

# THE EUROPE MEDIA MONITOR

Web Text Mining for News

8-9 aprile 2021

 LIVE SU WEBEX

**SESSIONE 1**  
**8 APRILE | h14.00**

# **JENS LINGE AND GUILLAUME JACQUET**

- ★ Introduction to the Europe Media Monitor
- ★ Presentation of some specific elements inside the EMM
- ★ Presentation of some use cases related to health domain



**SESSIONE 2**  
**9 APRILE | h14.00**

# **MARIA LETIZIA ALBERGATI AND HARALD PROELL**

- ★ Presentation of DG.COMM, one of the key users of EMM
- ★ Description how EMM can be integrated in a specific environment
- ★ Hands-on to practice how EMM can be used

# EMM

# Europe Media Monitor

**Guillaume Jacquet, Jens Linge**  
**JRC.I.3 – Text and Data Mining Unit**  
**European Commission – Joint Research Centre**

modulo Jean Monnet CoEur, 8th and 9th of April 2021



# The European Commission's science and knowledge service

Joint Research Centre

Directorate I - Competences  
JRC I.3 Text and Data Mining Unit



# DG Joint Research Centre



Established in

**1957**



Around

**3000**

Staff



**10**

Directorates



**6**

Locations in 5  
Member  
States



Seville

Petten  
Geel  
Brussels  
Karlsruhe  
Ispra



**>1000**

Publications  
per year

**42**

Large scale  
facilities



European  
Commission

# JRC.I.3 TEXT AND DATA MINING



**Data4Policy**



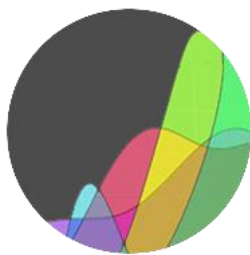
**Statistics and IT for  
Anti-Fraud, Security and Trade**



**Tools for Innovation  
Monitoring**



**Big Data Platform**



**Text Mining and Analysis  
Competence Centre**



**Europe Media Monitor**

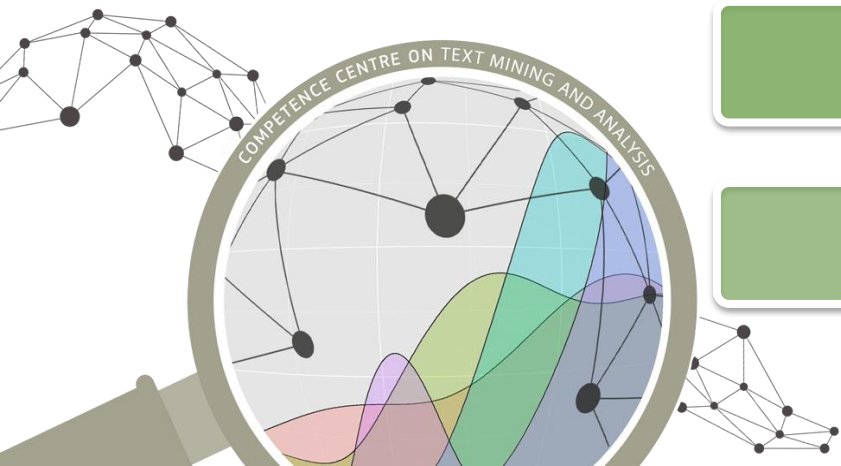
# EMM: Media Monitoring for Open Source information analysis

- Started in 2002 to provide Internet Media Monitoring to complement/replace traditional cut-and paste
- Information for decision makers, spokes persons: mainly "political events"
- Initially main customer DG COMM. Evolved into monitoring for other kinds of events: disasters and crisis, security, health threats, conflict etc.



# EMM: INFORMATION FROM TEXT

Objective: Find knowledge that is hidden in text



**Gathers**



**Enriches**



**Selects**



**Presents**

Results: Find perhaps  
100 relevant articles a  
day from 350 000

# WHAT IS EMM?

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

## Wide Coverage

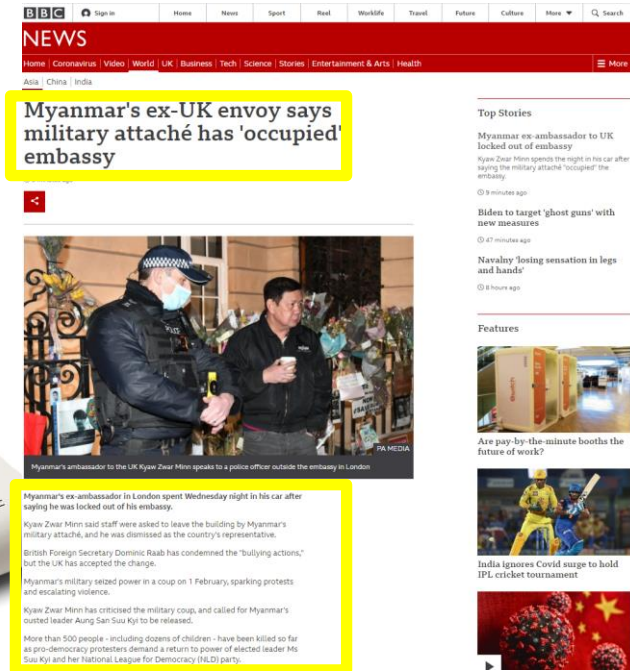
- Local, Regional, National and International
- Multi-Lingual

## Rapid Updates

- High Frequency (5min)


## Sources Covered

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



# EMM Architecture


## Custom Domain (Sources and Topics)



Gathers  
**~300000**  
new news articles  
per day

**70** in  
languages  
~8000  
news portals  
world-wide

**Classifies all news according  
to hundreds of subjects and  
countries**

 **24**  
hours  
per day

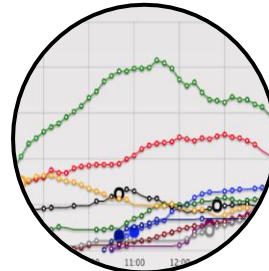
**7**  
days  
a week

Operational since  
**2002**

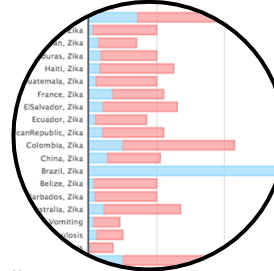
## EMM Open Source Monitoring Engine



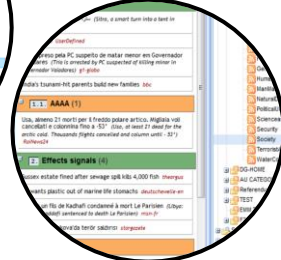
Newsbrief



Medisys



NewsDesk



Other  
Uses



## End Use Applications

# WHAT CAN EMM DO FOR YOU

## **Reduce amount of data to be analysed**

Filter out relevant information based on predefined categories

## **Provide additional information**

Other categories, places, people, quotes

## **Provide automatic alerting functionality**

Simple email Breaking news

Breaking news per category

## **Replaces manual searching in selected sources**

# WHAT Does the EMM Engine Do?

Provides real time analysis and alerting functions

- Topic identification
- Entity recognition
- Automatically detects stories and follows them.
- Breaking news detection
- ... much much more



