

# Overview of the Style Change Detection Task at PAN 2021

Eva Zangerle<sup>1</sup>, Maximilian Mayerl<sup>1</sup>, Martin Potthast<sup>2</sup> and Benno Stein<sup>3</sup>

<sup>1</sup>Universität Innsbruck

<sup>2</sup>Leipzig University

<sup>3</sup>Bauhaus-Universität Weimar

pan@webis.de    <http://pan.webis.de>

## Abstract

Style change detection means to identify text positions within a multi-author document at which the author changes. Detecting these positions is considered a key enabling technology for all tasks involving multi-author documents as well as a preliminary step for reliable authorship identification. In this year's PAN style change detection task, we asked the participants to answer the following questions: (1) Given a document, was it written by a single or by multiple authors? (2) For each pair of consecutive paragraphs in a given document, is there a style change between these paragraphs? (3) Find all positions of writing style change, i.e., assign all paragraphs of a text uniquely to some author, given the list of authors assumed for the multi-author document. The outlined task is performed and evaluated on a dataset that has been compiled from an English Q&A platform. The style change detection task, the underlying dataset, a survey of the participants' approaches, as well as the results are presented in this paper.

## 1. Introduction


Style change detection is a multi-author writing style analysis to determine for a given document both the number of authors and the positions of authorship changes. Previous editions of PAN featured already multi-author writing style analysis tasks: in 2016, participants were asked to identify and cluster text segments by author [1]. In 2017, the task was two-fold, namely, to detect whether a given document was written by multiple authors, and, if this was the case, to identify the positions at which authorship changes [2]. At PAN 2018, the task was relaxed to a binary classification task that aimed at distinguishing between single- and multi-author documents [3]. The PAN edition in 2019 broadened the task and additionally asked participants to predict the number of authors for all detected multi-author documents [4]. In 2020 the participants were asked to detect whether a document was written by a single or by multiple authors, and to determine the positions of style changes at the paragraph level. This year we asked participants (1) to find out whether the text is written by a single author or by multiple authors, (2) to detect the position of the changes on the paragraph-level, and (3) to assign all paragraphs of the text uniquely to some author out of the number of authors they assume for the multi-author document.

---

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

The remainder of this paper is structured as follows. Section 2 discusses previous approaches to style change detection. Section 3 introduces the PAN 2021 style change detection task, the underlying dataset, and the evaluation procedure. Section 4 summarizes the received submissions, and Section 5 analyzes and compares the achieved results.

## 2. Related Work

Style change detection is related to problems from the fields of stylometry, intrinsic plagiarism detection, and authorship attribution. Solutions typically create stylistic fingerprints of authors, which may rely on lexical features such as character  $n$ -grams [5, 6], or word frequencies [7], syntactic features such as part-of-speech tags [8], or structural features such as the use of indentation [9]. By computing such fingerprints on the sentence- or paragraph-level, style changes at the respective boundaries can be detected by computing pairwise similarities [10, 11], clustering [1, 12], or by applying outlier detection [13]. Recently, also deep learning models have been employed for these tasks [14, 15, 16, 17].

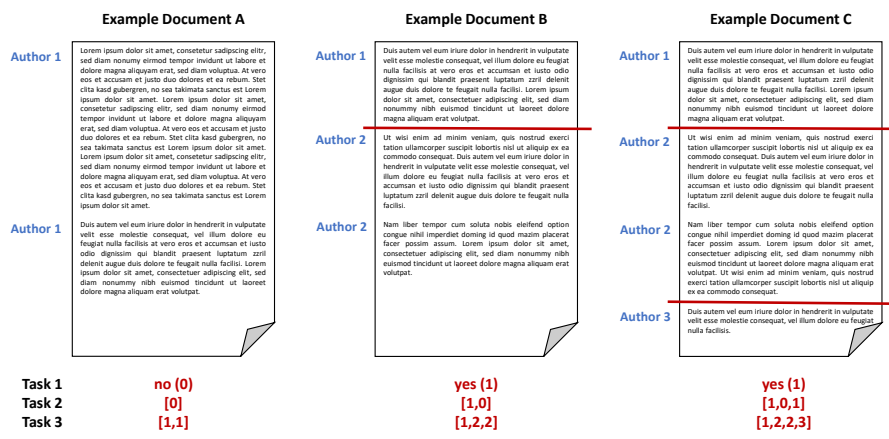
One of the first works on identifying inconsistencies of writing style was presented by Glover and Hirst [18]. Notably, Stamatatos [19] utilized  $n$ -grams to create stylistic fingerprints for quantifying variations in writing style. The task of intrinsic plagiarism detection was first tackled by Meyer zu Eißén and Stein [20, 21, 22]. Koppel et al. [23, 24] and Akiva and Koppel [25, 26] proposed to use lexical features as input for clustering methods to decompose documents into authorial threads. Tschuggnall et al. [27] proposed an unsupervised decomposition approach based on grammar tree representations. Gianella [28] utilizes Bayesian modeling to split a document into segments, followed by a clustering approach to cluster segments by author. Dauber et al. [29] presented an approach to tackle authorship attribution on multi-author documents based on multi-label classification on linguistic features. Aldebei et al. [30] and Sarwar et al. [31] used hidden Markov models and basic stylometric features to build a so-called co-authorship graph. Rexha et al. [32] predicted the number of authors of a text using stylistic features.

