
CENTRIC
Centre of Excellence in Terrorism,
Resilience, Intelligence and
Organised Crime Research

Retrieval and Analysis of Heterogeneous Online Content for Terrorist Activity Recognition

Babak Akhgar

What is?

- Retrieval and Analysis of Heterogeneous Online Content for Terrorist Activity Recognition
- EU H2020 Funded project
- 3 Year project commenced in Sept, 2016

Hochschule für den öffentlichen Dienst in Bayern

KEMEA

NCA National Crime Agency

ibz

Police & Crime Commissioner West Yorkshire

Generalitat de Catalunya Departament d'Interior

ECS European Organisation for Security

Linguatec Language Technologies

FINMECCANICA

Cybercrime Research Institute GmbH

upf. Universitat Pompeu Fabra Barcelona

rinicom secure communications

CERTH Centre for Research & Technology Hellas

THALES

CENTRIC

Project Focus

Provide significant key operational capabilities for the ever increasing technically challenging terrorism domain

CENTRIC

Common Characteristics LEAs ICT Requirements

- Large Datasets (3Vs principal - large volume of data, large variety of data and high velocity of data)
- Integration of traditional Intel with OSINT
- Data Visualisation
- Enhance Decision Making Capability (speed and impact)
- Predictive Analytics
- Situational Awareness
- Advance ICT Training
- And most importantly, developing relevant knowledge and actionable intelligence

CENTRIC

Situational Awareness in security and policing context is the capability to identify (people, events, materials, locations...), contextualise, visualise, process and comprehend the critical elements of intelligence about particular areas of concern. Areas of concern can be anything from Cyber investigation to the management of a major crisis.

Akhgar and Bayerl (2015)

CENTRIC

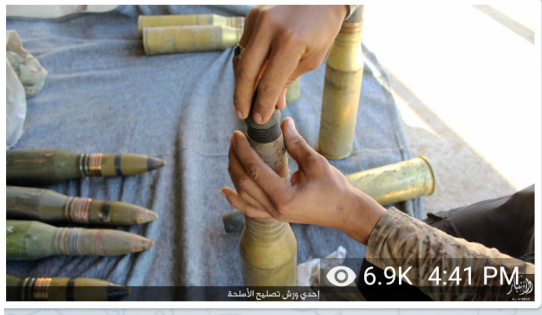
Core requirements for Situational Awareness Platforms

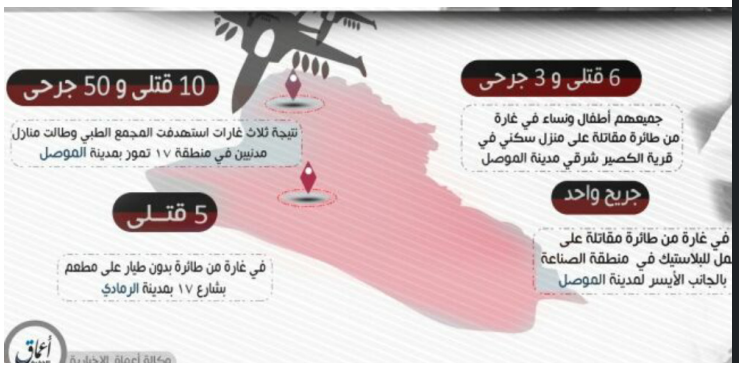
- Platforms should be active, not only during particular crises or incidents, but also provide monitoring capability for **preventative measures**
- Run real-time to near **real-time** analytical capability
- Provision for **pushing and pulling intelligence** rather than information
- Multimodal information visualisation
- Capability to manipulate and corroborate information from **various sources** (e.g. Social media to dark web)
- Provision for integration with OSINT
- Taxonomy-driven and semantic navigation
- Reliability indicators
- **Dynamic and multimodal search and traceability** of intelligent to its source
- Verification functionality
- Legal and ethical constrain - GDPR (from chain of evidence to presentation in court of law)