Full List :

Automatic language recognition

Entity extraction

Quote extraction

Geotagging

Tonality

Duplicate detection

Categorisation

Filtering

Clustering

Statistics

Event extraction



# MACHINE TRANSLATION



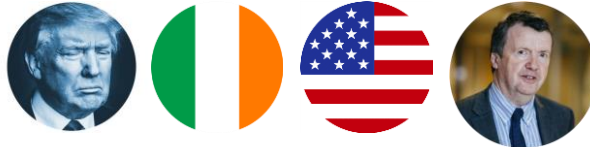
- Statistical Machine Translation (*planning for NMT*)
- Only from original language to English.
- 14 EU languages
  - Czech, Danish, Dutch, Finnish, French, German, Greek, Italian, Latvian, Lithuanian, Polish, Portuguese, Spanish and Swedish
- + Russian, Farsi, Chinese\* and Arabic.

\* Translation from Chinese is done using a commercial tool

# ENTITY DETECTION

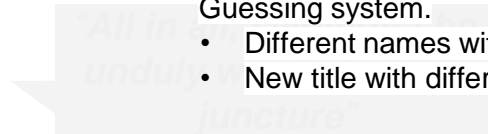


US president Donald Trump intention to cut the US corporate tax rate to 15 per cent may not be achievable with the border adjustment tax, which might have funded the reduction, now off the table. Irish tax experts remain convinced that US tax reform, whatever shape it takes, will not adversely damage Ireland's economic prospects, suggesting incentives for US multinationals to base operations here will continue. "All in all, I wouldn't be unduly worried at this juncture," Alan McQuaid said.



Our entity DB is built on an Entity Guessing system.

- Different names with same title
- New title with different name



# QUOTE DETECTION



US president Donald Trump intention to cut the US corporate tax rate to 15 per cent may not be achievable with the border adjustment tax, which might have funded the reduction, now off the table. Irish tax experts remain convinced that US tax reform, whatever shape it takes, will not adversely damage Ireland's economic prospects, suggesting incentives for US multinationals to base operations here will continue. "All in all, I wouldn't be unduly worried at this juncture," Alan McQuaid said.



Quotes are both "From & "About" in 19 Languages:

ar,bg,da,de,em,es,et,fr,it,nl  
no,pl,pt,ro,ru,sl,sv,sw,tr



*"All in all, I wouldn't be unduly worried at this juncture"* Alan McQuaid

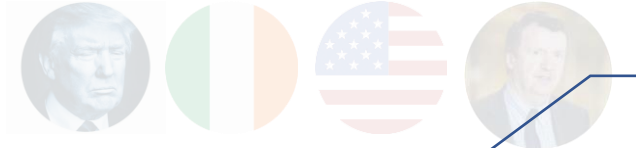




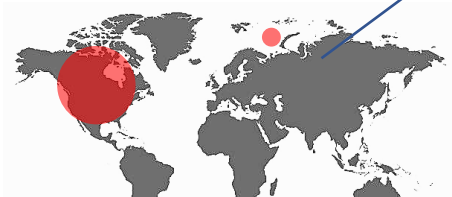
# GEO DETECTION



US president Donald Trump intention to cut the US corporate tax rate to 15 per cent may not be achievable with the border adjustment tax, which might have funded the reduction, now off the table. Irish tax experts remain convinced that US tax reform, whatever shape it takes, will not adversely damage Ireland's economic prospects, suggesting incentives for US multinationals to base operations here will continue. "All in all, I wouldn't be unduly worried at this juncture," Alan McQuaid said.



Database with  
1.3 million entries



# SENTIMENT ANALYSIS



US president Donald Trump intention to cut the US corporate tax rate to 15 per cent may not be achievable with the border adjustment tax, which might have funded the reduction, now off the table. Irish tax experts remain convinced that US tax reform, whatever shape it takes, **will not adversely damage Ireland's economic** prospects, suggesting incentives for US multinationals to base operations here will continue. **"All in all, I wouldn't be unduly worried at this juncture,"** Alan McQuaid said.



*"All in all, I wouldn't be unduly worried at this juncture"*



Sentiment:  
+  
**POSITIVE**



European  
Commission

# CATEGORIZATION

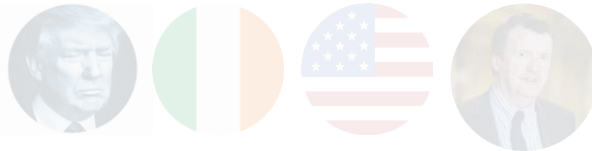


*Put in small example of a category*

US president Donald Trump intention to cut the US **corporate tax** rate to 15 per cent may not be achievable with the **border adjustment tax**, which might have funded the reduction, now off the table. Irish **tax experts** remain convinced that US **tax reform**, whatever shape it takes, will not adversely damage Ireland's **economic prospects**, suggesting incentives for US multinationals to base operations here will continue. "All in all, I wouldn't be unduly worried at this juncture," Alan McQuaid said.



**Categorization:**  
**Taxation**  
**Economy**  
**EU-US trade**



*"All in all, I wouldn't be unduly worried at this juncture"*



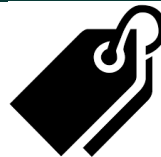
European  
Commission

# Structured data: logical view



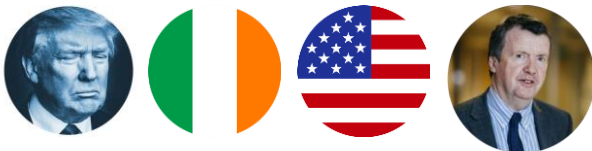
Language:  
English

US president Donald Trump intention to cut the US corporate tax rate to 15 per cent may not be achievable with the border adjustment tax, which might have funded the reduction, now off the table. Irish tax experts remain convinced that US tax reform, whatever shape it takes, will not adversely damage Ireland's economic prospects, suggesting incentives for US multinationals to base operations here will continue. *"All in all, I wouldn't be unduly worried at this juncture,"* Alan McQuaid said.