## 3. Style Change Detection Task

This section details the style change detection task, the dataset constructed for the task, and the employed performance measures.

### 3.1. Task Definition

Goal of the style change detection task is to identify text positions within a given multi-author document at which the author switches, and to assign each paragraph to an author. In a first step we suggest to check the document in question for writing style changes; the result is then used as predictor for single- or multi-authorship. If a document is considered a multi-author document, the exact positions at which the writing style (and probably the authorship changes) are to be determined, and, finally, paragraphs may be assigned to their alleged author.



**Figure 1:** Documents that illustrate different style change situations and the expected solution for Task 1 (single vs. multiple), Task 2 (change positions), and Task 3 (author attribution).

Given a document, we ask participants to answer the following three questions:

- *Single vs. Multiple.* Given a text, determine whether the text is written by a single author or by multiple authors (Task 1).
- *Change Positions.* Given a text, determine all positions within that text where the writing style changes (Task 2). For this task, such changes can only occur between paragraphs.
- *Author Attribution.* Given a text, assign all its paragraphs to some author out of the set of authors participants assume for the given text (Task 3).

Figure 1 shows documents and the results of the three tasks for these documents. Document A is written by a single author and does not contain any style changes. Document B contains a single style change between the Paragraphs 1 and 2, and Document C contains two style changes. As indicated in Figure 1, Task 1 is a binary classification task determining whether the document was written by multiple authors. For Task 2 we ask participants to provide a binary value indicating whether there is a change in authorship between each pair of consecutive paragraphs for each document. For Task 3 we ask participants to assign each paragraph uniquely to an author from a list of authors in question.

All documents are written in English and consist of paragraphs each of which written by a single author out of a set of four authors. A document can contain a number of style changes between paragraphs but no style changes within a paragraph.

We asked participants to deploy their software on the TIRA platform [33]. This allows participants to test their software on the available training and validation dataset, as well as to self-evaluate their software on the unseen dataset. TIRA enables blind evaluation, thus foreclosing optimization against the test data.

**Table 1**

Parameters for constructing the style change detection dataset.

| Parameter                       | Configurations     |
|---------------------------------|--------------------|
| Number of collaborating authors | 1-4                |
| Document length                 | 1,000-10,000       |
| Minimum paragraph length        | 100                |
| Minimum number of paragraphs    | 2                  |
| Change positions                | between paragraphs |
| Document language               | English            |

### 3.2. Dataset Construction

The dataset for the Style Change Detection task was created from posts taken from the popular StackExchange network of Q&A sites. This ensures that results are comparable with past editions of the tasks, which rely on the same data source [4, 34]. In the following, we outline the dataset creation process.