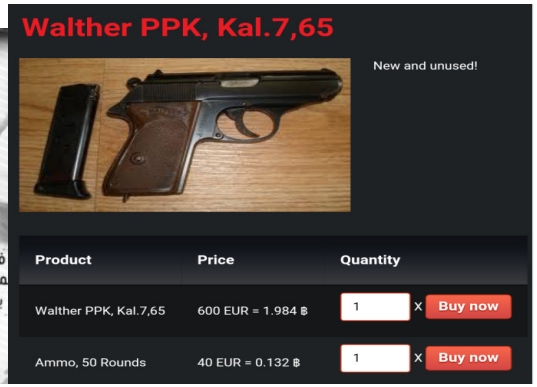
CENTRIC

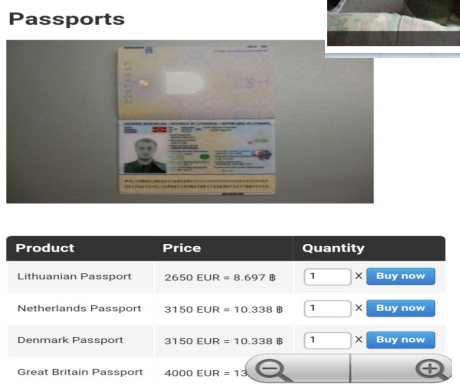
TENSOR proposition

- Bring significant advantages to LEA operational capabilities through a dynamic intelligence platform
- Detect, classify and report terrorist content
- Enhance reliable planning and prevention functionalities for the early detection
- Automate the workload, clarifying critical information from noise
- Analysis and classification of text, image, video, and audio documents.









CENTRIC

Core functionality

- Preparation
  - *Taxonomies, Ontologies and Indicators, Entry Points and Use Cases*
- Acquisition and Mark Up:
  - *Surface, deep and dark web crawling and target searching*
  - *Autonomous dialogue enabled bots - interact via forums*
  - *Concept extraction from text and multimedia*
- Analytics
  - *Further concept extraction, classification, clustering, SNA, credibility, priority and multimedia forensics*
- Summarization and Presentation
  - *Multilingual summaries, secure storage, audited access, visual analytics, dashboard, translation*

