



AI supporting content verification and analysis

Session: Automated Fact-Checking: The way to go?

Stuart E. Middleton

University of Southampton, Electronics and Computer Science
www.ecs.soton.ac.uk/people/sem

SciCar 2018

25th Sept 2018

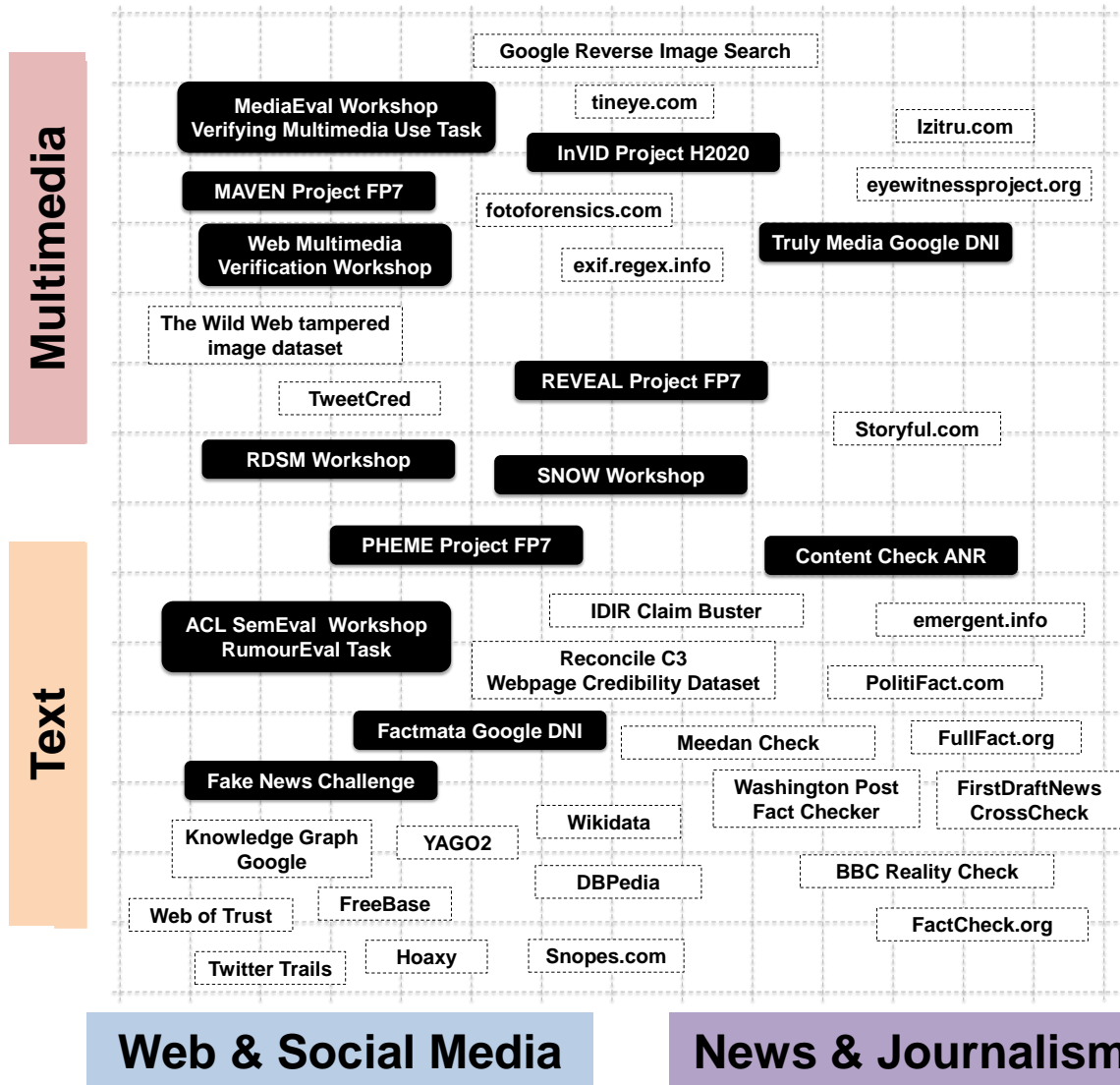
Overview

- Verification tools and research
- AI supported content verification
 - Social media and web scale analytics
 - Digital image forensics
 - Digital text forensics
- Trends around AI and content verification

Speaker

- Dr Stuart E. Middleton
 - Senior research engineer
 - University of Southampton, Electronics and Computer Science (ECS), IT Innovation Centre
- Research
 - Computational linguistics and information extraction
- Interdisciplinary
 - Journalists (Deutsche Welle, REVEAL)
 - Law enforcement agencies (UK Border Force, FloraGuard; UK National Crime Agency, VIVACE)
 - Intelligence analysts from UK DSTL and UK Ministry of Defence