The dataset for this year’s task consists of 16,000 documents. The text were drawn from a dump of questions and answers from various sites in the StackExchange network. To ensure topical coherence of the dataset, the considered sites revolve around topics focusing on technology.<sup>1</sup> We cleaned all questions and answers from these sites by removing questions and answers that were edited after they were originally submitted, and by removing images, URLs, code snippets, block quotes as well as bullet lists. Afterward, the questions and answers were split into paragraphs, dropping all paragraphs with fewer than 100 characters. Since one of the goals for this year’s edition of the task was to reduce the impact of topic changes within a document, which could inadvertently make the task easier, we constructed documents from these paragraphs by only taking paragraphs belonging to the same question/answer thread within a single document: we randomly chose a question/answer thread and also randomly chose a number  $n \in \{1, 2, 3, 4\}$ , denoting how many authors the resulting document should have. Following that, we took a random subset of size  $n$  of all the authors that contributed to the chosen question/answer thread that we wanted to draw paragraphs from. We took all paragraphs written by this subset of authors, shuffled them, and concatenated them together to form a document. If a generated document consisted of one paragraph only, or if it was fewer than 1,000 or more than 10,000 characters long, it was discarded.

We ensured that the number of authors was equally distributed across the documents—i.e., there are as many single-author documents as documents with two authors, three authors, and four authors. We split the resulting set of documents into a training set, a test set, and a validation set. The training set consists of 70% of all documents (11,200), and the test set and the validation set consist of 15% of all documents each (2,400). The parameters used for creating the dataset are given in Table 1, and an overview of the three dataset splits can be seen in Table 2.

---

<sup>1</sup>Code Review, Computer Graphics, CS Educators, CS Theory, Data Science, DBA, DevOps, GameDev, Network Engineering, Raspberry Pi, Superuser, and Server Fault.

**Table 2**

Dataset overview. Text length is measured as average number of tokens per document.

| Dataset | #Docs  | Documents / #Authors |              |              |              | Length / #Authors |       |       |       |
|---------|--------|----------------------|--------------|--------------|--------------|-------------------|-------|-------|-------|
|         |        | 1                    | 2            | 3            | 4            | 1                 | 2     | 3     | 4     |
| Train   | 11,200 | 2,800<br>25%         | 2,800<br>25% | 2,800<br>25% | 2,800<br>25% | 1,519             | 1,592 | 1,795 | 2,059 |
| Valid.  | 2,400  | 600<br>25%           | 600<br>25%   | 600<br>25%   | 600<br>25%   | 1,549             | 1,599 | 1,785 | 2,039 |
| Test    | 2,400  | 600<br>25%           | 600<br>25%   | 600<br>25%   | 600<br>25%   | 1,512             | 1,564 | 1,793 | 2,081 |

### 3.3. Performance Measures

To evaluate the submitted approaches and to compare the obtained results, submissions are evaluated by the  $F_\alpha$ -Measure for each document, where  $\alpha = 1$  equally weighs the harmonic mean between precision and recall. Across all documents, we compute the macro-averaged  $F_\alpha$ -Measure. The three tasks are evaluated independently based on the obtained accuracy measures.

## 4. Survey of Submissions

For the 2021 edition of the style change detection task we received five submissions, which are described in the following.

### 4.1. Style Change Detection on Real-World Data using LSTM-powered Attribution Algorithm

Deibel and Löfflad [35] propose the use of multi-layer perceptrons and bidirectional LSTMs for the style change detection task. The approach relies on textual features widely used in authorship attribution (mean sentence length in words, mean word length, or corrected type-token ratio) and pretrained fastText word embeddings. For Task 1, the approach uses a multi-layer perceptron, with three hidden, fully connected feed forward layers with per-document embeddings as input. For Task 2, the authors employ a two-layered bidirectional LSTM. Based on the style change predictions for Task 1 and Task 2, the approach iterates for Task 3 over all pairs of paragraphs to attribute each paragraph to an author. If no style change is detected between paragraphs, the current paragraph is attributed to the author of the previous paragraph. For an alleged style change between paragraphs the current paragraph is compared to all previously attributed paragraphs in order to either assign it to an already known author or to attribute it to a new author.

## 4.2. Style Change Detection using Siamese Neural Networks

The approach proposed by Nath [36] utilizes Siamese neural networks to compute paragraph similarities for the detection of style changes. Paragraphs are transformed into numerical vectors by lowercasing, removing all punctuation, tokenizing each paragraph and then, representing each vocabulary word as an integer id. For the pairwise similarity comparison of paragraphs the vector representation of the two paragraphs and the label (style change or not) are used as input. The Siamese network features a GloVe embedding layer, a bidirectional LSTM layer, distance measure layer, and a dense layer with sigmoid activation to compute the actual final label.