Categorization:  
Taxation  
Economy  
EU-US trade

Entities:



Quotes:

*"All in all, I wouldn't be unduly worried at this juncture"*

Geo:



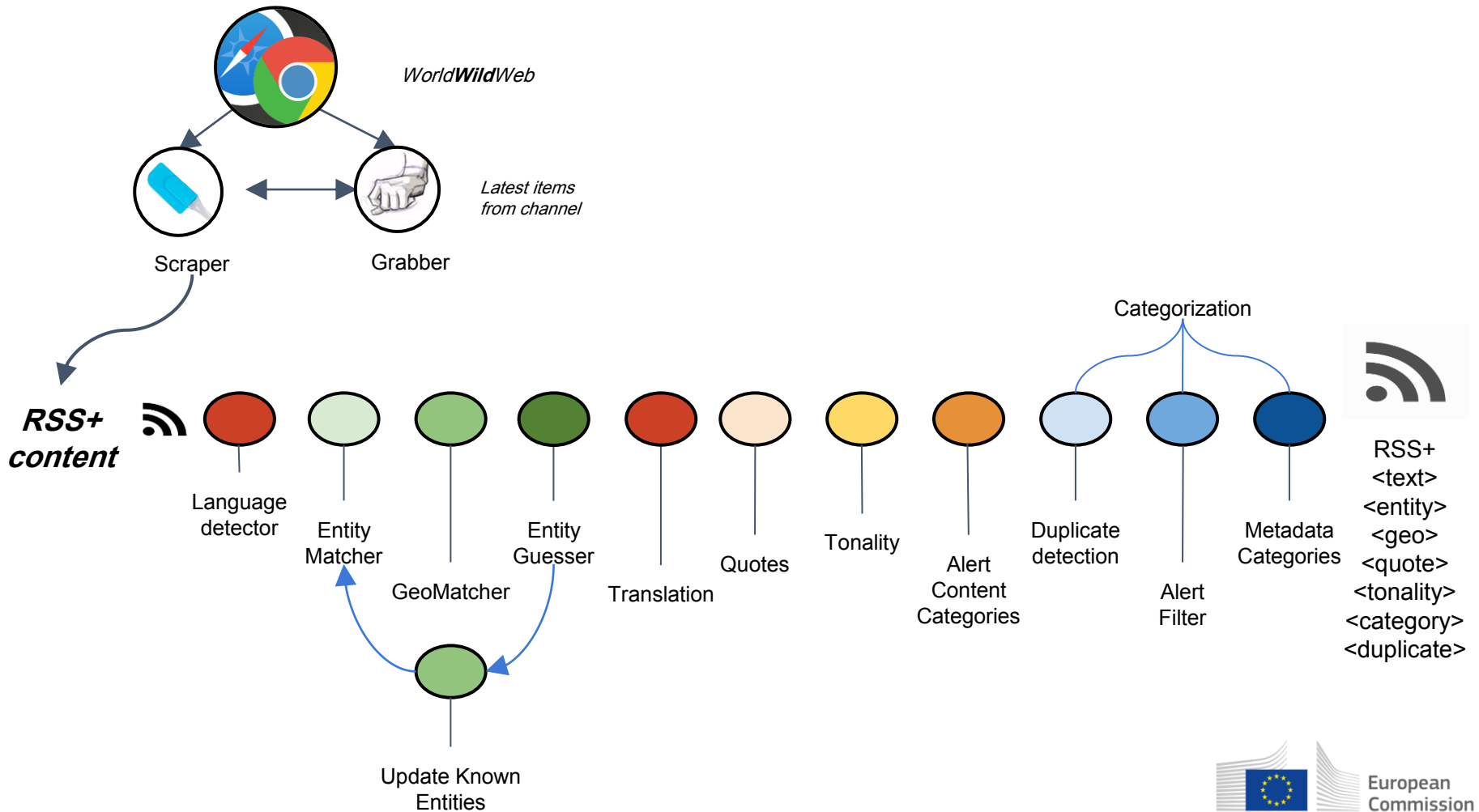
Sentiment:

+

POSITIVE



European  
Commission



# EVENT METADATA EXTRACTION

Turning natural language  
into structured data.

Today applicable only to a  
predefined set of event types  
(e.g. man-made or natural  
disasters)



## At least 11 protesters killed in northwestern Myanmar in clashes with security forces

Articles: 3, Last update: 2021-04-08T10:15+0200, Start: 2021-04-08T09:42+0200



Event type: Disorder/Protest/Mutiny

Who is dead: 11|people

Number killed: 11

Perpetrator: troops

Snippet: Security forces kill 11 protesters in northwestern Myanmar 26 minit yang ...

Place: Myanmar. [i](#)

## At least 11 protesters killed in northwestern Myanmar in clashes with security forces

Thursday, April 8, 2021 10:38:00 AM CEST

Myanmar security forces clashed with anti-coup demonstrators in the town of Taze in the northwest overnight and at least 11 of the protesters were killed, domestic media said on Thursday. The troops fired at protesters in a clash that started on Wednesday evening, the Irrawaddy and Myanmar Now news outlets said....

[More articles...](#)

<https://emm.newsbrief.eu/NewsBrief/eventedition/en/latest.html>



European  
Commission

# APPLICATION AREA: NEWS MONITORING

EMM

- 11.000 News Sites
- 350.000 articles per day
- 70 Languages
- 3000 Categories/Topics
- 60000 keywords
- Runs 24/7
- Visitors 25000/day

Domains:

- Border Security, Cybercrime & Fighting Crime
- Socio Political Event Monitoring
- Public Health Threats & Food Safety
- Conflict Early Warning
- Science and the Citizen



Myanmar's ex-UK envoy says military attaché has 'occupied' embassy



Myanmar's ex-ambassador in London spent Wednesday night in his car after saying he was locked out of his embassy.  
Kyau Zwar Min said staff were asked to leave the building by Myanmar's military attaché, and he was dismissed as the country's representative.  
British Foreign Secretary Dominic Raab has condemned the 'bullying actions', but the UK has accepted the change.  
Myanmar's military seized power in a coup on 1 February, sparking protests and escalating violence.  
Kyau Zwar Min has criticised the military coup, and called for Myanmar's ousted leader Aung San Suu Kyi to be released.  
More than 500 people - including dozens of children - have been killed so far as pro-democracy protesters demand a return to power of elected leader Ms Suu Kyi and her National League for Democracy (NLD) party.