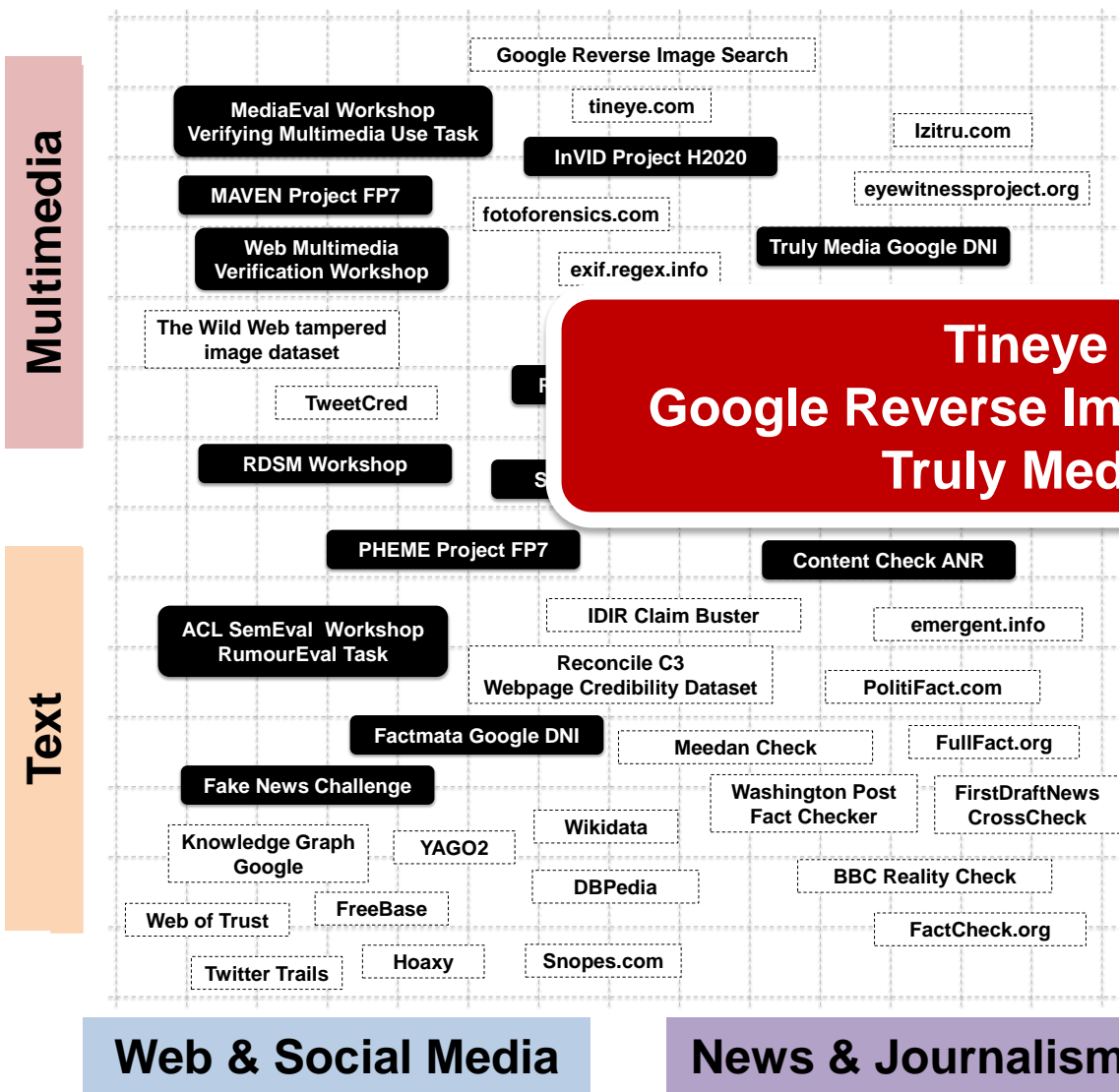
Verification tools and research



Key Datasets and Fact Checking Sites



Verification tools and research



Key
 [Dashed Box] Datasets and Fact Checking Sites
 [Black Box] Workshops and Projects



EU Referendum: The lies they tell...
eureferendum.blogspot.com



LA BOÎTE À IMAGES : LA GU...
archive.li



Adnan Hajj photographs controversy - Wikiwand
wikiwand.com



The Jawa Report: New Plane Makes Maiden V...
mypejawa.mu.nu



Lemonodor: February 2008 Archives
lemonodor.com



EU Referendum: The lies they tell...
eureferendum.blogspot.com



hgodlife Instagram followi...
piknu.com

Verification Workshop

exif.regex.info

Tineye Google Reverse Image Search Truly Media

TinEye Upload or enter Image URI

28 results
Searched over **31.2 billion images** in 18.8 seconds.
for: fig1a-IAFplaneLebanon.jpg

Best match Filter by domain/collection

www.fourandsix.com
Filename: **AnotherHajj.jpg**
Found on: photo-tampering-history/tag/media
Page crawled on Sep 02, 2015
Found on: photo-tampering-history/tag/photojour...
Page crawled on Sep 02, 2015

www.ynetnews.com
Filename: **1l_wa.jpg**
Found on: articles/0%2C7340%2CL-3287774%2C00.html
Page crawled on Nov 27, 2014
Found on: articles/0,7340,L-3287774,00.html
Page crawled on Apr 19, 2008

TRULY Demo Collection Jochen

Type URL to add...

Search for a collection

Sort by: Date added

VIEW: [Grid]

Item 1: Name: castle germany, Original Source: castle germany, GRAPHIC, Add tag, Verify

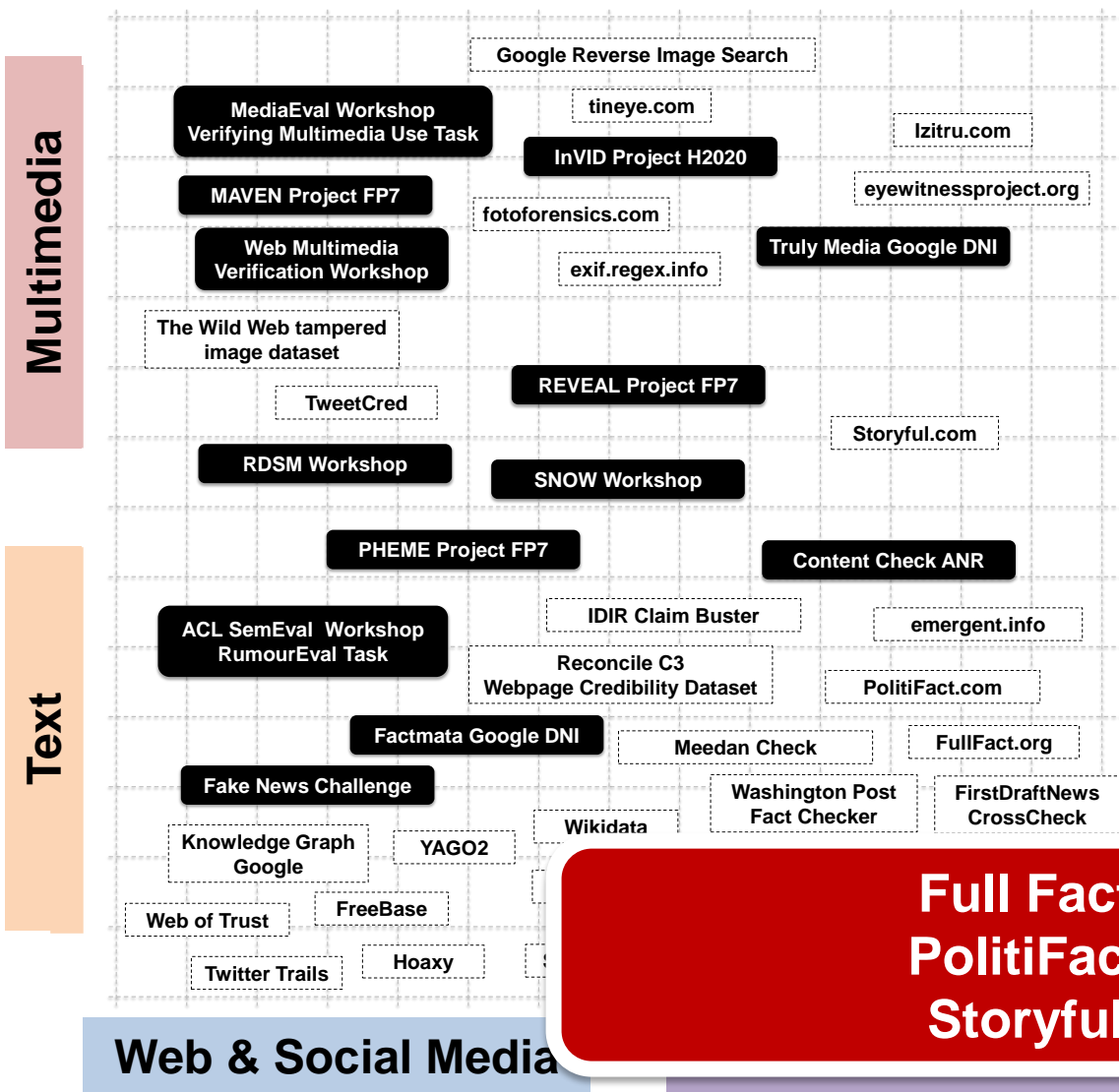
Item 2: Name: TVVenezuela Noticias, Original Source: EN VIVO: Situación en Venezuela - Consulta Popular 16 de Julio, 286.4k, 2.5k, 133, Add tag, Verify