## 4.3. Writing Style Change Detection on Multi-Author Documents

The approach by Singh et al. [37] is based on an approach for authorship verification submitted to PAN 2020 by Weerasinghe et al. [38]. The core of the approach hence is an authorship verification model which the authors use to determine whether two given paragraphs are written by the same author. In this regard they extract features for both paragraphs, including tf-idf features, n-grams of part of speech tags, and vocabulary richness measures among others. Then, the difference between the feature vectors for both paragraphs and take the magnitude of the resulting difference vector is computed. This magnitude is fed into a logistic regression classifier to determine whether both paragraphs have the same author. They then use this model to answer the three tasks posed in this year's style change detection task as follows. For Task 1, they use their verification model to predict whether all consecutive paragraphs in the document were written by the same author. If the average of the classifier scores for all consecutive paragraphs in a document is greater than 0.5, the document is classified as multi-author document. For Task 2, the author again use their verification model on each consecutive pair of paragraphs, and predict a style change between all paragraphs for which the model determines that they were not written by the same author. Finally, for Task 3, they ran their verification model on all pairs of paragraphs in a document, and used hierarchical clustering on a distance matrix created from classifier scores to group paragraphs written by the same author together.

## 4.4. Multi-label Style Change Detection by Solving a Binary Classification Problem

The approach developed by Strøm [39] is based on BERT embedding features and stylistic features previously proposed by Zlatkova et al. [40]. The embeddings are generated on a sentence-level and subsequently, sentence embeddings are aggregated to the paragraph-level by adding the sentence embeddings of each paragraph. Text features are extracted on the paragraph-level. To identify style changes between two paragraphs to solve tasks 1 and 2, binary classification via a stacking ensemble is performed. This ensemble uses a meta-learner trained on the predictions computed by base level classifiers for stylistic and embedding features. For the multi-label classification for Task 3, the author proposes a recursive strategy that is based on the predictions for Task 1 and Task 2. The algorithm iterates over all paragraphs, and computes the probability that each pair of paragraphs was written by the same author. If

this probability exceeds the threshold of 0.5, the paragraphs are attributed to the same author; otherwise to different authors.

#### 4.5. Using Single BERT For Three Tasks Of Style Change Detection

Zhang et al. [41] rely on a pretrained BERT model (specifically, BERT-Base as provided by Google). They model Task 3 as a binary classification task. Therefore, for each paragraph and each of its preceding paragraphs, they compute whether there is a style change to augment the amount of training data. These labels are used for fine-tuning the BERT model. The resulting weights are then saved and used for the actual predictions for the tasks 1–3. Labels for Task 2 and Task 3 are predicted, and the results for Task 1 are inferred from the results of Task 2.

### 5. Evaluation Results

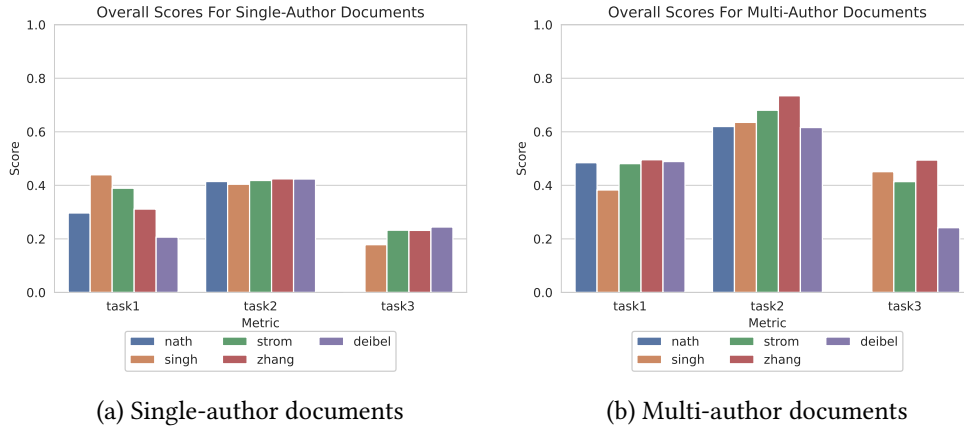
Table 3 shows the evaluation results of all submitted approaches as well as a baseline in form of  $F_1$  scores. The baseline approach uses a uniformly random prediction for Task 3, and infers the results for Tasks 1 and 2 from the predictions for Task 3. The predictions for Task 3 take into account that authors must be labeled with increasing author identifiers. As can be seen, all approaches significantly outperform the baseline on all tasks, except for the approach by Deibel et al. [35] for Task 3, which scores lower than the baseline. The best performance for Task 1—determining whether a document has one or multiple authors—was achieved by Strøm [39], whereas the best performance for the actual style change detection tasks, Task 2 and Task 3, was achieved by Zhang et al. [41]. In all cases, the best performing approach substantially outperforms all other submitted approaches.