## Top Stories

Myanmar ex-ambassador to UK locked out of embassy  
Kyau Zwar Min spends the night in his car after saying the military attaché 'occupied' the embassy.  
9 minutes ago

Biden to target 'ghost guns' with new measures  
47 minutes ago

Navalny 'losing sensation in legs and hands'  
8 hours ago

## Features



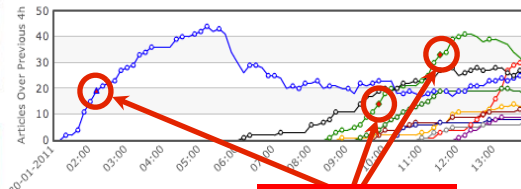
Are pay-by-the-minute booths the future of work?



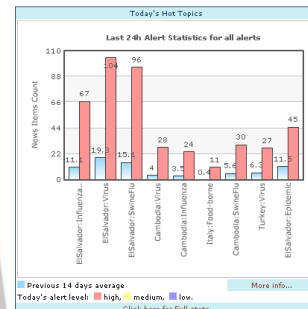
India ignores Covid surge to hold IPL cricket tournament



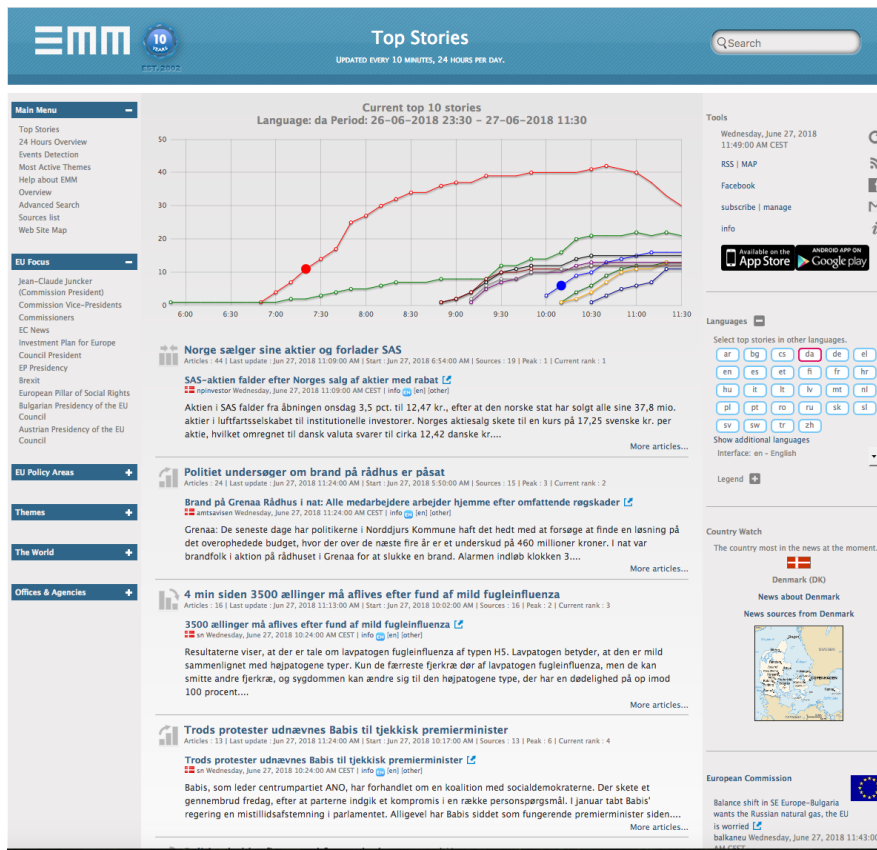
Current top 10 stories  
Language: en Period: Jan 30, 2011 1:50 AM - Jan 30, 2011 1:50 PM



Breaking News

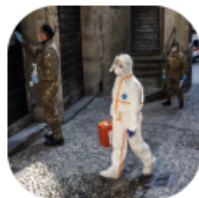


# NewsBrief AUTOMATED CONTENT MANAGEMENT





## Top Stories





### 6.25am Coronavirus likely of animal origin, not made in a lab: WHO



Articles: 85 | Last update: 2 minutes ago | Start: Apr 21, 2020 12:31 PM | Sources: 0 | Peak: New | Current rank: 1

6.25am Coronavirus likely of animal origin, not made in a lab: WHO

 theage Apr 21, 2020 11:05 PM | Info 

The World Health Organisation says that all available evidence suggests the novel coronavirus originated in animals in China late last year and was not manipulated or produced in a laboratory. US President Donald Trump said last week that his government was trying to determine whether the virus....



### More deaths, no benefit from malaria drug in VA virus study



Articles: 40 | Last update: 2 minutes ago | Start: Apr 21, 2020 6:16 PM | Sources: 0 | Peak: 2 | Current rank: 2

More deaths, no benefit from malaria drug in VA virus study

## Articles (570)



« 1 2 3 ... 55 56 57 »