CENTRIC

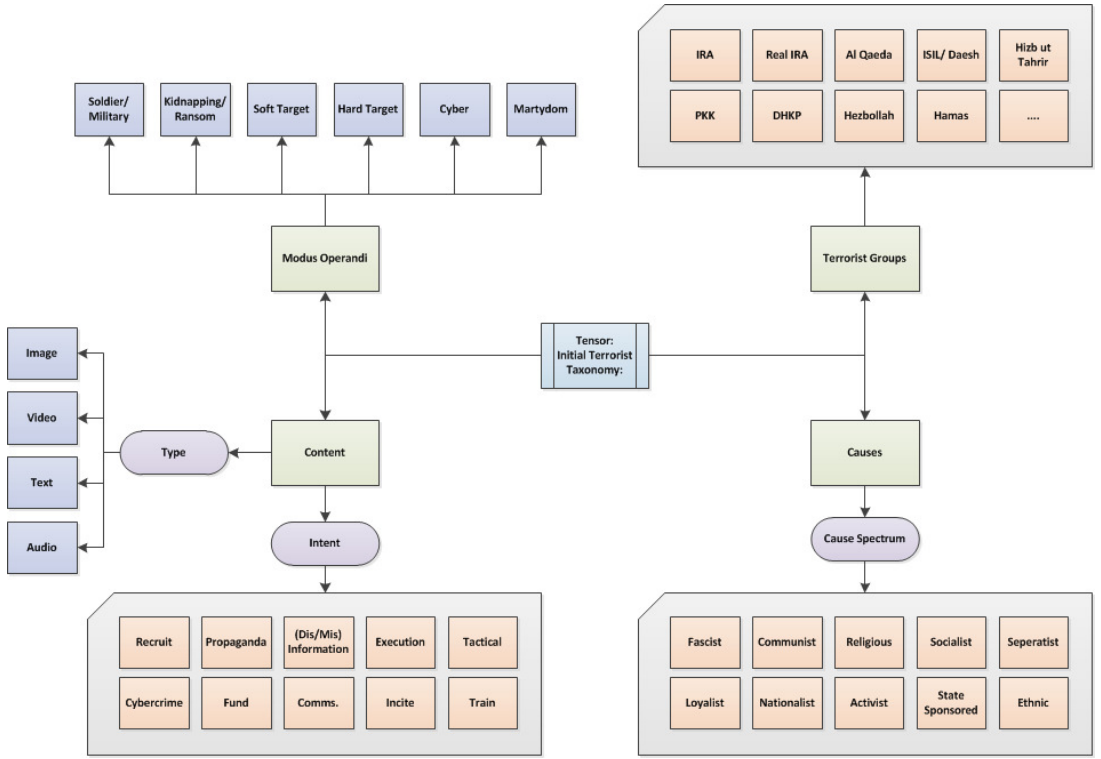
TENSOR components

- Identification of misinformation and disinformation in relation to terrorist generated content
- Web crawling and scraping across the surface, deep and dark webs
- Concept extraction using the taxonomy and ontology
- Aggregation based on formal concept analysis
- Multimedia processing components
- Data store and content repository
- Legal guidelines

CENTRIC

Categorisation

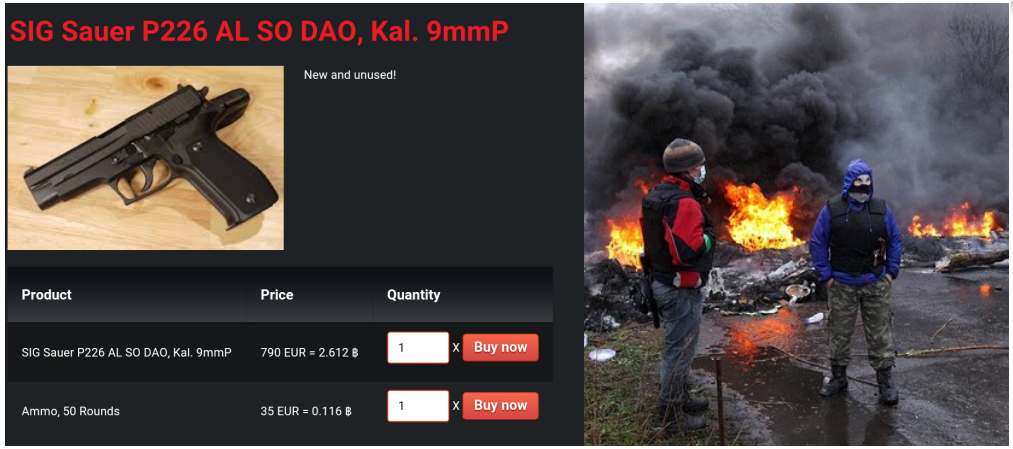
- Taxonomy
  - *building, registering and classifying over 35 active international terrorist groups including:*
- ideology,
- sub-defining characteristics,
- regions of operation,
- background information
- styles of operation/ action.

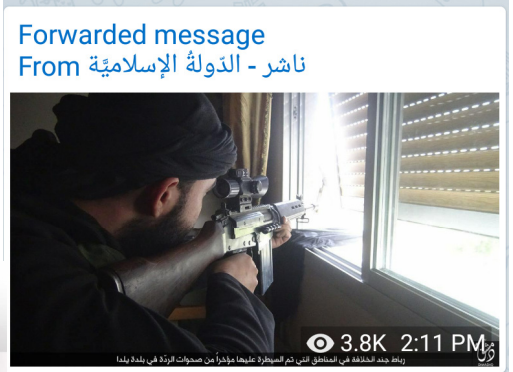


CENTRIC

Identifying Terrorist Communities

- Intelligent bots capable of interacting in forums and online pages of interest, harvesting dialogue.
- Social media networks analysis to identify connections between vulnerable and malicious users. (recruitment, funding, indoctrination)
- Analysis of metadata, multimodal and multilingual content from the surface and dark web. (speech and image recognition)