**Table 3**

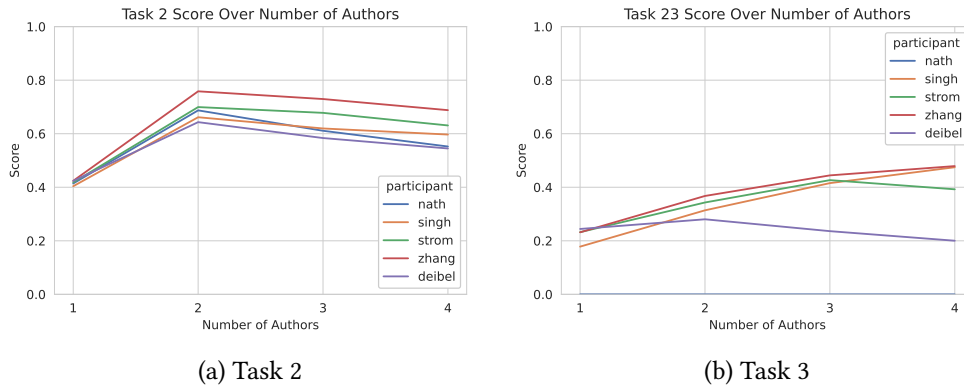
Overall results for the style change detection task, ranked by average performance across all three tasks.

| Participant   | Task1 $F_1$  | Task2 $F_1$  | Task3 $F_1$  |
|---------------|--------------|--------------|--------------|
| Zhang et al.  | 0.753        | <b>0.751</b> | <b>0.501</b> |
| Strøm         | <b>0.795</b> | 0.707        | 0.424        |
| Singh et al.  | 0.634        | 0.657        | 0.432        |
| Deibel et al. | 0.621        | 0.669        | 0.263        |
| Nath          | 0.704        | 0.647        | —            |
| Baseline      | 0.457        | 0.470        | 0.329        |

In addition to the overall evaluation given in Table 3, we further analyzed the performance of all submitted approaches separately for single-author and multi-author documents. The results for this analysis are given in Figure 2. There are a number of observations we can make from those results. For Task 1, the approach submitted by Singh et al. has the best performance out of all approaches for single-author documents, but the worst performance for multi-author documents. Looking at the results for Task 2, we can see that all approaches show almost the same performance for single-author documents. This means that the difference in overall performance between those approaches stems only from multi-author documents. A similar observation, though not quite as pronounced, can be made for Task 3.



**Figure 2:** Scores (F<sub>1</sub>) for all tasks separately for single-author and multi-author documents.



**Figure 3:** Scores (F<sub>1</sub>) for all Task 2 and Task 3, depending on the true number of authors in a document.

Finally, we looked at how the performance of the submitted approaches changes depending on the true number of authors per document. We performed this analysis for Task 2 and Task 3. The results can be seen in Figure 3. Looking at the results, we can see that the performance for Task 2 peaks at two authors for all approaches. In other words, all submitted approaches are best at determining style changes between paragraphs when the document was written by two authors. A different picture presents itself for Task 3. For two of the submitted approaches (Zhang et al. and Singh et al.), the performance keeps increasing with a growing number of authors. They perform best if the document was written by four authors. This suggests it may be interesting to increase the maximum number of authors per document for a future edition of the task.



## 6. Conclusion

For the style change detection task at PAN 2021, we asked participants to determine (1) whether a document was in fact written by several authors, (2) style changes between consecutive paragraphs (3) the most likely author for a paragraph. Altogether five participants submitted their approaches. For Task 1, the best performing approach relies on BERT embeddings and stylistic features, utilizing a stacking ensemble. For Task 2 and Task 3, the highest  $F_\alpha$ -Measure was obtained by fine-tuning pretrained BERT embeddings based on augmented data gained from permuting the paragraphs of each document.

