Overview of PAN 2024: Multi-Author Writing Style Analysis, Multilingual Text Detoxification, Oppositional Thinking Analysis, and Generative AI Authorship Verification

Extended Abstract

Janek Bevendorff,[1] Xavier Bonet Casals,[2] Berta Chulvi,[3] Daryna Dementieva,[4] Ashaf Elnagar,[5] Dayne Freitag,[6] Maik Fröbe,[7] Damir Korenčić,[3] Maximilian Mayerl,[8] Animesh Mukherjee,[9] Alexander Panchenko,[10] Martin Potthast,[1,11] Francisco Rangel,[12] Paolo Rosso,[3,13] Alisa Smirnova,[14] Efstathios Stamatatos,[15] Benno Stein,[16] Mariona Taulé,[2] Dmitry Ustalov,[14] Matti Wiegmann,[16] and Eva Zangerle[8]

[1]Leipzig University, Germany, [2]Universitat de Barcelona, Spain, [3]Univ. Politècnica de València, Spain, [4]Technical University of Munich, Germany, [5]University of Sharjah, United Arab Emirates, [6]SRI International, USA, [7]Friedrich Schiller University Jena, Germany, [8]University of Innsbruck, Austria, [9]Indian Institute of Technology Kharagpur, India, [10]Skolkovo Institute of Science and Technology, Russia, [11]ScaDS.AI, Germany, [12]Symanto Research, Spain, [13]ValgrAI - Valencian Graduate School and Research Network of Artificial Intelligence, Spain, [14]Toloka, Switzerland, [15]University of the Aegean, Greece, [16]Bauhaus-Universität Weimar, Germany

pan@webis.de pan.webis.de

Abstract The paper gives a brief overview of the four shared tasks organized at the PAN 2024 lab on digital text forensics and stylometry to be hosted at CLEF 2024. The goal of the PAN lab is to advance the state-of-the-art in text forensics and stylometry through an objective evaluation of new and established methods on new benchmark datasets. Our four tasks are: (1) multi-author writing style analysis, which we continue from 2023 in a more difficult version, (2) multilingual text detoxification, a new task that aims to translate and re-formulate text in a non-toxic way, (3) oppositional thinking analysis, a new task that aims to discriminate critical thinking from conspiracy narratives and identify their core actors, and (4) generative AI authorship verification, which formulates the detection of AI-generated text as an authorship problem, one of PAN's core tasks. As with the previous editions, PAN invites software submissions as easy-to-reproduce docker containers; more than 400 pieces of software have been submitted from PAN'12 through PAN'23 combined, with all recent evaluations running on the TIRA experimentation platform [8].

1 Introduction

PAN is a workshop series and a networking initiative for stylometry and digital text forensics. PAN hosts computational shared tasks on authorship analysis, computational ethics, and the originality of writing. Since the workshop's inception in 2007, we organized 64 shared tasks[1] and assembled 55 evaluation datasets[2] plus nine datasets contributed by the community.

In 2023, our four tasks concluded with 49 submissions and 35 notebook papers. The *Multi-Author Writing Style Analysis* task was revamped for 2023 with a new dataset and structured around topical heterogeneity as an indicator for difficulty. The task attracts consistent participation of high technical quality, while the problem is still relevant and offers room for improvements, hence we continue the task with only slight modifications in 2024. The trigger detection task was newly introduced in 2023 and concluded with a variety of different solutions. While we see value in continuing to refine the task and study other promising approaches, we postpone its renewal until further ground truth can be assembled. Instead, we introduce the new *Multilingual Text Detoxification* task to better align with the interest of our community on countering toxicity and in generative tasks. The profiling cryptocurrency influencer task continued a series of author profiling tasks and concluded with high attendance and satisfying technical results. Since no significant progress is expected, we replace the task with *Oppositional Thinking Analysis* to study critical thinking and conspiracy theories in online messages. The cross-discourse type authorship verification task concluded its second iteration with mixed results and limited progress. Discriminating authorship across discourse types is difficult despite the advanced methods employed by participants. We do not expect systems to improve without further theoretical deliberation, hence we discontinue the task. Instead, we focus on the new *Generative AI Authorship Verification* task, which aims to distinguish authorship between humans and generative AI—a task of high urgency. We briefly outline the upcoming tasks in the sections that follow.