Item 3: Name: iKcKyc, Original Source: iKcKyc, GRAPHIC, Add tag, Verify

Workshops and Projects



Verification tools and research



**Full Fact
PolitiFact
Storyful**

Key

- Datasets and Fact Checking Sites
- Workshops and Projects

Verification tools and research



Multimedia

Text

Google Reverse Image Search

MediaEv
Verifying Mul

MAVEN Pro

Web M
Verification

The Wild Web ta
image data

RDSM V

ACL SemEval
RumourE

Fake News Challenge

Home About Blog Contact Toolkit Donate THE UK'S INDEPENDENT FACTCHECKING CHARITY

Economy Europe Health Crime Education Immigration Law

How many cyclists are killed each year?

Published: 17th Aug 2018

SHARE 712 TWEET

FEATURED

Leaked 'fake news' inquiry report - our response and how you can help

Claim 1 of 3

Claim

In 2016, 18,477 cyclists were injured in road accidents, including 3,499 who were killed or seriously injured.

Conclusion

Correct for Britain. The 18,477 figure includes deaths, of which there were 102.

POLITIFACT
WINNER OF THE PULITZER PRIZE

EDITIONS x TRUTH-O-METER™ x PEOPLE x PROMISES x PANTS ON FIRE x ABOUT US x



"Puerto Rico's electric grid and their electric generating plant was dead before the storms ever hit. It was in very bad shape. It was in bankruptcy. It had no money. It was largely -- you know, it was largely closed."

— Donald Trump on Tuesday, September 11th, 2018 in press conference



Wikidata

Washington Post
Fact Checker

FirstDraftNews
CrossCheck

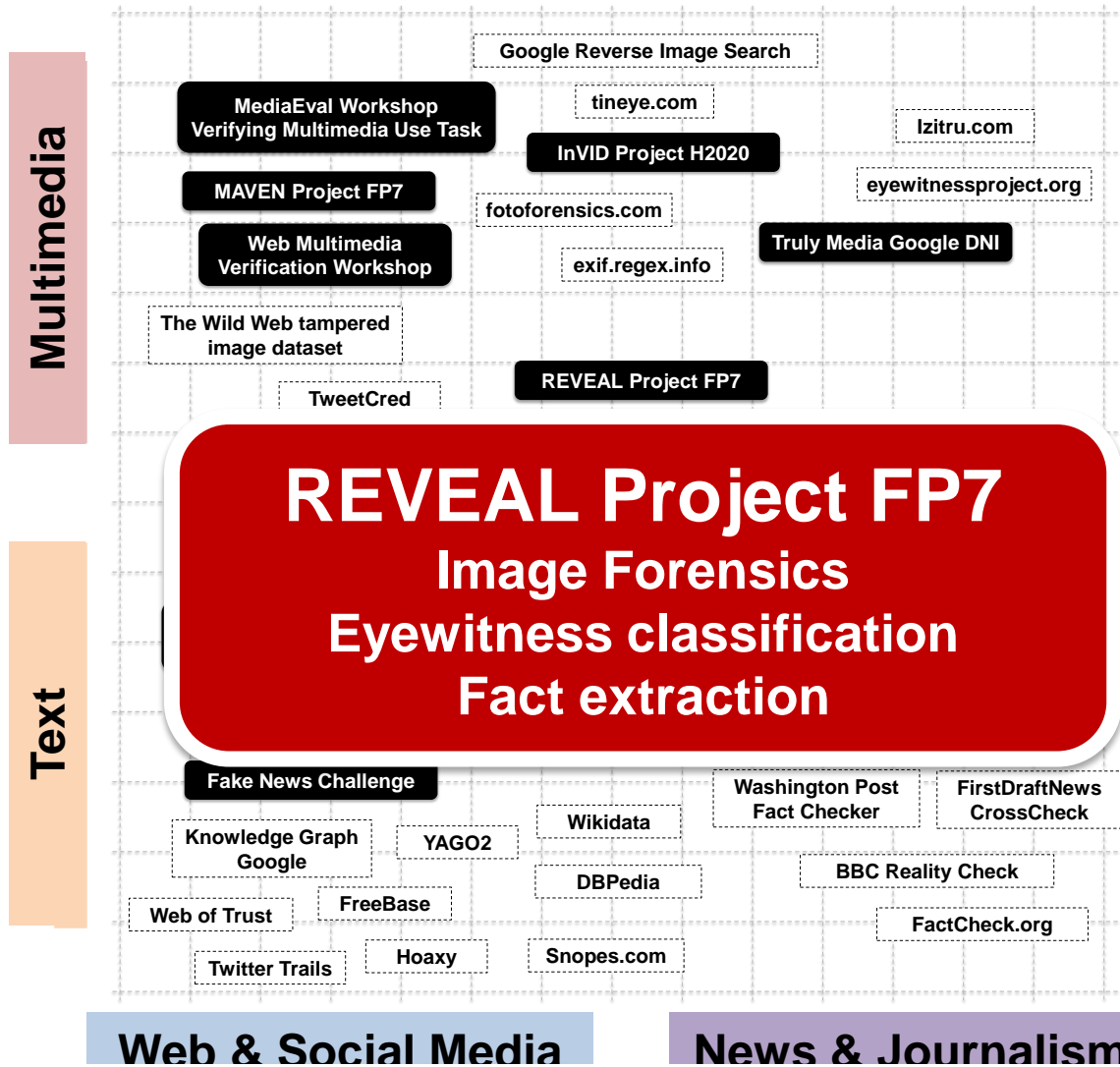
**Full Fact
PolitiFact
Storyful**

Key

Datasets and Fact Checking Sites

Workshops and Projects

AI Verification Research



C. Boididou, S. E. Middleton, Z. Jin, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, and Y. Kompatsiaris. "Verifying information with multimedia content on Twitter", Multimedia Tools and Applications, Sep. 2017, Pages 1-27, Springer 2017