## EMM News sources by country



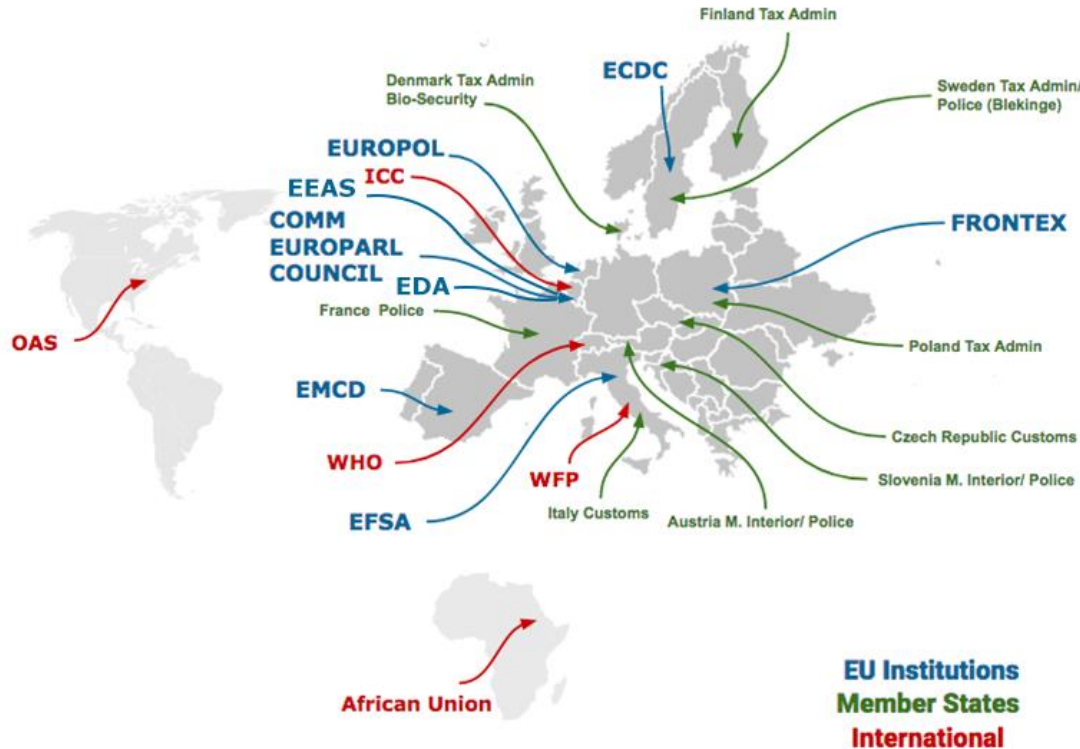
Map of sources of 570 articles



# NEWSLETTERS/REPORTS/PRESS REVIEW



# SOME EXISTING CLIENTS



**EU Institutions,  
EU Agencies,  
UN Agencies,  
African Union,  
Organization of  
American States,  
Member States  
Tax Agencies and  
Customs**

from Member States have  
already started working with  
EMM and OSINT tools.

# **Inside the EMM processing chain**

## **One example:**

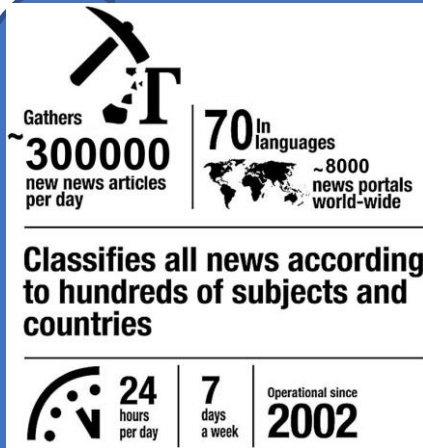
# **Named Entity Recognition**

# **in a highly Multilingual Environment**

**Text and Data Mining Unit**  
**European Commission - Joint Research Centre**

# EMM Architecture

## Custom Domain (Sources and Topics)



**Gathers**  
~300000  
new news articles  
per day

**70** in  
languages  
~8000  
news portals  
world-wide

**Classifies all news according  
to hundreds of subjects and  
countries**

**24**  
hours  
per day

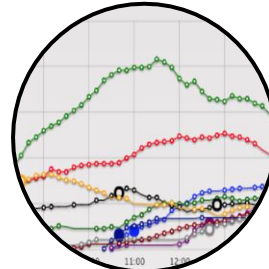
**7**  
days  
a week

Operational since  
**2002**

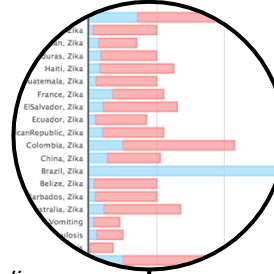
## EMM Open Source Monitoring Engine



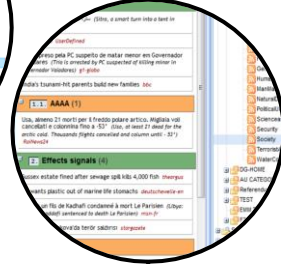
Newsbrief



Medisys



NewsDesk



Other  
Uses



## End Use Applications



European Commission

Content shown for the following countries:  
AT, BE, BG, HR, CY, CZ, DK, EE, FI, FR, DE, GR, HU, IE, IT, LV, LT, LU, MT, NL, PL, PT, RO, SK, SI, ES, SE, GB, USContent shown for the following languages:  
cs, da, de, en, es, fr, it, lt, lv, nl, pl, pt, sv, ar, fa (2 more)

Home

Add a new channel...

Refinement filters...

Get Newsletter

Source list

Hide similar items

EU Press Releases

Press Agencies

Main Sources

Online Sources

&gt;&gt; Guillaume &gt; Anti-Fraud (OLAF) &gt;



pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28



### 5 pounds of pot and nearly \$30K in cash seized in North Jersey town, cops say

Secaucus police arrested, from left, Arquemedes Robles, Marcia Turk, and Robert Stafford on Saturday morning on a variety of drug and money laundering charges. ( -- Police seized five pounds of marijuana and nearly \$30,000 in cash when they arrested a trio of out-of-townners Saturday morning outside a local hotel, authorities said.

Trigger words: money laundering

nj \_ONLINE SOURCES lang:en pub: 29 Jan 2018, 18:37



### Honduras: Hernandez inauguration 'austere', security tight

Honduran President Juan Orlando Hernandez accepted the blue-and-white presidential sash from Congress president Mauricio Oliva in a low-key inauguration ceremony held in the National Stadium in Tegucigalpa on 27 January, becoming the first person under the 1982 Honduran constitution to be sworn-in for a second term, Honduran media reported.

Trigger words: anti-corruption, Corruption, fraud, corruption(FightagainstFraud)anti-corruption, Corruption, fraud, corruption, (EU)EU, European Union

bbcmonitoring \_ONLINE SOURCES lang:en pub: 29 Jan 2018, 18:35



### "Corruzione apocalisse della democrazia". Papa Francesco parla a noi

Inizia la campagna elettorale e il tema di corruzione e mafia, torna prepotentemente all'attenzione della Chiesa per iniziativa dello stesso Papa Francesco, proprio mentre sembra essere scivolato in secondo piano nel dibattito politico. Giovedì 1 febbraio, presso il Dicastero vaticano per il...

Trigger words: corruzione, Corruzione

huffingtonpost-italia \_ONLINE SOURCES lang:it pub: 29 Jan 2018, 18:34



### Bulgarian prosecution investigates owners of luxury property

The prosecution in Bulgaria is investigating owners of luxury properties and expensive vehicles purchased in 2015 and 2016. According to the prosecutor's office in Sofia, this measure is being taken in order to find out where and how the citizens in question had acquired the funds.