2 Multi-Author Writing Style Analysis

The purpose of multi-author writing style analysis is to identify the positions of authorship changes within a document. Using authors' writing style has been shown to allow segmenting documents into parts written by different authors, essentially conducting an intrinsic style analysis task and paving the way for intrinsic plagiarism detection (i.e., detecting plagiarism without the use of a reference corpus).

Multi-author writing style analysis has been part of PAN since 2016. Originally, participants had to identify and group the authors of fragments of a document [18]. In 2017, participants had to assess whether a document was written by a single or multiple authors [22] and, for multi-author documents, to find the

[1] Find PAN's past shared tasks at `pan.webis.de/shared-tasks.html`
[2] Find PAN's datasets at `pan.webis.de/data.html`

exact positions of authorship changes. In 2018, the task was relaxed to a binary classification task aiming to assess whether a document was written by one or more authors [11]. This classification was also part of the task in 2019–2021. In 2019, the classification task was extended to determine the number of authors of multi-authored documents [27]. In 2020, participants additionally had to find changes in authorship between paragraphs [26]. In 2021, participants had to find all style changes on the paragraph level and assign all paragraphs to authors [24]. In 2022, this was extended from paragraph to sentence level [25]. In 2023, the task was relaxed to paragraph level but controlled for the simultaneous change of authorship and topic.

In the 2024 edition of the writing style analysis task, we will continue balancing "real" style changes among paragraphs and the topical similarity of paragraphs as a signal of style change. We will ask participants to solve the following intrinsic style change detection task: "For a given text, find all positions of writing style change on the paragraph level" (i.e., determine whether a style change occurred for all pairs of consecutive paragraphs). This task will be carried out on three datasets with increasing topical similarity among paragraphs and hence, increasing difficulty levels: (1) "Easy dataset": The paragraphs of a document cover various topics, allowing to infer information about authorship changes based on topic changes; (2) 'Medium dataset": The number of topics covered in a document is limited. This requires approaches to focus on style changes (rather than topic changes) to solve the detection task effectively; (3) "Hard dataset": Every paragraph in the document has the same topic.

3 Multilingual Text Detoxification

Text detoxification is a subtask of text style transfer where the style of text should be changed from toxic to neutral while preserving the content. As language modeling advances, there is growing concern about the potential unintended consequences of this technology. One such concern is the possibility of harmful or biased texts, which could perpetuate negative stereotypes or misinformation [13]. This has led to a growing interest in AI safety and the need for approaches to mitigating these risks [3]. This presents a major challenge for researchers and practitioners in language model safety, who need to develop effective detoxification techniques that can be applied to many languages.

Our first contribution to the field of text detoxification was the creation of the first parallel corpus for English together with a language-agnostic collection pipeline called ParaDetox [16]. We used this pipeline to collect a Russian parallel corpus, which was used in the first shared task on text detoxification: RUSSE-2022 [5]. The participants had to train their models based on 7k parallel toxic↔neutral pairs in Russian. The evaluation was done in two setups—automatic and manual. For both setups, three main parameters were assessed: (1) style transfer accuracy (STA), (2) content similarity (SIM), and (3) fluency (FL). Models were ranked via the geometric mean of these three parameters. Lastly, we compiled the best practices in the evaluation of text style transfer models by comparing the relationship between automatic and manual assessment [15]. The essential challenge for detoxification is that corpora with toxic↔neutral are scarce and that cross-lingual transfer of detoxification knowledge to new languages is challenging, as shown by our preliminary experiments [17].