Middleton, S.E. Kordopatis-Zilos, G. Papadopoulos, S. Kompatsiaris, Y. "Location Extraction from Social Media: Geoparsing, Location Disambiguation, and Geotagging", ACM Transactions on Information Systems (TOIS) 36, 4, Article 40 (June 2018)

Metadata summary

JPEG	JpegComment	JPEG	File
Compression	Baseline	Version	1.1
Type	8 bits	Resolution Units	none
Data Precision	723 pixels	X Resolution	1 dot
Image Height	550 pixels	Y Resolution	1 dot
Image Width	3	Thumbnail Width Pixels	0
Number of Components	Y component: Quantization table 0, Sampling factors 2, hortic/2 vert	Thumbnail Height Pixels	0
Component 1			

MAITA

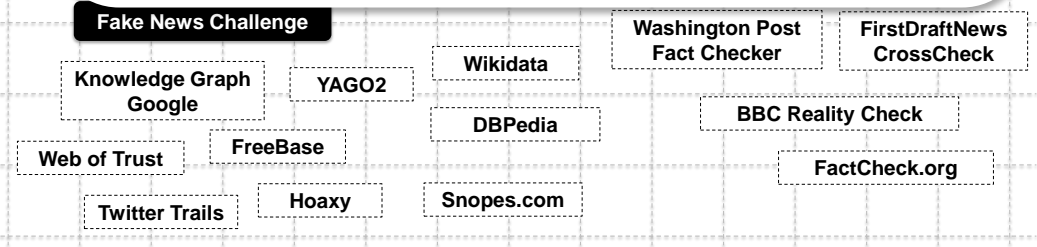
REVEAL Project FP7

Image Forensics

Eyewitness classification

Fact extraction

Text



Web & Social Media

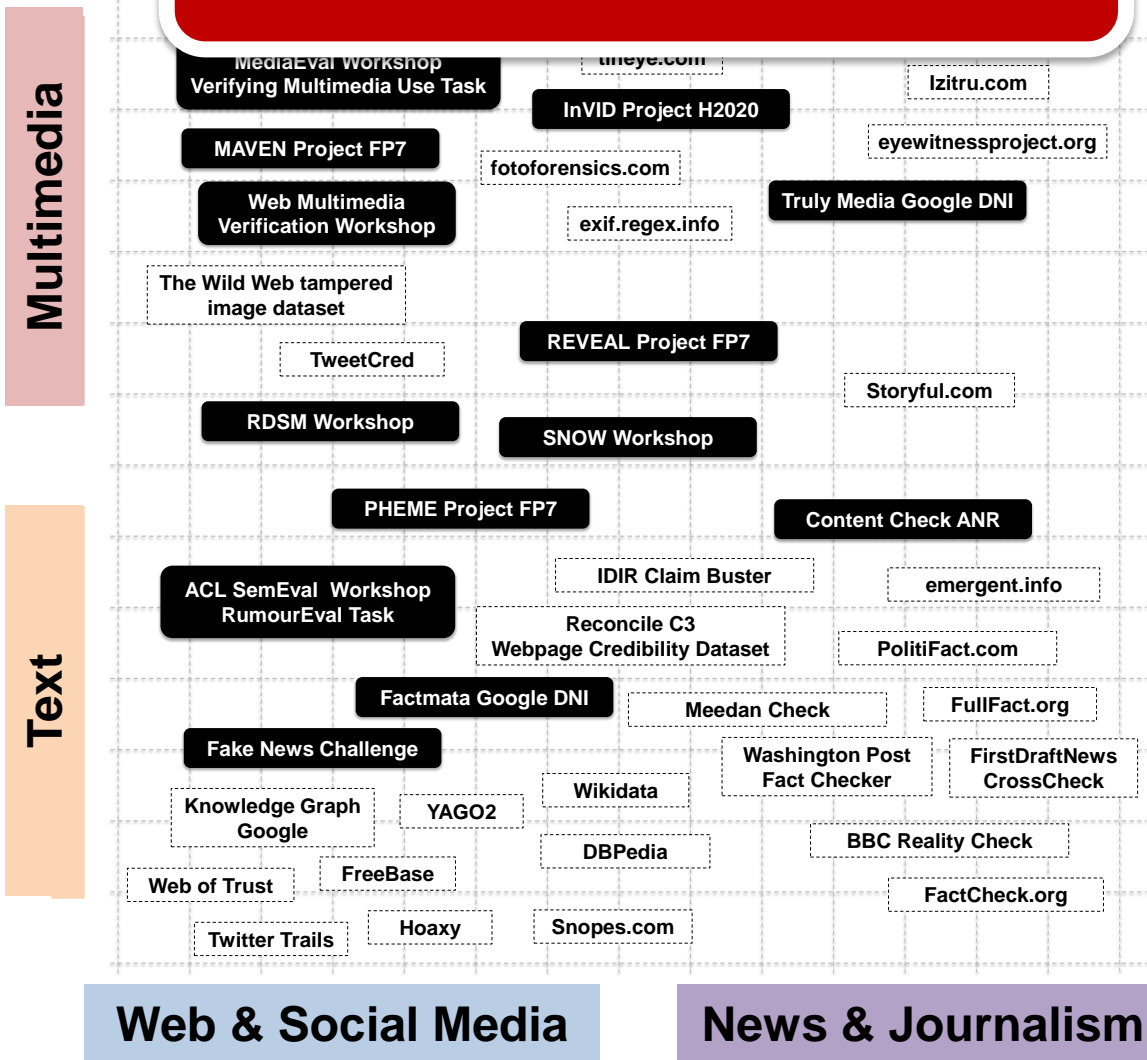
News & Journalism

C. Boididou, S. E. Middleton, Z. Jin, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, and Y. Kompatsiaris. "Verifying information with multimedia content on Twitter", Multimedia Tools and Applications, Sep. 2017, Pages 1-27, Springer 2017

Middleton, S.E. Kordopatis-Zilos, G. Papadopoulos, S. Kompatsiaris, Y. "Location Extraction from Social Media: Geoparsing, Location Disambiguation, and Geotagging", ACM Transactions on Information Systems (TOIS) 36, 4, Article 40 (June 2018)



InVID Project H2020 Video Forensics



Key Datasets and Fact Checking Sites



Verifi

Search

InVID Project H2020 Video Forensics

MediaEval Workshop
Verifying Multimedia Use Task
theeye.com
Izitr.com



- ANALYSIS
- KEYFRAMES
- THUMBNAILS
- SEARCH
- MAGNIFIER
- METADATA
- FORENSIC
- ABOUT

Video contextual verification

Copy and paste a Youtube or Facebook url

Submit

A plugin to debunk fake news and to verify videos and images

Twitter Trails Hoaxy Snopes.com

Web & Social Media

News & Journalism

Key Datasets and Fact Checking Sites

AI support for content verification

- Social media and web scale analytics
 - Information extraction
 - Tending/emerging events & factual claims
 - Context (profile, comments, network) > Verify source & claim
 - Classify image & videos from posted comments
 - Eyewitness, Fake/Genuine, First person report



