PAN

CLEF 2023 Thessaloniki

Digital Text Forensics and Stylometry

Janek Bevendorff Bauhaus-Universität Weimar

Ian Borrego-Obrador Symanto Research

Mara Chinea-Ríos Symanto Research

Marc Franco-Salvador Symanto Research

Maik Fröbe Friedrich Schiller University Jena

Annina Heini Aston University

Krzysztof Kredens Aston University

Maximilian Mayerl University of Innsbruck

Piotr Pęzik Aston University

Martin Potthast Leipzig University and ScaDS.AI

Francisco Rangel Symanto Research

Paolo Rosso Universitat Politècnica de València

Efstathios Stamatatos University of the Aegean

Benno Stein Bauhaus-Universität Weimar

Matti Wiegmann Bauhaus-Universität Weimar

Magdalena Wolska Bauhaus-Universität Weimar

Eva Zangerle University of Innsbruck

PAN - History

- Started in 2009 (held with SEPLN)
- Since 2010 it is held with CLEF
- Almost 70 shared tasks

Statistics	SEPLN	CLEF												
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Follower	78	151	181	232	286	302	333	337	347	368	476	486	496	502	506
Registrations	21	53	52	68	110	103	148	143	191	227	373	306	173	117	70
Runs/Software	14	27	27	48	58	57	54	37	34	41	76	109	84	81	67
Notebooks	11	22	22	34	47	36	52	29	30	35	66	46	52	52	36

PAN - History

Originality:

- Plagiarism detection (2009-15)

Computational ethics:

- Wikipedia vandalism detection (2010-11)
- Wikipedia quality flaw prediction (2012)
- Sexual predator identification (2012)

Authorship analysis:

- Identification (2011-12, 2018-19)
- Profiling (2013-22)
- Verification (2013-15,2020-22)
- Clustering (2016-17)
- Diarization (2016-22)
- Obfuscation (2016-18)