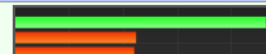
#### GEO location

Map Satellite



396 336 3731 267 8

#### Attributes

EU  
ANTI-FRAUD  
FC News

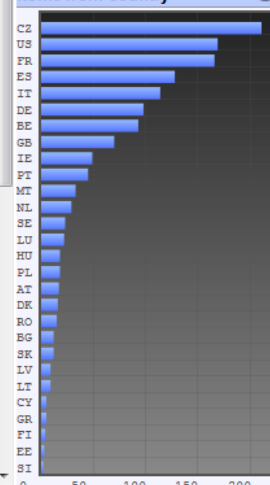
#### Channels:

- EU Anti-Fraud (OLAF) Policy
- Main sources: Anti-Fraud (OLAF)
- Fight against Fraud
- OLAF
- Food Fraud

Add new channel

#### Sources

#### Items from country







European Commission

Content shown for the following countries:  
AT, BE, BG, HR, CY, CZ, DK, EE, FI, FR, DE, GR, HU, IE, IT, LV, LT, LU, MT, NL, PL, PT, RO, SK, SI, ES, SE, GB, USContent shown for the following languages:  
cs, da, de, en, es, fr, it, lt, lv, nl, pl, pt, sv, ar, fa (2 more)

Home

Add a new channel...

Refinement filters...

Get Newsletter

Source list

Hide similar items

EU Press Releases

Press Agencies

Main Sources

Online Sources

&gt;&gt; Guillaume &gt; Anti-Fraud (OLAF) &gt;



pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28



### 5 pounds of pot and nearly \$30K in cash seized in North Jersey town, cops say

Secaucus police arrested, from left, Arquimedes Robles, Marcia Turk, and Robert Stafford on Saturday morning on a variety of drug and money laundering charges. ( -- Police seized five pounds of marijuana and nearly \$30,000 in cash when they arrested a trio of out-of-townners Saturday morning outside a local hotel, authorities said.

Trigger words: money laundering

nj \_ONLINE SOURCES lang:en pub: 29 Jan 2018, 18:37

7 4 2

### Honduras: Hernandez inauguration 'austere', security tight

Honduran President Juan Orlando Hernandez accepted the blue-and-white presidential sash from Congress president Mauricio Oliva in a low-key inauguration ceremony held in the National Stadium in Tegucigalpa on 27 January, becoming the first person under the 1982 Honduran constitution to be sworn-in for a second term, Honduran media reported.

Trigger words: anti-corruption, Corruption, fraud, corruption(FightagainstFraud)anti-corruption, Corruption, fraud, corruption, (EU)EU, European Union

bbcmonitoring \_ONLINE SOURCES lang:en pub: 29 Jan 2018, 18:35

9 6 9

### "Corruzione apocalisse della democrazia". Papa Francesco parla a noi

Inizia la campagna elettorale e il tema di corruzione e mafia, torna prepotentemente all'attenzione della Chiesa per iniziativa dello stesso Papa Francesco, proprio mentre sembra essere scivolato in secondo piano nel dibattito politico. Giovedì 1 febbraio, presso il Dicastero vaticano per il...

Trigger words: corruzione, Corruzione

huffingtonpost-italia \_ONLINE SOURCES lang:it pub: 29 Jan 2018, 18:34

3 3 3

### Bulgarian prosecution investigates owners of luxury property

The prosecution in Bulgaria is investigating owners of luxury properties and expensive vehicles purchased in 2015 and 2016. According to the prosecutor's office in Sofia, this measure is being taken in order to find out where and how the citizens in question had acquired the funds.

#### Entity analysis

- EU
- Andrej Babiš
- European Commission
- Jaroslava Faltýnka
- European Parliament
- Anti Fraud Office
- Liviu Dragnea
- Alena Schillerová
- Front National
- Volkswagen
- Marine Le Pen
- Jean Marie Le Pen
- Giovanni Kessler
- Miroslav Kalousek
- Miloš Zeman
- European Investment Bank
- John Dalli
- Court of Auditors
- Fondurilur Europene
- Drum SA

396 336 3731 267 8

#### Attributes

EU  
ANTI-FRAUD  
FC News

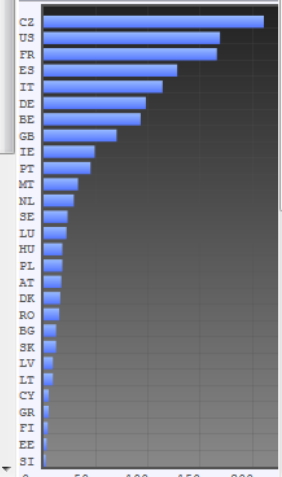
#### Channels:

- EU Anti-Fraud (OLAF) Policy
- Main sources: Anti-Fraud (OLAF)
- Fight against Fraud
- OLAF
- Food Fraud

Add new channel

#### Sources

#### Items from country



# Multilingual name recognition and variant mapping

it      contro il regime del presidente **Bashar al Assad** in Siria.



# Multilingual name recognition and variant mapping

es protestas pacíficas contra el presidente **Bashar el Asad**.

fr la Lybie conquise pour la Syrie de **Bachar Al-Assad**.

de syrischen Führung unter **Präsident Baschar al-Assad** sei.