CENTRIC

TWITTER COMMUNITIES

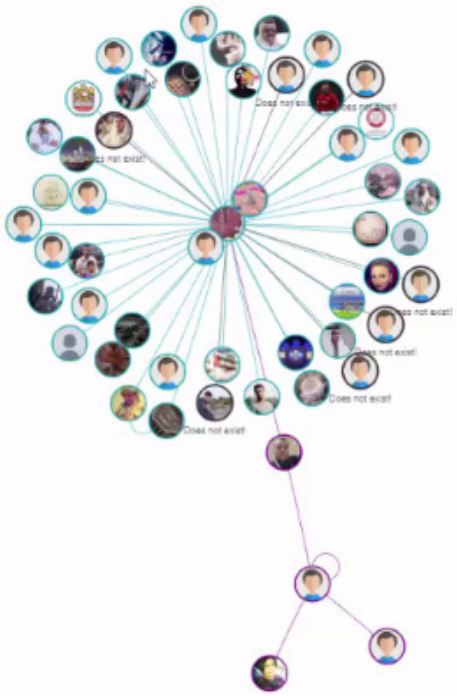
From 16/08/2017 To 19/08/2017 Get Communities

STATISTICS

466 Total Users
124 Suspended Users
30 Non-existent Accounts
408 Twitter Pairs
87 Components
90 Communities

CONNECTED COMPONENTS

#1: 76 users, 1 community (C1)
#2: 54 users, 2 communities (C2 C
#3: 40 users, 1 community (C3)
#4: 29 users, 2 communities (C5 C
#5: 19 users, 2 communities (C4 C
#6: 11 users, 1 community (C6)
#7: 9 users, 1 community (C7)
#8: 8 users, 1 community (C8)
#9: 8 users, 1 community (C9)
#10: 6 users, 1 community (C12)



KEY PLAYERS

#1	Ola_Zein	Community #1 (76 users)
#2	AwdeBilal	Community #1 (76 users)
#3	BassemSalame	Community #1 (76 users)
#4	Rana_al_bassam	Community #1 (76 users)
#5	zainab_hijazy	Community #1 (76 users)
#6	malak_9f99	Community #1 (76 users)
#7	hanan_syria	Community #1 (76 users)
#8	sarah_Naser90	Community #1 (76 users)
#9	~~zainab_hijazy~~	Community #1 (76 users)
#10	Nidal_Zein11	Community #1 (76 users)

CENTRIC

OSINT Focus aspect of TENSOR

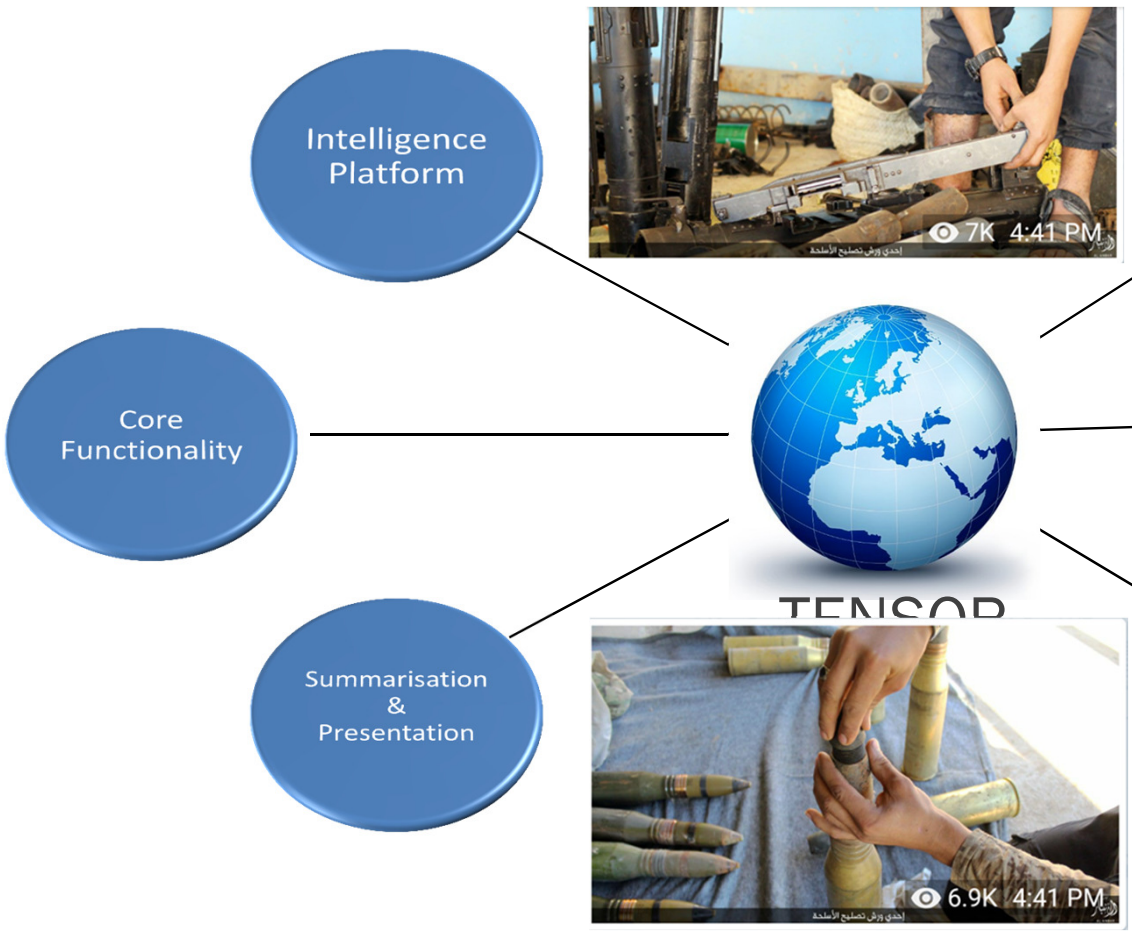
- Increasing best practice, capabilities and experience in information security and converting them to **effective practices**
- Exploration and exploitation of surface, deep , dark web, social media and GP
- Empowering investigators through technology (such as virtualisation)