AI support for content verification


- Social media and web scale analytics
 - Information extraction
 - Topic & event models, Entity extraction, OpenIE, Geoparsing
 - Social network analysis, Temporal traffic analysis
 - Classify image & videos from posted comments
 - Supervised & semi-supervised learning, Language models




AI support for content verification

MALTA Refresh Data ABOUT


Content Items (duplicates pruned)



ABC News Politics @ABCPolitics
MALTA: hostage situation with hijacked plane now over, the Maltese Prime Minister's spokesman confirms:
3:07 PM - 23 Dec 2016



DannyDK @LoveNServeAll
JUST IN: Malta hostage situation with hijacked plane now over, the Maltese Prime Minister's spokesman confirms:
3:08 PM - 23 Dec 2016



Settings

Thumbnail Metric: incident

Start time: 2000-01-01T00:00:00

End time: 3000-01-01T00:00:00

Admin level: Country

Top N Posts: 10

Top N Thumbnails: 15

AI support for content verification


- Digital image forensics
 - Manual content inspection
 - Background landmarks, weather, signposts, insignia
 - Image metadata analysis
 - Manipulation detection algorithms

Home | Tutorial | About | Github

Metadata summary ⓘ <>

JPEG		JpegComment		JFIF		File	
description	value	description	value	description	value	description	value
Compression Type	Baseline	CREATOR	gd-jpeg v1.0 (using IJG JPEG v62), quality = 75	Version	1.1	File Name	Raw
Data Precision	8 bits	JPEG Comment		Resolution Units	none	File Size	53539 bytes
Image Height	733 pixels			X Resolution	1 dot	File Modified Date	Tue Jul 12 14:39:03 +03:00 2016
Image Width	550 pixels			Y Resolution	1 dot		
Number of Components	3			Thumbnail Width Pixels	0		
Component 1	Y component: Quantization table 0, Sampling factors 2 horiz/2 vert			Thumbnail Height Pixels	0		

send this image to [Google reverse image search](#)



Map 0: Double JPEG quantization inconsistencies (DQ) - What does it mean?

Map 1: JPEG Ghosts (GHOST) - What does it mean?

Map 2: JPEG blocking artifact inconsistencies (BLOCK) - What does it mean?

Map 3: Error Level Analysis (ELA) - What does it mean?

Map 4: Median filtering noise residue (MEDIAN) - What does it mean?

Map 5: High frequency noise (WAVELET) - What does it mean?

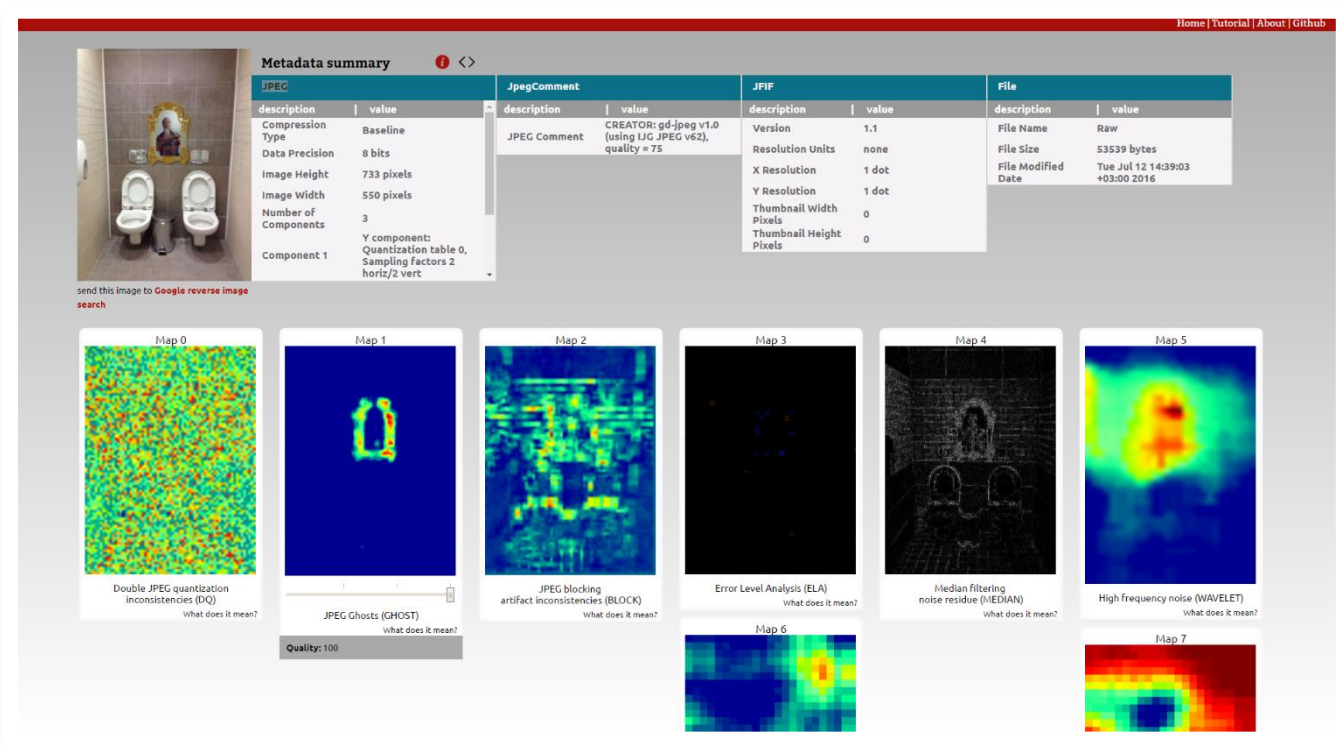
Map 6: [Unlabeled]

Map 7: [Unlabeled]

Quality: 100

AI support for content verification

- Digital image forensics
 - Manual content inspection
 - EXIF metadata e.g. camera data for checking with author
 - Manipulation detection algorithms
 - Splice/copy-move detection using JPEG compression traces



Home | Tutorial | About | Github

Metadata summary ⓘ <>

JPEG		JpegComment		JFIF		File	
description	value	description	value	description	value	description	value
Compression Type	Baseline	description	value	Version	1.1	File Name	Raw
Data Precision	8 bits	JPEG Comment	CREATOR: gd-jpeg v1.0 (using IJG JPEG v62), quality = 75	Resolution Units	none	File Size	53539 bytes
Image Height	733 pixels			X Resolution	1 dot	File Modified Date	Tue Jul 12 14:39:03 +03:00 2016
Image Width	550 pixels			Y Resolution	1 dot		
Number of Components	3			Thumbnail Width Pixels	0		
Component 1	Y component: Quantization table 0, Sampling factors 2 horiz/2 vert			Thumbnail Height Pixels	0		

send this image to [Google reverse image search](#)

Map 0: Double JPEG quantization inconsistencies (DQ) - What does it mean?