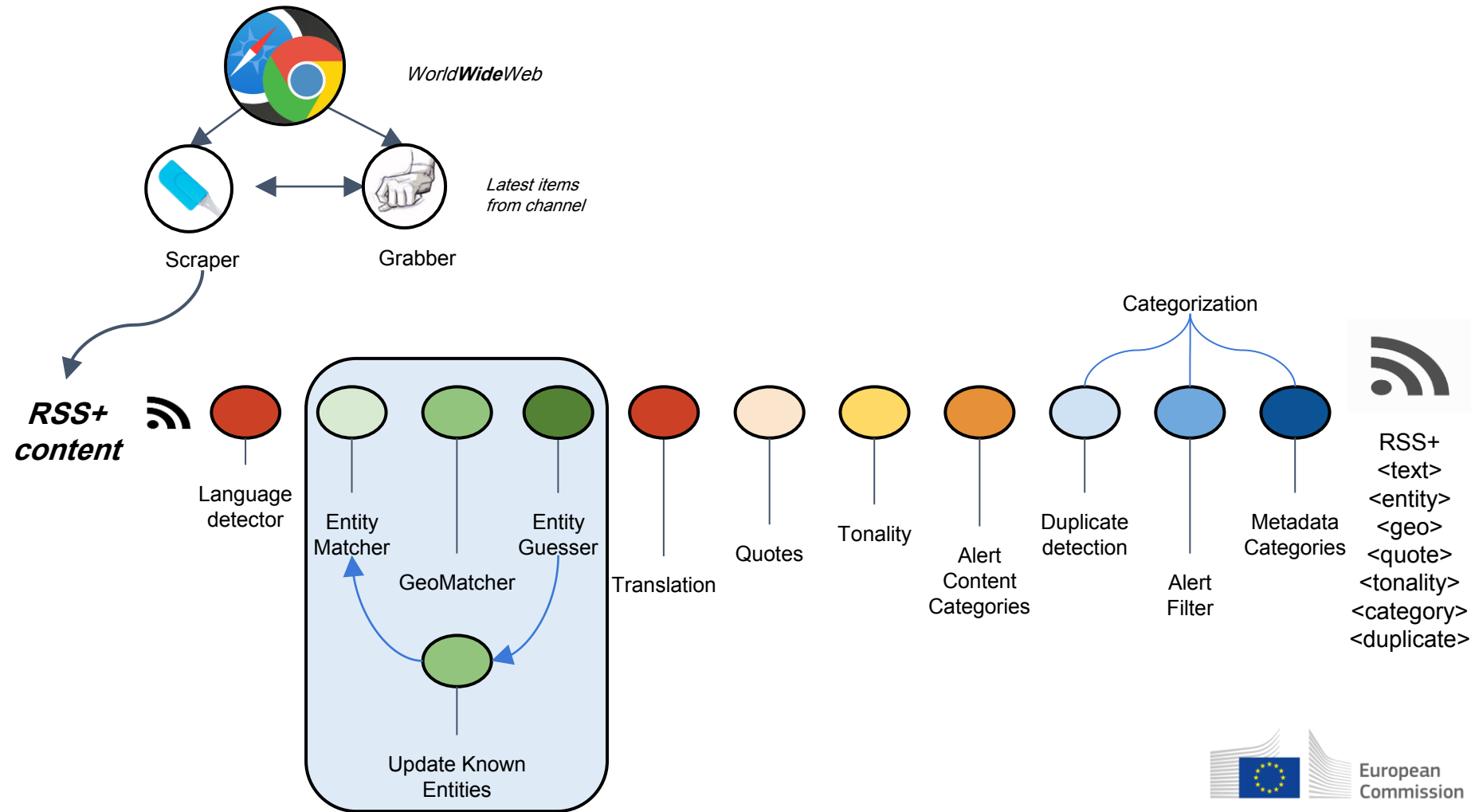
it contro il regime del presidente **Bashar al Assad** in Siria.

cs trvající snahu o zničení režimu **Baššara Asada**

hu sok felkelő csatlakozott a szíriai **Bassár el-Aszad** elnök

ru чтобы тот уничтожил **Башара Асада** в Сирии

ar احتجاجات سلمية على حكم الرئيس **بشار الأسد**.



# NER for EMM

- Hybrid approach combining Rule based and Machine Learning components
  - Highly multilingual: 70 languages
  - Specific lexical resources for 21 languages
  - Around 200 rules including 30 language specific rules  
e.g. : "<function> + (punct) + <first name> + UCpattern<1,2> + notUCpattern"

# NER for EMM

## ■ Person Name guesser evaluation

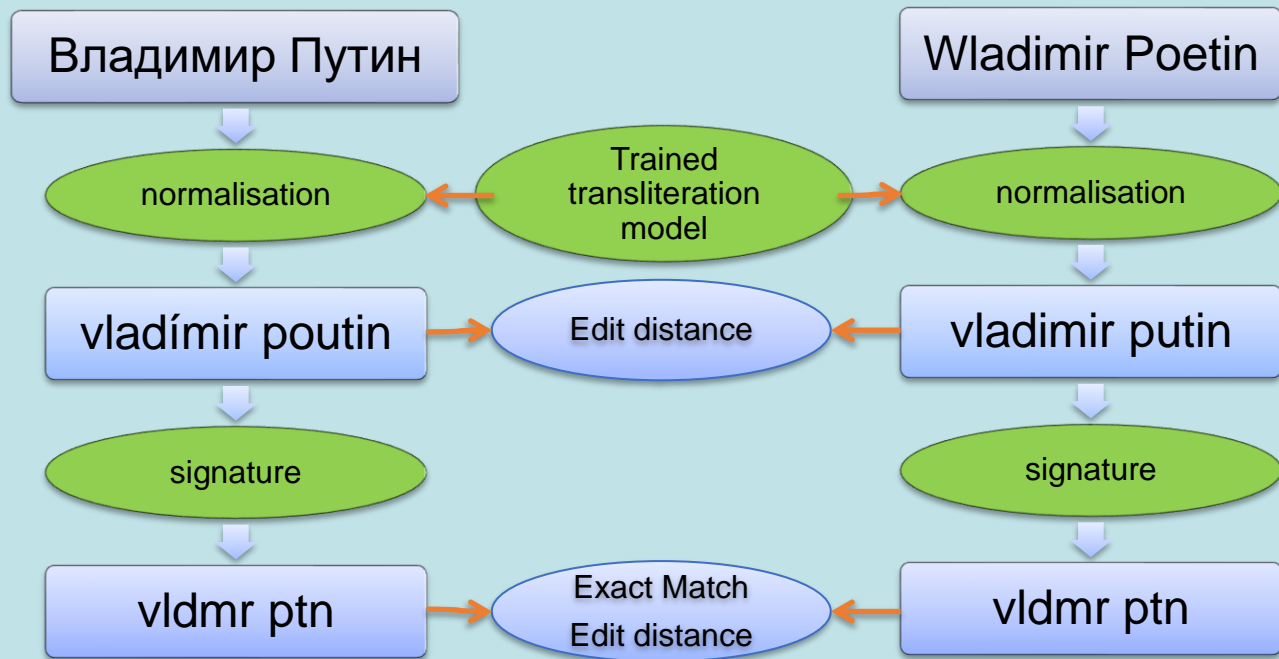
language code	Reference entities	Entities found by the system	Correct entities	Wrong entities	Forgotten entities	Precision	Recall	F-measure
de	2770	1613	1487	126	1242	92.19%	53.68%	67.85%
nl	4716	2344	2105	239	2468	89.80%	44.64%	59.63%
es	4252	2582	2395	187	1824	92.76%	56.33%	70.09%
en	6560	3815	3694	121	2843	96.83%	56.31%	71.21%
ro	336	226	205	21	128	90.71%	61.01%	72.95%
hu	678	589	501	88	163	85.06%	73.89%	79.08%
it	4545	1954	1789	165	2698	91.56%	39.36%	55.05%
hu	3746	1975	1779	196	1885	90.08%	47.49%	62.19%
pt	740	458	382	76	337	83.41%	51.62%	63.77%
tr	388	140	123	17	262	87.86%	31.70%	46.59%
Ttl	28731	15696	14460	1236	13850	92.13%	50.33%	65.10%