In this first edition of the shared task on multilingual text detoxification, we want to extend the covered languages by adding Ukrainian, German, Chinese, Arabic, Amharic, and Italian. We provide participants with development sets of 1,000 parallel pairs for each of these languages. In addition, we provide the best metric for automatic evaluation for each language: (1) STA: binary toxicity classifier; (2) cosine similarity based on text embeddings; (3) either binary fluency classifier or perplexity measurement for fluency depending on the resources available for languages. The challenge for participants will be to perform cross-lingual detoxification: use a small parallel corpus in each target language, the languages metric, and the large English-Russian parallel corpus and transfer knowledge from the resource-rich to the resource-poor language. We welcome the participants to explore any multilingual large language models [21, 4]. For cross-lingual knowledge transfer, approaches like back-translation [7], corpus translation [23], and adapter layers [14] training can be solid baselines.

To make a fair final evaluation on the test set, we will repeat the manual evaluation pipeline from RUSSE-2022 [5] and again utilize crowd-sourcing at Toloka.ai platform for manual evaluation. The obtained manual assessments will allow again to investigate correlations between automatic and manual metrics not only for Russian and English but for all other 6 mentioned above languages. Such corpus of human vs automatic metrics can provide base more accurate toxicity classifiers, content similarities, and fluency estimation models development.

4 Oppositional Thinking Analysis

Conspiracy theories are complex narratives that attempt to explain the ultimate causes of significant events as cover plots orchestrated by secret, powerful, and malicious groups [6]. A challenging aspect of identifying conspiracy with NLP models [9] stems from the difficulty of distinguishing critical thinking from conspiratorial thinking in automatic content moderation. This distinction is vital because labeling a message as conspiratorial when it is only oppositional could drive those who were simply asking questions into the arms of the conspiracy communities.

At PAN 2024 we aim at analyzing oppositional thinking, and more concretely, at discriminating conspiracy from critical narratives from a *stylometry* perspective. The task will address two new challenges for the NLP research community: (1) to distinguish the conspiracy narrative from other oppositional narratives that do not express a conspiracy mentality (i.e., critical thinking); and (2) to identify in online messages the key elements of a narrative that fuels the intergroup conflict in oppositional thinking. Accordingly, we propose two sub-tasks:

Sub-task 1 is a binary classification task differentiating between (1) critical messages that question major decisions in the public health domain, but do not promote a conspiracist mentality; and (2) messages that view the pandemic or public health decisions as a result of a malevolent conspiracy by secret, influential groups.

Sub-task 2 is a token-level classification task aimed at recognizing text spans corresponding to the key elements of oppositional narratives. Since conspiracy narratives are a special kind of causal explanation, we developed a span-level annotation scheme that identifies the goals, effects, agents, and the groups-in-conflict in these narratives.

For the creation of the corpus, we first manually compiled a list of 2,273 public *Telegram* channels in *English* and *Spanish* that contain oppositional non-mainstream views on the COVID-19 pandemic.

For the second task, a new fine-grained annotation scheme was developed with the goal of identifying, at the text span level, how oppositional and conspiracy narratives use inter-group conflict. The annotation will be performed for the described 5,000 binary-labeled messages per language. We identify the following six categories of narrative elements at the span level: *Agents* (the hidden power that pulls the strings of the conspiracy. In critical messages, agents are actors that design the mainstream public health policies: Government, WHO, . . .); *Objectives* (parts of the narrative that answer the question "What is intended by the agents of the CT or by the promoters of the action being criticized from a critical thinking perspective?"); *Consequences* (parts of the narrative that describe the effects of the agent's actions); *Facilitators* (the facilitators are those who collaborate with the conspirators; in critical messages, facilitators are those who implement the measures dictated by the authorities); *Campaigners* (in conspiracy messages, the campaigners are the ones who uncover the conspiracy theory; in critical messages, campaigners are those who resist the enforcement of laws and health instructions; and *Victims*, the people who are deceived into following the conspiratorial plan or the ones who suffer due to the decisions of the authorities.

5 Voight-Kampff Generative AI Authorship Verification

Authorship verification is a fundamental task in author identification. All cases of questioned authorship can be decomposed into a series of verification instances, be it in a closed-set or open-set scenario [12]. Recent editions of PAN studied authorship verification from a *cross-domain* perspective [1, 2, 20] with very high validation rates in recent years [1, 2]. The latest two editions of this task studied the still challenging *cross-discourse type verification* setting [19]. Since PAN has been continuously organizing Authorship verification tasks since 2011, we are in a prime position to investigate a currently ubiquitous challenge of the highest societal importance: identifying and attributing the authorship of large language models in contrast to human authors.