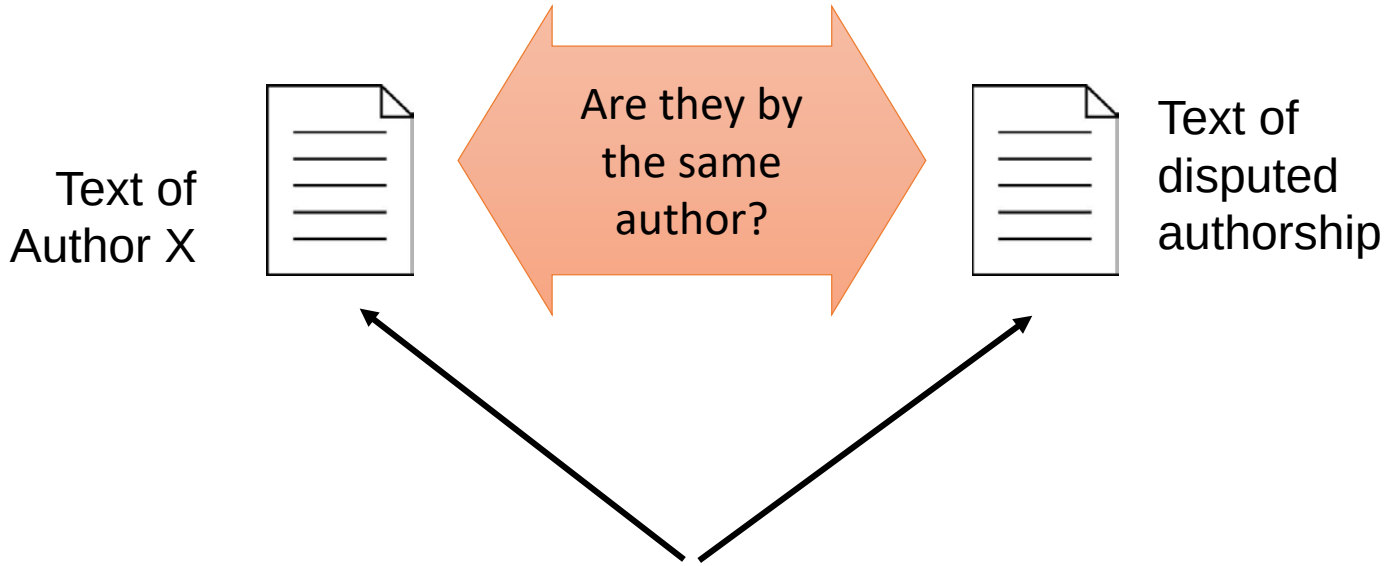
PAN 2023 Tasks

- Authorship verification (7th edition)
  - Cross-discourse type authorship verification
- Author diarization (8th edition)
  - Multi-author writing style analysis
- Author profiling (11th edition)
  - Profiling cryptocurrency influencers with few-shot learning
- Trigger detection (New!)

- Software submissions
  - Deployed (**new**: docker containers) and evaluated in TIRA

M. Fröbe, et al. 2023. **Continuous Integration for Reproducible Shared Tasks with TIRA.io**, In Proc. of ECIR-2023.

Cross-Discourse Type Authorship Verification



- The texts belong to different discourse types (e.g. essays, emails)
- **New in 2023**: Both written and spoken discourse types are included

Authorship Verification 2023 Dataset

	Training	Test
Text pairs		
Positive	4,418 (50.0%)	4,828 (50.0%)
Negative	4,418 (50.0%)	4,828 (50.0%)
Email - Speech transcription	1,036 (11.7%)	1,074 (11.1%)
Essay - Email	1,454 (16.5%)	1,618 (16.8%)
Essay - Interview	884 (10.0%)	938 (9.7%)
Essay - Speech transcription	256 (2.9%)	206 (2.1%)
Interview - Email	4,564 (51.7%)	5,214 (54.0%)
Speech transcription - Interview	642 (7.3%)	606 (6.3%)
Text length (avg. chars)		
Email	2,308	2,346
Essay	9,894	10,770
Interview	2,503	2,501
Speech transcription	2,395	2,537