# Update known entities

- Guessed entities are used to enrich and keep up-to-date the known entity resource
- 600 unknown names per day
- 120 are automatically merged with known entities

# Update known entities

- 600 new names from the guesser are compared daily with ~1,500,000 known name variants



# Geomatcher

- Matching

Milan Komar went to Venice during his visit in Italy

- Disambiguation

~~Milan:Lombardia:Italy cls:2~~ Entity

Venice:Veneto:Italy cls:-3

Venice:Venice:Veneto:Italy cls:2

The most specific wins!

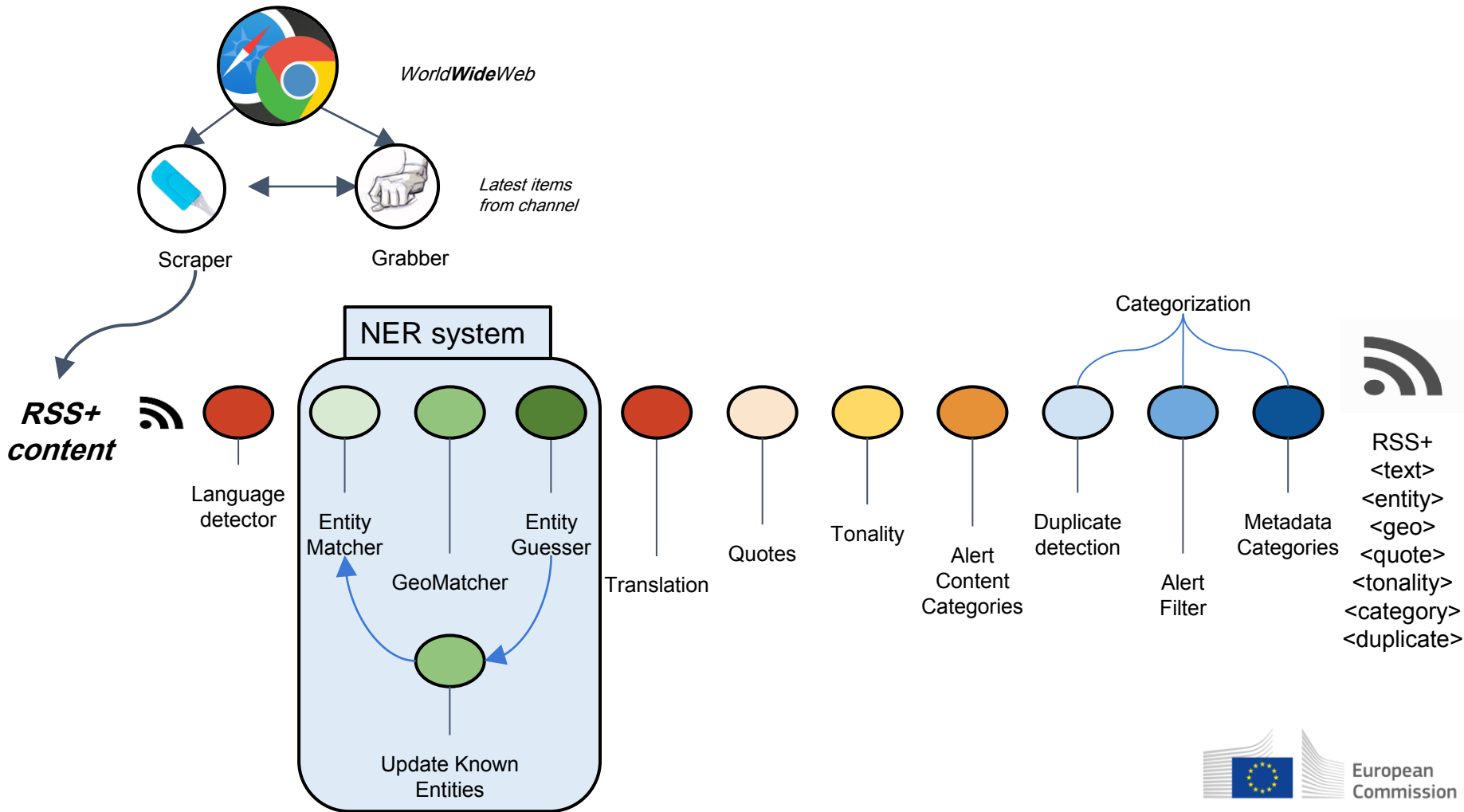
~~Venice:Madison:Illinois:USA cls:5~~

Nothing USA related

Italy cls: -1

Result:

georss: Venice:Venice:Veneto:Italy fullgeo: Italy







# Any questions?

EMM support: [\*\*JRC-EMM-SUPPORT@ec.europa.eu\*\*](mailto:JRC-EMM-SUPPORT@ec.europa.eu)

## Top Stories





### 6.25am Coronavirus likely of animal origin, not made in a lab: WHO



Articles: 85 | Last update: 2 minutes ago | Start: Apr 21, 2020 12:31 PM | Sources: 0 | Peak: New | Current rank: 1

6.25am Coronavirus likely of animal origin, not made in a lab: WHO

 theage Apr 21, 2020 11:05 PM | Info 

The World Health Organisation says that all available evidence suggests the novel coronavirus originated in animals in China late last year and was not manipulated or produced in a laboratory. US President Donald Trump said last week that his government was trying to determine whether the virus....



### More deaths, no benefit from malaria drug in VA virus study



Articles: 40 | Last update: 2 minutes ago | Start: Apr 21, 2020 6:16 PM | Sources: 0 | Peak: 2 | Current rank: 2

More deaths, no benefit from malaria drug in VA virus study

## Articles (570)



« 1 2 3 ... 55 56 57 »

## EMM News sources by country



Map of sources of 570 articles