Map 1: JPEG Ghosts (GHOST) - What does it mean?

Map 2: JPEG blocking artifact inconsistencies (BLOCK) - What does it mean?

Map 3: Error Level Analysis (ELA) - What does it mean?

Map 4: Median filtering noise residue (MEDIAN) - What does it mean?

Map 5: High frequency noise (WAVELET) - What does it mean?

Map 6: [Unlabeled]

Map 7: [Unlabeled]

Quality: 100

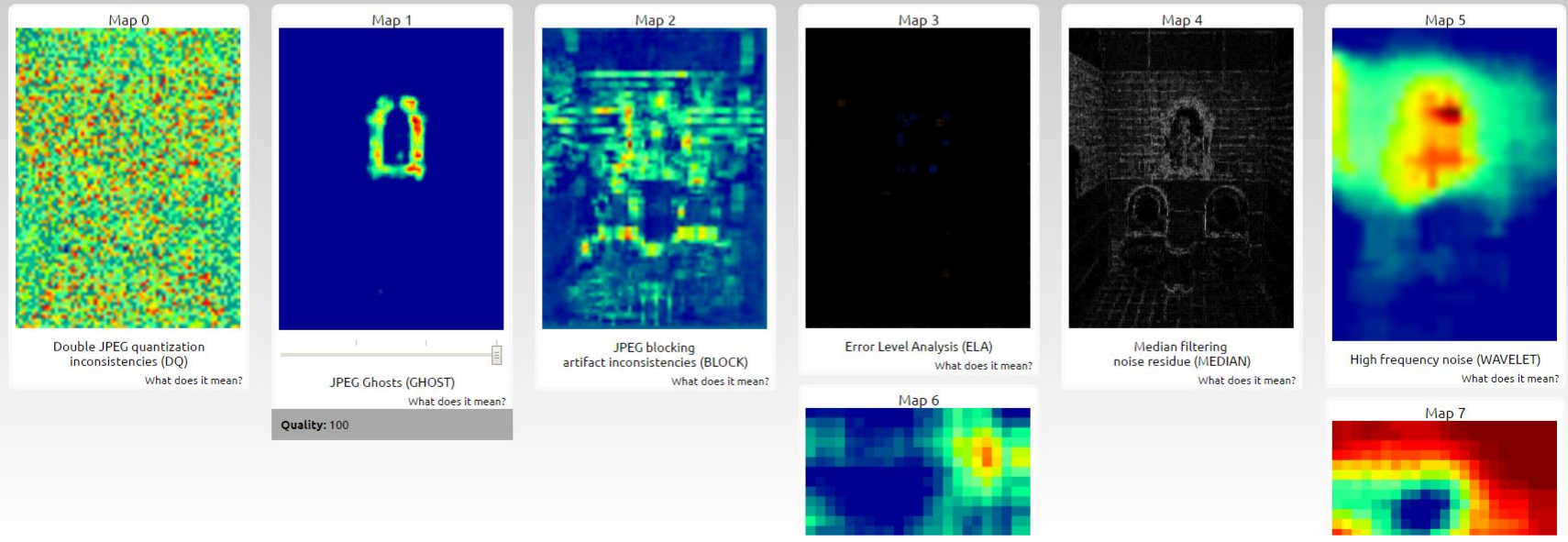
AI support for content verification



send this image to [Google reverse image search](#)

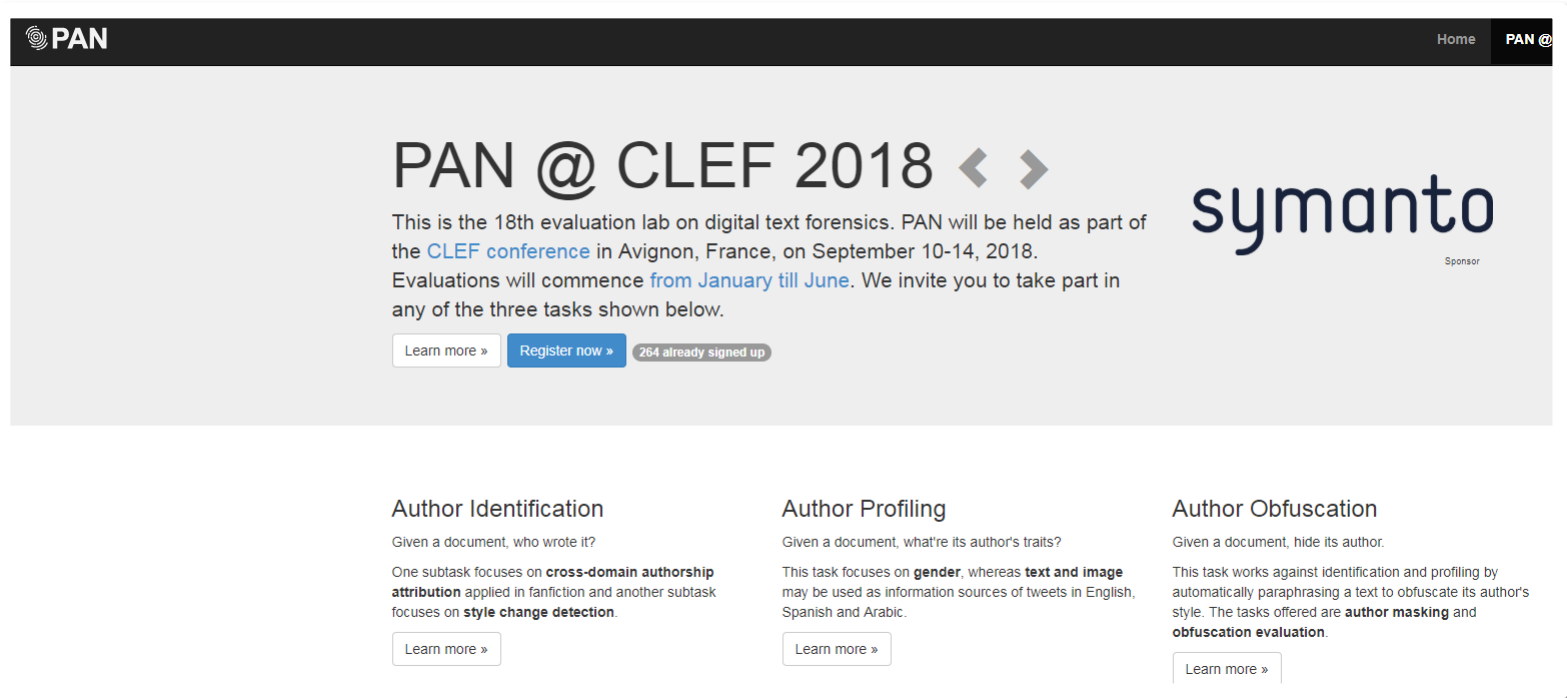
Metadata summary

JPEG		JpegComment		JFIF		File	
description	value	description	value	description	value	description	value
Compression Type	Baseline	JPEG Comment	CREATOR: gd-jpeg v1.0 (using IJG JPEG v62), quality = 75	Version	1.1	File Name	Raw
Data Precision	8 bits			Resolution Units	none	File Size	53539 bytes
Image Height	733 pixels			X Resolution	1 dot	File Modified Date	Tue Jul 12 14:39:03 +03:00 2016
Image Width	550 pixels			Y Resolution	1 dot		
Number of Components	3			Thumbnail Width Pixels	0		
Component 1	Y component: Quantization table 0, Sampling factors 2 horiz/2 vert			Thumbnail Height Pixels	0		