- Based on Aston 100 Idiolects Corpus in English

Authorship Verification Results

- 11 participants
- 27 runs
- 4 baselines

Systems	AUROC	c@1	F_1	$F_{0.5u}$	Brier	Overall
Ibrahim, et al. (reduced-graph) [19]	**0.616**	**0.572**	0.617	0.562	0.746	**0.623**
Ibrahim, et al. (resolving-globe) [19]	**0.616**	**0.572**	0.617	0.562	0.746	**0.623**
Guo, et al. (irregular-strategist) [14]	0.581	0.557	0.621	**0.571**	0.742	0.614
Ibrahim, et al. (golden-ottoman) [19]	0.598	0.546	0.622	0.550	0.744	0.612
BASELINE (cngdist)	0.516	0.499	**0.666**	0.555	0.741	0.595
Petropoulos (graceful-chianti) [40]	0.526	0.514	0.624	0.549	0.743	0.591
Petropoulos (clever-daemon) [40]	0.525	0.516	0.622	0.550	0.743	0.591
BASELINE (galicia22)	0.504	0.502	0.650	0.552	0.740	0.589
Valdez Valenzuela, et al. (GNN-SHORT) [70]	0.511	0.508	0.655	0.555	0.705	0.587
Sun, et al. (SDML epoch 8) [66]	0.504	0.502	0.632	0.546	0.747	0.586
Sun, et al. (SDML epoch 24) [66]	0.505	0.501	0.601	0.536	0.749	0.578
Guo, et al. (uniform-reward) [14]	0.595	0.555	0.460	0.527	0.723	0.572
Valdez Valenzuela, et al. (GNN-FULL) [70]	0.517	0.512	0.628	0.549	0.644	0.570
Sun, et al. (SDML epoch 35) [66]	0.511	0.508	0.558	0.526	0.749	0.570
Valdez Valenzuela, et al. (GNN-MED) [70]	0.503	0.502	0.602	0.534	0.709	0.570
BASELINE (najafi22)	0.601	0.569	0.466	0.543	0.595	0.555
Huang, et al. (isochoric-paint) [18]	0.563	0.563	0.511	0.550	0.563	0.550
Liu, et al. (coincident-sound) [30]	0.548	0.548	0.544	0.547	0.548	0.547
Lv (radioactive-copyright) [33]	0.553	0.553	0.504	0.540	0.553	0.541
Huang, et al. (steel-coriander) [18]	0.500	0.500	0.651	0.551	0.500	0.540
Li, et al. (wan-ocean) [28]	0.500	0.500	0.646	0.550	0.500	0.539
Lv, et al. (tender-bugle) [33]	0.551	0.551	0.501	0.537	0.551	0.538
Lv, et al. (cold-rotor) [33]	0.550	0.550	0.465	0.524	0.550	0.528
Qiu, et al. (corn-mall) [42]	0.540	0.540	0.421	0.499	0.540	0.508
Qiu, et al. (poky-deck) [42]	0.540	0.540	0.421	0.499	0.540	0.508
Liu, et al. (perpendicular-field) [30]	0.534	0.534	0.421	0.493	0.534	0.503
Liu, et al. (foggy-raster) [30]	0.533	0.533	0.424	0.493	0.533	0.503
BASELINE (compressor)	0.506	0.051	0.626	0.076	**0.750**	0.402
Sanjesh, et al. (calm-lyrics) [58]	0.525	0.500	0.030	0.068	0.729	0.370
Sanjesh, et al. (null-midpoint) [58]	0.523	0.499	0.031	0.066	0.730	0.370
Sanjesh, et al. (Multi-Feature Classifier) [58]	0.501	0.01	0.000	0.000	0.750	0.252

Multi-author Writing Style Analysis

- Identify all positions of writing style changes on the paragraph level
- **New in 2023**: control topic similarity between paragraphs
  - Easy: paragraphs cover a variety of topics
  - Medium: topical variety is small
  - Hard: all paragraphs are on the same topic

Multi-author Writing Style Analysis 2023 Dataset

- A new dataset based on Reddit
  - *r/worldnews*
  - *r/politics*
  - *r/askhistorians*
  - *r/legaladvice*
- Synthetic texts
  - Concatenation of paragraphs
  - Control of thematic similarity
- 6,000 texts
  - 70% training
  - 15% validation
  - 15% test

Multi-author Writing Style Analysis: Results

- 6 participants

Systems	Easy F_1	Medium F_1	Hard F_1
Ye et al. [76]	0.983	0.830	**0.821**
Hashemi et al. [15]	**0.984**	**0.843**	0.812
Kucukkaya et al. [24]	0.982	0.810	0.772
Huang et al. [17]	0.968	0.806	0.769
Chen et al. [6]	0.914	0.820	0.676
Jacobo et al. [20]	0.793	0.591	0.498

Profiling Cryptocurrency Influencers with Few-shot Learning

- SubTask1: Low-resource influencer profiling:
  - Profile authors according to their degree of influence
  - [non-influencer, nano, micro, macro, mega]
- SubTask2: Low-resource influencer interest profiling:
  - Profile authors according to their main interests or areas of influence
  - [technical information, price update, trading matters, gaming, other]
- SubTask3: Low-resource influencer intent profiling:
  - Profile authors according to the intent of their messages
  - [subjective opinion, financial, information, advertising, announcement]

Author Profiling 2023 Dataset

Task	Partition	Total number of users per class
1	train	macro:32, mega:32, micro:30, nano:32, non-influencer:32
	test	macro:42, mega:45, micro:46, nano:45, non-influencer:42
2	train	technical information:64; trading matters:64; price update:64; gaming:64; other:64
	test	technical information:42; trading matters:112; price update:108; gaming:40; other:100
3	train	announcement:64; subjective opinion:64; financial information:64; advertising:64
	test	announcement:37; subjective opinion:160; financial information:43; advertising:52