## References

- [1] E. Stamatatos, M. Tschuggnall, B. Verhoeven, W. Daelemans, G. Specht, B. Stein, M. Potthast, Clustering by Authorship Within and Across Documents, in: Working Notes Papers of the CLEF 2016 Evaluation Labs, CEUR Workshop Proceedings, CLEF and CEUR-WS.org, 2016. URL: <http://ceur-ws.org/Vol-1609/>.
- [2] M. Tschuggnall, E. Stamatatos, B. Verhoeven, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the Author Identification Task at PAN 2017: Style Breach Detection and Author Clustering, in: L. Cappellato, N. Ferro, L. Goeuriot, T. Mandl (Eds.), Working Notes Papers of the CLEF 2017 Evaluation Labs, volume 1866 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017. URL: <http://ceur-ws.org/Vol-1866/>.
- [3] M. Kestemont, M. Tschuggnall, E. Stamatatos, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the Author Identification Task at PAN-2018: Cross-domain Authorship Attribution and Style Change Detection, in: L. Cappellato, N. Ferro, J.-Y. Nie, L. Soulier (Eds.), Working Notes Papers of the CLEF 2018 Evaluation Labs, volume 2125 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018. URL: <http://ceur-ws.org/Vol-2125/>.
- [4] E. Zangerle, M. Tschuggnall, G. Specht, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2019, in: L. Cappellato, N. Ferro, D. Losada, H. Müller (Eds.), CLEF 2019 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2019. URL: <http://ceur-ws.org/Vol-2380/>.
- [5] E. Stamatatos, Intrinsic Plagiarism Detection Using Character n-gram Profiles, in: Notebook Papers of the 5th Evaluation Lab on Uncovering Plagiarism, Authorship and Social Software Misuse (PAN), Amsterdam, The Netherlands, 2011.
- [6] M. Koppel, J. Schler, S. Argamon, Computational methods in authorship attribution, *Journal of the American Society for Information Science and Technology* 60 (2009) 9–26.
- [7] D. I. Holmes, The Evolution of Stylometry in Humanities Scholarship, *Literary and Linguistic Computing* 13 (1998) 111–117.
- [8] M. Tschuggnall, G. Specht, Countering Plagiarism by Exposing Irregularities in Authors' Grammar, in: Proceedings of the European Intelligence and Security Informatics Conference (EISIC), IEEE, Uppsala, Sweden, 2013, pp. 15–22.
- [9] R. Zheng, J. Li, H. Chen, Z. Huang, A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques, *Journal of the American Society for Information Science and Technology* 57 (2006) 378–393.