OUR Project is driven by

JAPAN

- J- Action JUSTIFIED
- A- AUTHORITIES require - in place ? (e.g. DSA)
- P-Action PROPORTIONATE TO THREAT
- A-Is everything AUDITABLE ? in what form?
- N- Is what we are doing NECESSARY

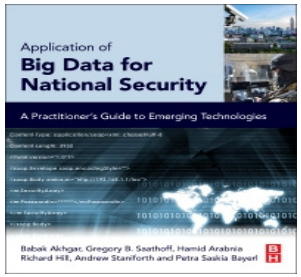
Project Objectives

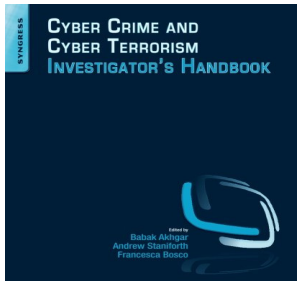


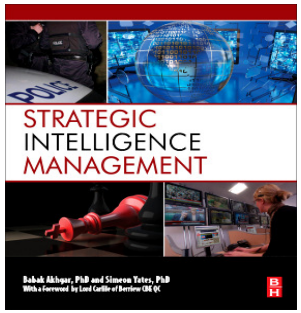
- Video
  - Extraction of subtitles
  - Iconography identification
  - Extraction of audio for processing
  - Classification of content
  - Insertion of time stamps
- Audio
  - Multi-language speech recognition
  - Conversion of audio to text for NLP
- Image
  - Image identification
  - Media forensics
  - Iconography Identification
- Natural Language & Text
  - Translation to Intelligence Officers language
  - Multilingual processing
  - Classification of sentiment / content

CENTRIC

CENTRIC







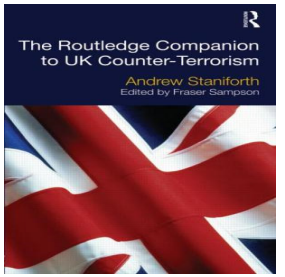
tensor-project.eu

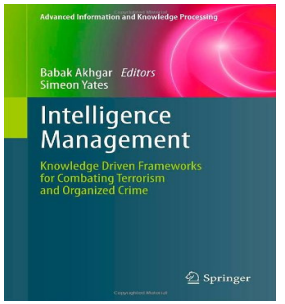
CENTRIC

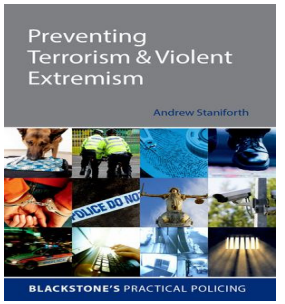
CENTRIC@shu.ac.uk

research.shu.ac.uk/centric

@CENTRICResearch







CENTRIC