- A new dataset based on English tweets
  - Task 1: number of followers indicate cryptocurrency influencer class
  - Task 2 and 3: Three human annotators label tweets
- Limitation of tweets per user
  - Task 1: 10 tweets
  - Task 2 and 3: 1 tweet

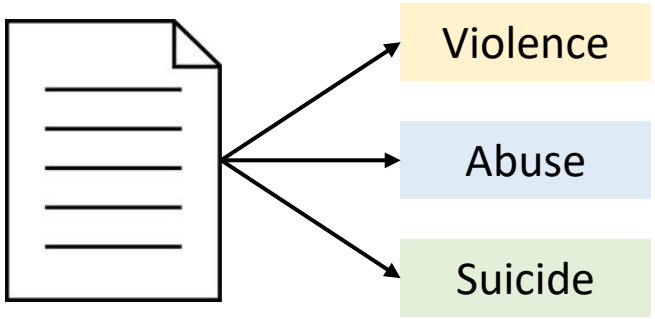
Author Profiling: Results

- 27 participants
- 5 baselines

Systems	Macro F_1		
	ST1 (Influence)	ST2 (Interest)	ST3 (Intent)
Cano-Caravaca (terra-classic)	61.14	63.15	**67.46**
Villa-Cueva et al. (stellar) [72]	58.44	**67.12**	64.46
(MRL-LLP)	57.44	62.00	65.74
Balanzá García (holo)	**62.32**	57.50	61.81
Giglou et al.(symbol) [12]	52.31	61.21	65.83
Cardona-Lorenzo (vechain)	55.51	60.16	60.28
Carbonell Granados (shiba-inu)	50.38	58.47	66.15
Ferri-Molla et al. (magic) [35]	57.14	55.68	61.62
Li et al.(neo) [29]	55.10	61.63	57.62
Iranzo Sánchez (iota)	54.43	64.55	50.62
t5-large (label tuning) - FS	49.34	56.48	59.91
Huallpa (hive)	52.94	51.48	59.08
Llanes Lacomba (api3)	49.18	46.07	63.12
Labadie et al.(dogecoin) [26]	50.80	51.72	52.59
Casamayor Segarra (tron)	50.13	49.77	53.43
user-char-lr	35.25	52.95	60.21
de Castro Isasi (terra)	48.74	44.60	54.83
Rodríguez Ferrero (harmony)	47.93	54.41	45.83
LDSE	50.20	44.92	51.96
Jaramillo-Hernández (waves)	55.06	42.35	49.21
Girish et al. [13]	37.92	46.66	50.42
Espinosa et al. (core) [9]	34.76	43.47	55.34
Coto et al. (ethereum)	46.68	-	55.94
García Bohigues (sushiswap)	46.64	19.23	22.58
t5-large (bi-encoders) - ZS	12.76	33.34	32.71
random	15.92	20.81	18.41
Kumar et al. [25]	50.21	-	-
Siino et al. (alchemy-pay) [60]	38.51	-	-
Siino et al. (nexo) [61]	38.34	-	-
Lomonaco et al. (wax) [31]	37.62	-	-
Valles Silva (solana)	15.92	-	-
Muslihuddeen et al. (icon) [38]	12.90	-	-