Together with the ELOQUENT [10] lab on the evaluation of generative language model quality, we organize a collaborative shared task in the builder-breaker style. PAN's participants will *build* systems to detect the authorship of language models and distinguish them from human-authored texts. ELOQUENT's participants will attempt to *break* these systems by constructing evaluation datasets designed to challenge the discriminative capabilities of the PAN participants' systems.

In the builder task, participants will develop authorship verification models to attribute a text to a human or a large language model. The texts, as produced by the the breakers, may, for example, source a selection of human-authored texts and modify or re-recreate them from sections, or use style transfer or error injection to make the generated text more human-like.

When including machine-generated text in authorship identification, the most desirable and hardest problem formulation is generative AI detection, where a single document is disputed without reference (see Figure 1, Task 7). It is unclear if this detection task can be solved; it is an escalation of the standard verification setting where the authorship of two documents is decided. We hence have defined a range of problems with raising difficulty level, where the possibilities in the assignment space are in different ways constrained. In the "easiest" problem (see Figure 1, Task 1), two documents with unknown authorship are given and we guarantee that exactly one is generated by a human, $\boxed{\text{A}}$, and a machine, $\boxed{\text{M}}$, respectively. This constraint is relaxed for the other tasks where, for example, both texts may also stem from a machine, $\{\boxed{\text{M}}, \boxed{\text{M}}\}$. Note that an additional level of difficulty can be introduced by restricting the text lengths.

Input / Task		**Possible Assignment Patterns**
1. $\{\boxed{?}, \boxed{?}\}$		1. $\{\boxed{\text{A}}, \boxed{\text{M}}\}$
2. $\{\boxed{?}, \boxed{?}\}$		2. $\{\boxed{\text{A}}, \boxed{\text{M}}\}, \{\boxed{\text{A}}, \boxed{\text{A}}\}$
3. $\{\boxed{?}, \boxed{?}\}$	$\longrightarrow$	3. $\{\boxed{\text{A}}, \boxed{\text{M}}\}, \{\boxed{\text{M}}, \boxed{\text{M}}\}$
4. $\{\boxed{?}, \boxed{?}\}$		4. $\{\boxed{\text{A}}, \boxed{\text{M}}\}, \{\boxed{\text{A}}, \boxed{\text{A}}\}, \{\boxed{\text{M}}, \boxed{\text{M}}\}$
5. $\{\boxed{?}, \boxed{?}\}$		5. $\{\boxed{\text{A}}, \boxed{\text{M}}\}, \{\boxed{\text{A}}, \boxed{\text{A}}\}, \{\boxed{\text{A}}, \boxed{\text{B}}\}$
6. $\{\boxed{?}, \boxed{?}\}$		6. $\{\boxed{\text{A}}, \boxed{\text{M}}\}, \{\boxed{\text{A}}, \boxed{\text{A}}\}, \{\boxed{\text{A}}, \boxed{\text{B}}\}, \{\boxed{\text{M}}, \boxed{\text{M}}\}$
7. $\boxed{?}$		7. $\boxed{\text{A}}, \quad \boxed{\text{M}}$

Figure 1. Hierarchy of authorship verification problems from "easy" (1) to "hard" (7), involving LLM-generated text. The difficulty results from the possible assignment patterns that are allowed to occur. $\boxed{\text{M}}$ denotes LLM-generated text, while $\boxed{\text{A}}$ and $\boxed{\text{B}}$ denote human-authored text, where the same letter encodes the same human author.