AI support for content verification

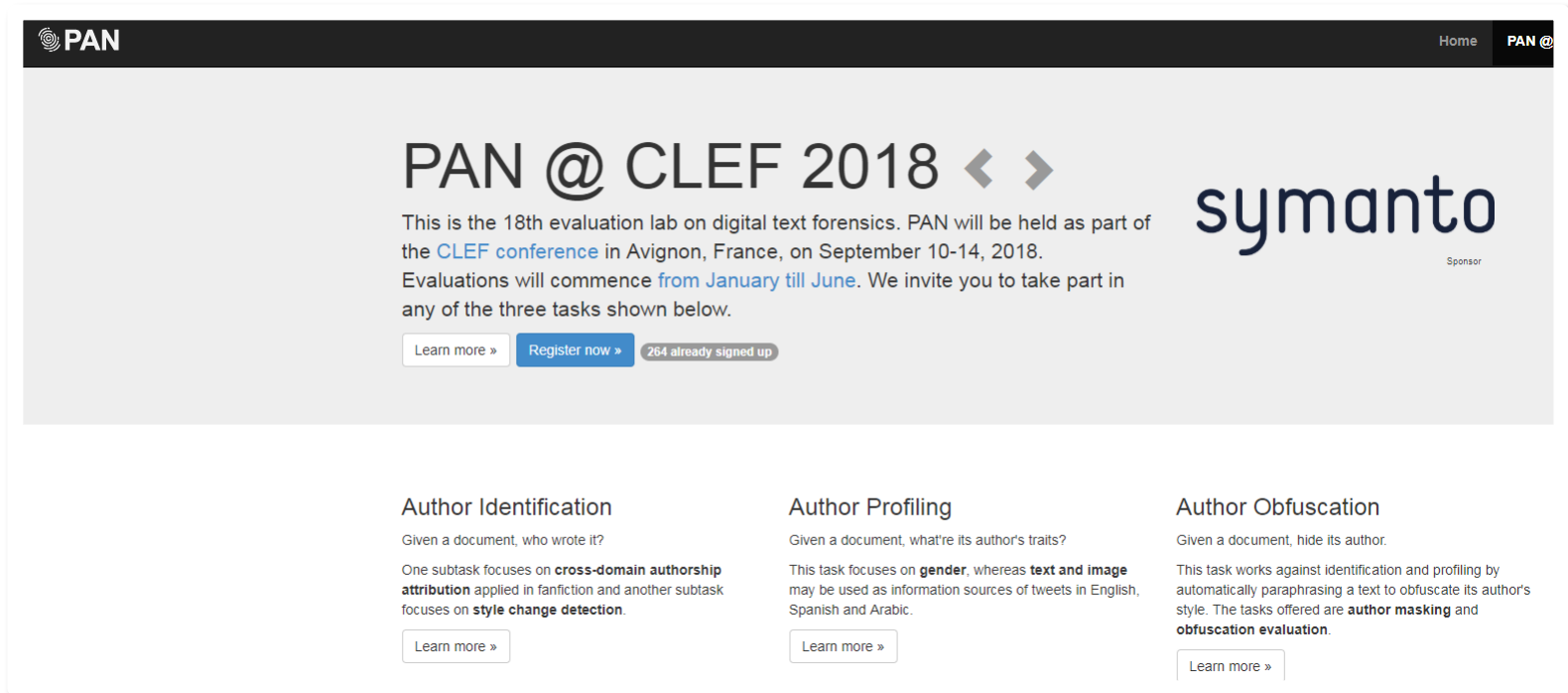
- Digital text forensics
 - Manual linguistic expert analysis
 - Author attribution, profiling & clustering
 - Academic > PAN conference series
 - Interest from law enforcement agencies (e.g. FloraGuard)



The screenshot shows the website for PAN @ CLEF 2018. The header includes the PAN logo and navigation links for Home and PAN @. The main content area features the title "PAN @ CLEF 2018" with left and right navigation arrows. Below the title, a paragraph states: "This is the 18th evaluation lab on digital text forensics. PAN will be held as part of the CLEF conference in Avignon, France, on September 10-14, 2018. Evaluations will commence from January till June. We invite you to take part in any of the three tasks shown below." To the right of this text is the Symantec logo with "Sponsor" written below it. Below the paragraph are three buttons: "Learn more >", "Register now >" (highlighted in blue), and "264 already signed up". The bottom section of the page is divided into three columns, each describing a task: "Author Identification", "Author Profiling", and "Author Obfuscation". Each column includes a brief description and a "Learn more >" button.

AI support for content verification

- Digital text forensics
 - Manual linguistic expert analysis
 - Author attribution, profiling & clustering
 - Supervised learning (word and character n-gram features), zip/checksum, repetitions, statistical profiling of vocabulary

