Welcome to the THESEUS website
Project supported by Hercule III

THESEUS is designed and maintained by JRC multiannual work programmes.
Public area publishes information about our aim and work overview, events, and info on frequently asked questions.
Restricted area publishes results obtained from work conducted as a part of the multiannual work programme 2015/2020, its preceding work programmes and the JRC project AMT (Automated Monitoring Tool) on External Trade.
THESEUS is financed by the HERCULE III anti-fraud programme of the European Union; the AMT project series is financed by OLAF and carried out in collaboration between OLAF and JRC.
OLAF is the primary customer of work published for the website.
Objective: to test further the concept and usefulness of the Europe Media Monitor (EMM) tool in monitoring open source news and identifying information relevant for customs fraud. The first test category on "illicit tobacco" was created in August 2018.

Activities:
- Training in EMM
- Verify and improve the "customs relevant" sources used
- Improve and extend the "illicit tobacco" query
- Create newsletter
- Analyse "meta-data"
- Look at Darkweb potential
EMM monitors
• ~320,000 news articles/day (open source)
• 24/7
• from ~10,000 sources

EMM categorises news according to more than 9,500 definitions and meta-definitions (6800 alerts, 2700 filters)
• Alerts based on keyword combinations
• Filters combine alerts with alerts, countries, sources, languages

Developed by JRC I.3
Consolidated project (started in 2002)
Reminder – how EMM works

World Wide Web

Scraper
Grabber

Latest items from channel

Initial RSS:
- <source>
- <title>
- <text>

Language detector
NEROne
Translation
Quotes
Sentiment
Keyword-based Categories
Duplicate detection
Metadata-based Classification

RSS+:
- <source>
- <title>
- <text>
- <entity>
- <geo>
- <quote>
- <tonality>
- <category>
- <duplicate>
- <tonality>

Event Metadata Extraction

Clusters
Cross-lingual linking
Time linking

Index

Event Metadata Extraction

Clusters
Cross-lingual linking
Time linking

Index
Illicit Tobacco Trade – EMM training

- 1 training session on EMM tools with OLAF and MS (20-21/03/2019)
- 1 dedicated NewsDesk group called "INTEL4CUSTAF" with 11 users from OLAF, JRC and MS (DK, CZ, ES, FR)
Illicit Tobacco Trade – EMM sources & queries

- 350 new **sources** added to EMM from May 2019
  - from 11 countries
  - trying to identify "customs relevant" sources

- 7 **categories** (topics) created and/or extended by EMM and OLAF
  - combination of new keywords
  - improved languages covered

- 2 new **categories** created by Ministry of Finance of the Czech Republic (about VAT and Tobacco Offer) in April 2019

- 3 **categories** EMM created for OLAF before 2019 about Tobacco Smuggling
Illicit Tobacco Trade (EMM) – where to find us

Personalise the results and get informed
Illicit Tobacco Trade mentioning EU Countries
Searching EMM (2)

Illicit Tobacco Trade reporting from EU Countries

Two men arrested by detectives investigating criminality linked to INLA.

(Taranto, in case 16kg of cigarette contraband: arrested) (Tarento, in home 16kg of contraband cigarettes: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested) (Taranto, in case 16 kg of contraband: arrested)

(Taranto, in case 16kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Taranto, in case 16 kg of cigarette contraband: arrested)

(Tarant
Illicit Tobacco Trade reporting from EU Countries mentioning "water pipe"
Illicit Tobacco Trade (EMM) - statistics & trend analysis

Illicit Tobacco Trade, EU vs. US
Reporting from EU and US compared, normalized values, last 3 months of data
Illicit Tobacco Trade (EMM) - statistics & trend analysis

Comparison of reporting in different languages
Comparing the reporting for 4 categories, last 30 days, normalized values
Illicit Tobacco Trade (EMM) - statistics & trend analysis

Reporting about Illicit Tobacco Trade
Countries breakdown, last 30 days
Illicit Tobacco Trade (EMM) - statistics & trend analysis

Illicit Tobacco Trade
Reporting from selected countries, 1 month data, normalized values
Illicit Tobacco Trade (EMM)  
User feedback

Negatives:
- Effective use requires expertise – NOT easy to "dip in and out"
- Press articles only informative in character and generally about the past
- Is the information from open source internet really true?

Positives:
+ EMM can provide a general worldwide overview of interesting criminal/fraudulent cases
+ It could be used to monitor:
  • legislative developments across countries (useful for Mutual Administrative Assistance)
  • court decisions in anti-fraud matters
  • other interesting issues
+ EMM could help to establish strategical customs risk analysis more effectively

PROPOSAL: create monthly reports on Customs Risks from the central point JRC/OLAF
Objective: to do text mining and categorization of web sites in the Dark Web using EMM technology

Initial list of websites (mostly e-commerce) provided by Danish Customs Agency about Illicit Tobacco Trade

JRC ran an initial feasibility test
The Dark Web Experiment

Positive outcomes:
- It is feasible to scrape the dark web (with some limitations)
- EMM can scrape the text from an onion website and analyse it

Drawbacks:
- EMM performs better on news websites rather than e-commerce
- Dark Web is very slow
- JRC prefers to avoid being traced (and blocked)
- A user account is always needed
- Onion URLs change regularly
- All e-commerce onion sites require a CAPTCHA
The Dark Web Experiment

- What we **cannot** do (currently):
  - EMM cannot scrape the DarkWeb (especially e-commerce onion sites) with the same frequency of traditional websites
  - EMM cannot register accounts to all monitored onion sites
  - EMM cannot cope with the constant updates of the onion URLs

- What we **could** do in future:
  - Connect to a list of onion websites (mostly with text or news)
  - Take a "snapshot" of each page, analyse it and show the result of the analysis/categorization to help users understand better where to investigate more
  - Look for different solutions to access the Dark Web (e.g. Tor2Web)
The Dark Web Experiment: the initial set
The Dark Web Experiment: another example
The Dark Web Experiment: another example
Conclusions

+ The pilot has been fruitful with positive outcomes and ideas for the future
+ Demonstrates EMM can identify relevant information across EU
+ Community role is vital to improving coverage and searches
+ Can provide valid complementary information/support for customs to monitor situation (early warning, situation awareness)

- EMM gives complementary information, with little grounds for new investigation
- EMM tool for category definition requires expertise/frequent use to make it easy to use
The way forward: personalized dashboard
Thanks for your attention!

Questions/comments?