- [10] J. Khan, Style Breach Detection: An Unsupervised Detection Model—Notebook for PAN at CLEF 2017, in: L. Cappellato, N. Ferro, L. Goeuriot, T. Mandl (Eds.), CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, CEUR-WS.org, 2017. URL: <http://ceur-ws.org/Vol-1866/>.
- [11] D. Karaś, M. Śpiewak, P. Sobiecki, OPI-JSA at CLEF 2017: Author Clustering and Style Breach Detection—Notebook for PAN at CLEF 2017, in: L. Cappellato, N. Ferro, L. Goeuriot, T. Mandl (Eds.), CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, CEUR-WS.org, 2017. URL: <http://ceur-ws.org/Vol-1866/>.
- [12] S. Nath, UniNE at PAN-CLEF 2019: Style Change Detection by Threshold Based and Window Merge Clustering Methods, in: L. Cappellato, N. Ferro, D. Losada, H. Müller (Eds.), CLEF 2019 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2019.
- [13] K. Safin, R. Kuznetsova, Style Breach Detection with Neural Sentence Embeddings—Notebook for PAN at CLEF 2017, in: L. Cappellato, N. Ferro, L. Goeuriot, T. Mandl (Eds.), CLEF 2017 Evaluation Labs and Workshop – Working Notes Papers, CEUR-WS.org, 2017. URL: <http://ceur-ws.org/Vol-1866/>.
- [14] N. Graham, G. Hirst, B. Marthi, Segmenting Documents by Stylistic Character, *Natural Language Engineering* 11 (2005) 397–415. URL: <https://doi.org/10.1017/S1351324905003694>. doi:10.1017/S1351324905003694.
- [15] A. Iyer, S. Vosoughi, Style Change Detection Using BERT, in: L. Cappellato, N. Ferro, A. Névóel, C. Eickhoff (Eds.), CLEF 2020 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2020.
- [16] M. Hosseinia, A. Mukherjee, A Parallel Hierarchical Attention Network for Style Change Detection—Notebook for PAN at CLEF 2018, in: L. Cappellato, N. Ferro, J.-Y. Nie, L. Soulier (Eds.), CLEF 2018 Evaluation Labs and Workshop – Working Notes Papers, CEUR-WS.org, 2018.
- [17] C. Zuo, Y. Zhao, R. Banerjee, Style Change Detection with Feedforward Neural Networks Notebook for PAN at CLEF 2019, in: L. Cappellato, N. Ferro, D. Losada, H. Müller (Eds.), CLEF 2019 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2019.
- [18] A. Glover, G. Hirst, *Detecting Stylistic Inconsistencies in Collaborative Writing*, Springer London, London, 1996, pp. 147–168. doi:10.1007/978-1-4471-1482-6\_12.
- [19] E. Stamatatos, Intrinsic Plagiarism Detection Using Character  $n$ -gram Profiles, in: B. Stein, P. Rosso, E. Stamatatos, M. Koppel, E. Agirre (Eds.), SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09), Universidad Politécnica de Valencia and CEUR-WS.org, 2009, pp. 38–46. URL: <http://ceur-ws.org/Vol-502>.
- [20] S. Meyer zu Eißén, B. Stein, Intrinsic Plagiarism Detection, in: M. Lalmas, A. MacFarlane, S. Rüger, A. Tombros, T. Tsikrika, A. Yavlinsky (Eds.), *Advances in Information Retrieval. 28th European Conference on IR Research (ECIR 2006)*, volume 3936 of *Lecture Notes in Computer Science*, Springer, Berlin Heidelberg New York, 2006, pp. 565–569. doi:10.1007/11735106\_66.
- [21] B. Stein, S. Meyer zu Eißén, Intrinsic Plagiarism Analysis with Meta Learning, in: B. Stein, M. Koppel, E. Stamatatos (Eds.), *1st Workshop on Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection (PAN 2007) at SIGIR*, 2007, pp. 45–50. URL: <http://ceur-ws.org/Vol-276>.