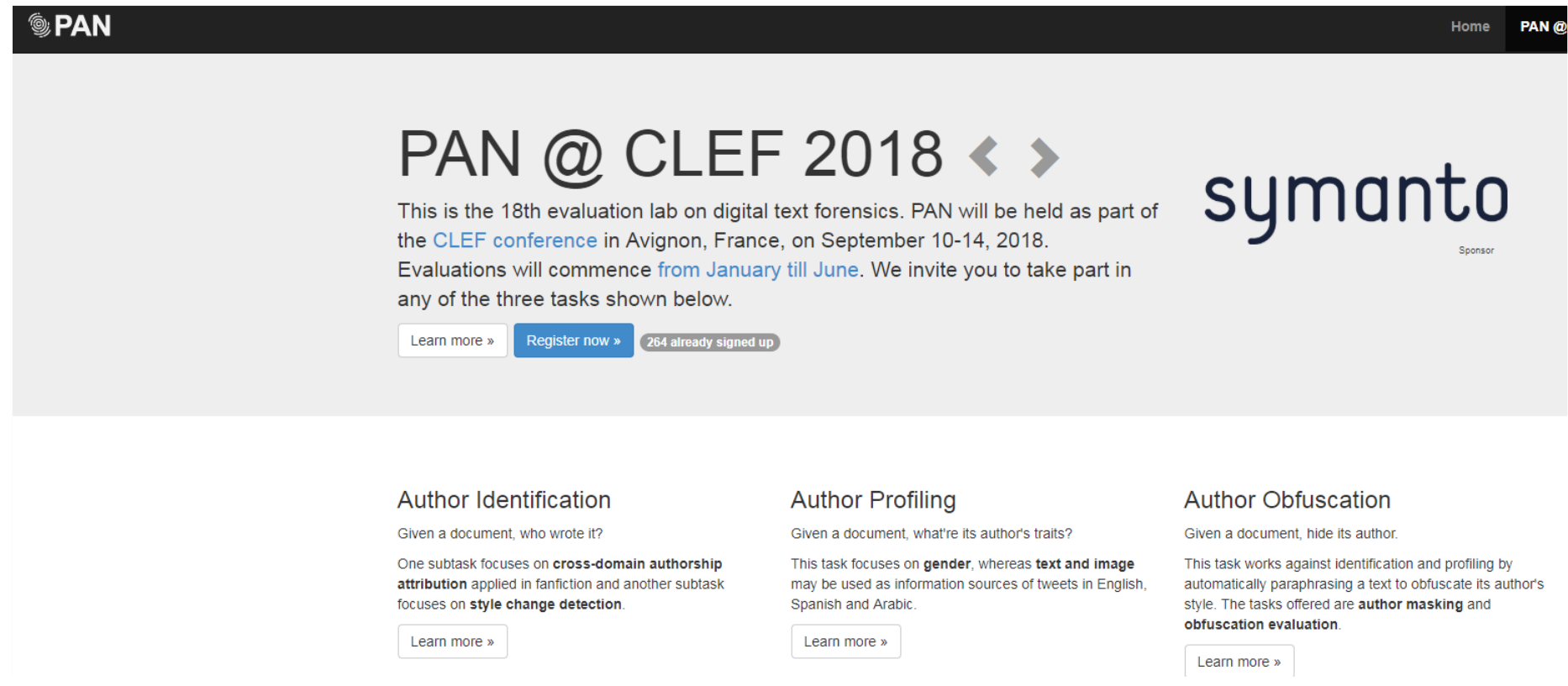


The screenshot shows the homepage for PAN @ CLEF 2018. At the top left is the PAN logo, and at the top right are links for 'Home' and 'PAN @'. The main heading is 'PAN @ CLEF 2018' with left and right navigation arrows. Below this, a paragraph states: 'This is the 18th evaluation lab on digital text forensics. PAN will be held as part of the CLEF conference in Avignon, France, on September 10-14, 2018. Evaluations will commence from January till June. We invite you to take part in any of the three tasks shown below.' To the right of this text is the Symantec logo with 'Sponsor' written below it. Below the main text are three buttons: 'Learn more >', 'Register now >' (highlighted in blue), and '264 already signed up'. The page is divided into three columns, each representing a task:

- Author Identification**: Given a document, who wrote it? One subtask focuses on **cross-domain authorship attribution** applied in fanfiction and another subtask focuses on **style change detection**. A 'Learn more >' button is at the bottom.
- Author Profiling**: Given a document, what're its author's traits? This task focuses on **gender**, whereas **text and image** may be used as information sources of tweets in English, Spanish and Arabic. A 'Learn more >' button is at the bottom.
- Author Obfuscation**: Given a document, hide its author. This task works against identification and profiling by automatically paraphrasing a text to obfuscate its author's style. The tasks offered are **author masking** and **obfuscation evaluation**. A 'Learn more >' button is at the bottom.

AI support for content verification

- Digital text forensics
 - Manual linguistic expert analysis
 - Author attribution, profiling & clustering



The screenshot shows the PAN @ CLEF 2018 website. The header includes the PAN logo and navigation links for Home and PAN @. The main content area features the title "PAN @ CLEF 2018" with left and right navigation arrows. Below the title is a paragraph of text: "This is the 18th evaluation lab on digital text forensics. PAN will be held as part of the CLEF conference in Avignon, France, on September 10-14, 2018. Evaluations will commence from January till June. We invite you to take part in any of the three tasks shown below." To the right of this text is the Symanto logo with the word "Sponsor" underneath. Below the text are three buttons: "Learn more »", "Register now »", and "264 already signed up".

The website is divided into three columns, each representing a task:

- Author Identification**: Given a document, who wrote it? One subtask focuses on **cross-domain authorship attribution** applied in fanfiction and another subtask focuses on **style change detection**. A "Learn more »" button is located below the text.
- Author Profiling**: Given a document, what're its author's traits? This task focuses on **gender**, whereas **text and image** may be used as information sources of tweets in English, Spanish and Arabic. A "Learn more »" button is located below the text.
- Author Obfuscation**: Given a document, hide its author. This task works against identification and profiling by automatically paraphrasing a text to obfuscate its author's style. The tasks offered are **author masking** and **obfuscation evaluation**. A "Learn more »" button is located below the text.

Trends around AI and content verification

- AI helping to handle larger content volumes
 - Machine learning increasingly automating pre-filtering of content prior to human verification work
 - Crowdsourcing being regularly used for fact checking
 - Eyewitness, Local experts, Volunteer fact checkers
 - Better integration of open data from trusted sources
 - Government open data, OpenStreetMap, Demographic/census data, Wikipedia
 - Use of distributed AI and big data to scale up algorithms
 - Big data solutions from companies owning big datasets (e.g. Facebook algorithms to take down terrorist or child sex abuse content)
- Platforms as gateways to original content
 - Social media platform API's increasingly the only access to UGC
 - Metadata stripping by platforms could prevent many forensic techniques that needs access to original content

Trends around AI and content verification

- AI producing better fake content
 - Generative Adversarial Networks (GAN) for face swap / deepfake
 - AI can be trained to spot fakes, but AI can also be trained to produce ever more realistic fake images and videos
 - Future content might never be trusted without provenance or corroborating evidence
- Systems being designed to enhance trust in AI
 - Increased focus on AI algorithm transparency, bias, ethics
 - Better interfaces to verification AI to allow human-machine collaboration
 - Explainability of AI results so they can be understood
 - Journalists (and courts of law) need verification methods to be explainable and causal links from findings back to original evidence



Thanks you for your attention!

Any questions?

Dr Stuart E. Middleton

University of Southampton, Electronics and Computer Science, IT Innovation Centre

email: sem03@soton.ac.uk

web: www.ecs.soton.ac.uk/people/sem

www.it-innovation.soton.ac.uk

twitter: [@stuart_e_middle](https://twitter.com/stuart_e_middle)