HOME, AU, *etc.*), epidemics (Medisys clients), disaster relief (ECHO); no solution from market

Where we are NEXUS is in PROD for 11 languages; quality is state-of-the-art (i.e. far from perfect)

Where we go experiment Machine Learning and Self-Learning (see Research Area H), more languages and event types, use CLiCL output

D. Sentiment Analysis and Argument Mining

Why it matters high demand for policy analysis (SG, JRC, Policy DGs) and for Disinfo (COMM, LP, *etc.*)

Where we are SEmo in PROD for EN (FR, DE, IT, ES, PT are ready), prototypes developed for Argument Mining and Claim Checking

Where we go test SEmo with BERT, extend to EU-24 languages at least, resume Argument/Claims works based on Disinfo requirements

E. Semantic Analysis of Texts

Why it matters enabler for topic mining, semantic search and Policy Analysis, Systematic Scientific Review, translation, summarisation and more

Where we are SeTA very close to PROD

Where we go build Policy Analysis prototype on top of SemLEX, test new "dialects" use cases based on input from OP, TIM (SemTECH), SG (SemPOL) and Social Media Mining activities (SemSOC)

New Paradigms

F. Knowledge Graph Extraction and Summarisation

Why it matters summarisation is requested by EMM clients (information overload); Knowledge Graph Extraction (KGE) is a cornerstone of EMM 2.0 vision (i.e. EMM as an Intelligence Analysis platform)

Where we go implement Extractive Summarisation as a quick win for EMM; lay foundations of KGE in EMM (to start with, enhance NERONE to this end)

G. Multi-modal Text Analysis

Why it matters expected advances for Social Media Mining (i.e. analyse full context incl. media + social graph) and for Text Analysis combined with Big Data

Where we go define specific use case requirements, e.g. for Disinfo, Political Discourse Analysis, Emotional Landscape Analysis, etc.

H. Self-leaning Models

Why it matters Machine Learning needs massive annotated data to increase performance; user feedback can provide such annotations at no cost

Where we go test this paradigm on systems for which teeback loops can be put in place, e.g. Event moderation, SEmo, NERONE (disambiguation)

I. Prediction based on past news

Why it matters Foresight is a key strategic topic for JRC and EC College; EMM archive is a key asset for identifying weak signals announcing past events

Where we go get inspired by similar approaches (e.g. applied to financial stock market forecasts) ; identify a proper use case (e.g. election results forecast?) which can be tested with EMM archive

Thank you for your attention

questions

are warmly welcome

European Commission - Joint Research Centre -
Competences Directorate -Text and Data Mining Unit

bertrand.de-longueville@ec.europa.eu