Acknowledgments

The work of members of the Universitat Politècnica de València is in the framework of the XAI-DisInfodemics research project on eXplainable AI for disinformation and conspiracy detection during infodemics, funded by MCIN/AEI and by European Union NextGenerationEU/PRTR ((Grant PLEC2021-007681). The work of members of the Bauhaus-Universität Weimar and Leipzig University was partially supported by the European Commission under grant agreement GA 101070014 (https://openwebsearch.eu)

Bibliography

[1] Bevendorff, J., Chulvi, B., la Peña Sarracén, G.L.D., Kestemont, M., Manjavacas, E., Markov, I., Mayerl, M., Potthast, M., Rangel, F., Rosso, P., Stamatatos, E., Stein, B., Wiegmann, M., Wolska, M., Zangerle, E.: Overview of PAN 2021: Authorship verification, profiling hate speech spreaders on twitter, and style change detection. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 12th International Conference of the CLEF Association, vol. 12880, pp. 419–431 (2021)

[2] Bevendorff, J., Ghanem, B., Giachanou, A., Kestemont, M., Manjavacas, E., Markov, I., Mayerl, M., Potthast, M., Pardo, F.M.R., Rosso, P., Specht, G., Stamatatos, E., Stein, B., Wiegmann, M., Zangerle, E.: Overview of PAN 2020: Authorship verification, celebrity profiling, profiling fake news spreaders on twitter, and style change detection. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 11th International Conference of the CLEF Association, vol. 12260, pp. 372–383 (2020)

[3] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H.S., Roff, H., Allen, G.C., Steinhardt, J., Flynn, C., hÉigeartaigh, S.Ó., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., Amodei, D.: The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. CoRR **abs/1802.07228** (2018)

[4] Costa-jussà, M.R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., et al.: No language left behind: Scaling human-centered machine translation. arXiv e-prints pp. arXiv–2207 (2022)

[5] Dementieva, D., Logacheva, V., Nikishina, I., Fenogenova, A., Dale, D., Krotova, I., Semenov, N., Shavrina, T., Panchenko, A.: RUSSE-2022: Findings of the first Russian detoxification task based on parallel corpora. In: Computational Linguistics and Intellectual Technologies (2022)

[6] Douglas, K.M., Sutton, R.M.: What are conspiracy theories? a definitional approach to their correlates, consequences, and communication. Annual Review of Psychology **74**(1), 271–298 (2023), https://doi.org/10.1146/annurev-psych-032420-031329

[7] El-Alami, F.z., El Alaoui, S.O., Nahnahi, N.E.: A multilingual offensive language detection method based on transfer learning from transformer fine-tuning model. Journal of King Saud University-Computer and Information Sciences **34**(8), 6048–6056 (2022)

[8] Fröbe, M., Wiegmann, M., Kolyada, N., Grahm, B., Elstner, T., Loebe, F., Hagen, M., Stein, B., Potthast, M.: Continuous Integration for Reproducible Shared Tasks with TIRA.io. In: Kamps, J., Goeuriot, L., Crestani, F., Maistro, M., Joho, H., Davis, B., Gurrin, C., Kruschwitz, U., Caputo, A. (eds.) Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023), pp. 236–241, Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York (Apr 2023), https://doi.org/10.1007/978-3-031-28241-6_20

[9] Giachanou, A., Ghanem, B., Rosso, P.: Detection of conspiracy propagators using psycho-linguistic characteristics. Journal of Information Science **49**(1), 3–17 (2023), https://doi.org/https://doi.org/10.1177/0165551520985486

[10] Karlgren, J., Dürlich, L., Gogoulou, E., Guillou, L., Nivre, J., Talman, A.: ELOQUENT CLEF shared tasks for evaluation of generative language model