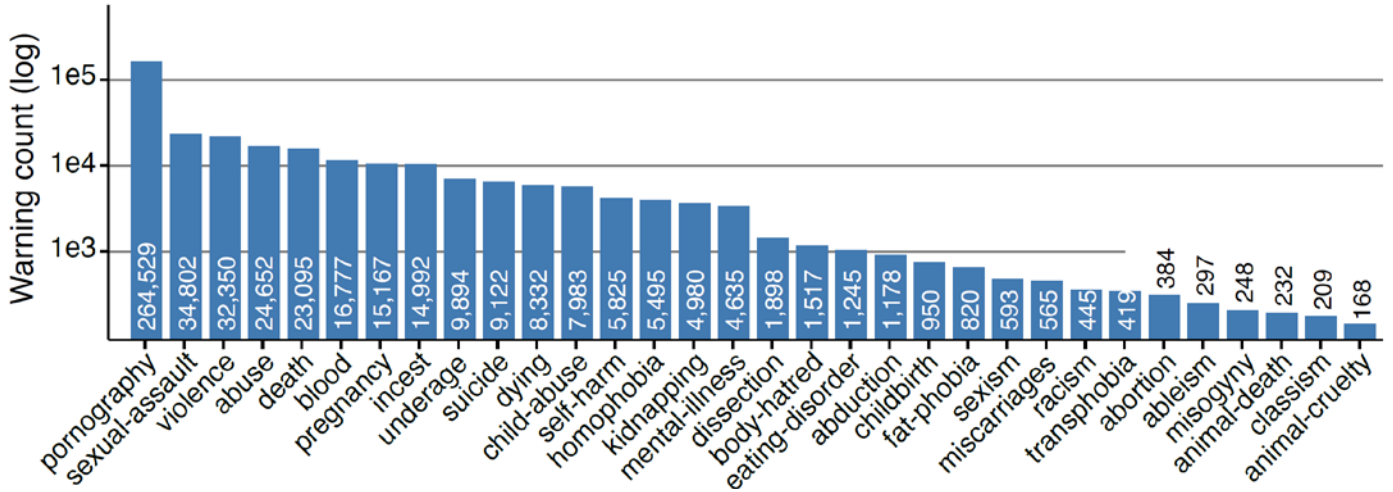
Trigger Detection

- Trigger: a stimulus that elicits negative emotions or feelings of distress
- Task definition:
  - Identifying whether or not a given document contains triggering content, and if so, of what type
  - Multi-label text classification



Trigger Detection: Dataset

- Fanfiction
  - Archive of our Own
- User-generated labels
  - 32 triggering labels



Training Dataset	
Total Works	307,102
< 512 words	15,233
< 4,096 words	261,156
Mean no. words	2,400
Median no. words	2,126
90pct no. words	4,579

Validation Dataset	
Total Works	17,104
< 512 words	861
< 4,096 words	14,571
Mean no. words	2,386
Median no. words	2,115
90pct no. words	4,550

Test Dataset	
Total Works	17,040
< 512 words	813
< 4,096 words	14,555
Mean no. words	2,388
Median no. words	2,101
90pct no. words	4,558

Trigger Detection: Results

- 6 participants
- 1 baseline

Systems	Macro			Micro			Acc
	Prec	Rec	F_1	Prec	Rec	F_1	
Sahin et al. [57]	0.37	0.42	**0.352**	0.73	0.74	0.74	0.59
Su et al. [65]	**0.54**	0.30	**0.350**	0.80	0.71	**0.75**	**0.62**
XGBoost baseline	0.52	0.25	0.301	**0.88**	0.57	0.69	0.53
Cao, H. et al. [5]	0.24	0.29	0.228	0.43	0.79	0.56	0.18
Cao, G. et al. [4]	0.28	0.22	0.225	0.58	0.66	0.62	0.32
Felser et al. [10]	0.11	**0.63**	0.161	0.27	**0.82**	0.40	0.27
Shashirekha et al. [27]	0.10	0.04	0.048	0.82	0.50	0.63	0.52

PAN Program

https://pan.webis.de/

- Monday, September 18
  - 16:10 – 17:40: Keynote-1 & Author profiling

    Yiannis Kompatsiaris: **Behavioural and Policy Aspects of Online Disinformation**
- Tuesday, September 19
  - 09:30 – 11:20: Keynote-2 & Author profiling

    Alejandro Martin: **Detection and Disrupting Disinformation: Social Network Analysis and NLP**
  - 14:00 – 15:30: Authorship verification
  - 16:00 – 17:30: Multi-author writing style analysis & Trigger detection

- All sessions in Room 1