quality. In: Advances in Information Retrieval: 46th European Conference on IR Research (ECIR) (2024)
[11] Kestemont, M., Tschuggnall, M., Stamatatos, E., Daelemans, W., Specht, G., Stein, B., Potthast, M.: Overview of the author identification task at PAN 2018: Cross-domain authorship attribution and style change detection. In: CLEF 2018 Labs and Workshops, Notebook Papers (2018)
[12] Koppel, M., Winter, Y.: Determining if two documents are written by the same author. Journal of the Association for Information Science and Technology **65**(1), 178–187 (2014)
[13] Kumar, S., Balachandran, V., Njoo, L., Anastasopoulos, A., Tsvetkov, Y.: Language generation models can cause harm: So what can we do about it? an actionable survey. CoRR **abs/2210.07700** (2022)
[14] Lai, H., Toral, A., Nissim, M.: Multilingual pre-training with language and task adaptation for multilingual text style transfer. In: Muresan, S., Nakov, P., Villavicencio, A. (eds.) Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, pp. 262–271, Association for Computational Linguistics (2022)
[15] Logacheva, V., Dementieva, D., Krotova, I., Fenogenova, A., Nikishina, I., Shavrina, T., Panchenko, A.: A study on manual and automatic evaluation for text style transfer: The case of detoxification. In: Proceedings of the 2nd Workshop on Human Evaluation of NLP Systems (HumEval), pp. 90–101, Association for Computational Linguistics, Dublin, Ireland (May 2022)
[16] Logacheva, V., Dementieva, D., Ustyantsev, S., Moskovskiy, D., Dale, D., Krotova, I., Semenov, N., Panchenko, A.: ParaDetox: Detoxification with parallel data. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 6804–6818, Association for Computational Linguistics, Dublin, Ireland (May 2022)
[17] Moskovskiy, D., Dementieva, D., Panchenko, A.: Exploring cross-lingual text detoxification with large multilingual language models. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 346–354, Association for Computational Linguistics, Dublin, Ireland (May 2022)
[18] Rosso, P., Rangel, F., Potthast, M., Stamatatos, E., Tschuggnall, M., Stein, B.: Overview of PAN'16—New Challenges for Authorship Analysis: Cross-genre Profiling, Clustering, Diarization, and Obfuscation. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. 7th International Conference of the CLEF Initiative (CLEF 16) (2016)
[19] Stamatatos, E., Kestemont, M., Kredens, K., Pezik, P., Heini, A., Bevendorff, J., Stein, B., Potthast, M.: Overview of the Authorship Verification Task at PAN 2022. In: Faggioli, G., Ferro, N., Hanbury, A., Potthast, M. (eds.) CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org (Sep 2022)
[20] Stamatatos, E., Potthast, M., Pardo, F.M.R., Rosso, P., Stein, B.: Overview of the PAN/CLEF 2015 evaluation lab. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction, vol. 9283, pp. 518–538 (2015)
[21] Tang, Y., Tran, C., Li, X., Chen, P.J., Goyal, N., Chaudhary, V., Gu, J., Fan, A.: Multilingual translation with extensible multilingual pretraining and finetuning (2020)
[22] Tschuggnall, M., Stamatatos, E., Verhoeven, B., Daelemans, W., Specht, G., Stein, B., Potthast, M.: Overview of the author identification task at PAN 2017:

style breach detection and author clustering. In: CLEF 2017 Labs and Workshops, Notebook Papers (2017)
[23] Wadud, M.A.H., Mridha, M.F., Shin, J., Nur, K., Saha, A.K.: Deep-bert: Transfer learning for classifying multilingual offensive texts on social media. Comput. Syst. Sci. Eng. **44**(2), 1775–1791 (2023)
[24] Zangerle, E., Mayerl, M., , Potthast, M., Stein, B.: Overview of the Style Change Detection Task at PAN 2021. In: Faggioli, G., Ferro, N., Joly, A., Maistro, M., Piroi, F. (eds.) CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org (2021)
[25] Zangerle, E., Mayerl, M., Potthast, M., Stein, B.: Overview of the Style Change Detection Task at PAN 2022. In: Faggioli, G., Ferro, N., Hanbury, A., Potthast, M. (eds.) CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org (Sep 2022)
[26] Zangerle, E., Mayerl, M., Specht, G., Potthast, M., Stein, B.: Overview of the Style Change Detection Task at PAN 2020. In: CLEF 2020 Labs and Workshops, Notebook Papers (2020)
[27] Zangerle, E., Tschuggnall, M., Specht, G., Stein, B., Potthast, M.: Overview of the Style Change Detection Task at PAN 2019. In: CLEF 2019 Labs and Workshops, Notebook Papers (2019)