- [22] B. Stein, N. Lipka, P. Prettenhofer, Intrinsic Plagiarism Analysis, *Language Resources and Evaluation (LRE)* 45 (2011) 63–82. doi:10.1007/s10579-010-9115-y.
- [23] M. Koppel, N. Akiva, I. Dershowitz, N. Dershowitz, Unsupervised decomposition of a document into authorial components, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Portland, Oregon, USA, 2011, pp. 1356–1364. URL: <https://www.aclweb.org/anthology/P11-1136>.
- [24] M. Koppel, N. Akiva, I. Dershowitz, N. Dershowitz, Unsupervised decomposition of a document into authorial components, in: D. Lin, Y. Matsumoto, R. Mihalcea (Eds.), *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA, The Association for Computer Linguistics, 2011*, pp. 1356–1364. URL: <http://www.aclweb.org/anthology/P11-1136>.
- [25] N. Akiva, M. Koppel, Identifying Distinct Components of a Multi-author Document, in: N. Memon, D. Zeng (Eds.), *2012 European Intelligence and Security Informatics Conference, EISIC 2012*, IEEE Computer Society, 2012, pp. 205–209. URL: <https://doi.org/10.1109/EISIC.2012.16>. doi:10.1109/EISIC.2012.16.
- [26] N. Akiva, M. Koppel, A Generic Unsupervised Method for Decomposing Multi-Author Documents, *JASIST* 64 (2013) 2256–2264. URL: <https://doi.org/10.1002/asi.22924>. doi:10.1002/asi.22924.
- [27] M. Tschuggnall, G. Specht, Automatic decomposition of multi-author documents using grammar analysis, in: F. Klan, G. Specht, H. Gamper (Eds.), *Proceedings of the 26th GI-Workshop Grundlagen von Datenbanken*, volume 1313 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2014, pp. 17–22. URL: [http://ceur-ws.org/Vol-1313/paper\\_4.pdf](http://ceur-ws.org/Vol-1313/paper_4.pdf).
- [28] C. Giannella, An Improved Algorithm for Unsupervised Decomposition of a Multi-Author Document, *JASIST* 67 (2016) 400–411. URL: <https://doi.org/10.1002/asi.23375>. doi:10.1002/asi.23375.
- [29] E. Dauber, R. Overdorf, R. Greenstadt, Stylometric Authorship Attribution of Collaborative Documents, in: S. Dolev, S. Lodha (Eds.), *Cyber Security Cryptography and Machine Learning - First International Conference, CSCML 2017, Proceedings*, volume 10332 of *Lecture Notes in Computer Science*, Springer, 2017, pp. 115–135. URL: [https://doi.org/10.1007/978-3-319-60080-2\\_9](https://doi.org/10.1007/978-3-319-60080-2_9). doi:10.1007/978-3-319-60080-2\_9.
- [30] K. Aldebei, X. He, W. Jia, W. Yeh, SUDMAD: Sequential and Unsupervised Decomposition of a Multi-Author Document Based on a Hidden Markov Model, *JASIST* 69 (2018) 201–214. URL: <https://doi.org/10.1002/asi.23956>. doi:10.1002/asi.23956.
- [31] R. Sarwar, C. Yu, S. Nutanong, N. Uraiertprasert, N. Vannaboot, T. Rakthanmanon, A scalable framework for stylometric analysis of multi-author documents, in: J. Pei, Y. Manolopoulos, S. W. Sadiq, J. Li (Eds.), *Database Systems for Advanced Applications - 23rd International Conference, DASFAA 2018, Proceedings, Part I*, volume 10827 of *Lecture Notes in Computer Science*, Springer, 2018, pp. 813–829. doi:10.1007/978-3-319-91452-7\_52.
- [32] A. Rexha, S. Klampfl, M. Kröll, R. Kern, Towards a more fine grained analysis of scientific authorship: Predicting the number of authors using stylometric features, in: P. Mayr, I. Frommholz, G. Cabanac (Eds.), *Proceedings of the Third Workshop on Bibliometric-*

- enhanced Information Retrieval co-located with the 38th European Conference on Information Retrieval (ECIR 2016), volume 1567 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2016, pp. 26–31. URL: <http://ceur-ws.org/Vol-1567/paper3.pdf>.
- [33] M. Potthast, T. Gollub, M. Wiegmann, B. Stein, TIRA Integrated Research Architecture, in: N. Ferro, C. Peters (Eds.), *Information Retrieval Evaluation in a Changing World, The Information Retrieval Series*, Springer, Berlin Heidelberg New York, 2019. doi:10.1007/978-3-030-22948-1\_5.
- [34] E. Zangerle, M. Mayerl, G. Specht, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2020, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéal (Eds.), *CLEF 2020 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2020. URL: <http://ceur-ws.org/Vol-2696/>.
- [35] R. Deibel, D. Löfflad, Style Change Detection on Real-World Data using LSTM-powered Attribution Algorithm, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), *CLEF 2021 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2021.
- [36] S. Nath, Style Change Detection using Siamese Neural Networks, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), *CLEF 2021 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2021.
- [37] R. Singh, J. Weerasinghe, R. Greenstadt, Writing Style Change Detection on Multi-Author Documents, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), *CLEF 2021 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2021.
- [38] J. Weerasinghe, R. Greenstadt, Feature Vector Difference based Neural Network and Logistic Regression Models for Authorship Verification—Notebook for PAN at CLEF 2020, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéal (Eds.), *CLEF 2020 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2020. URL: <http://ceur-ws.org/Vol-2696/>.
- [39] E. Strøm, Multi-label Style Change Detection by Solving a Binary Classification Problem, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), *CLEF 2021 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2021.
- [40] D. Zlatkova, D. Kopev, K. Mitov, A. Atanasov, M. Hardalov, I. Koychev, P. Nakov, An Ensemble-Rich Multi-Aspect Approach for Robust Style Change Detection, in: L. Cappellato, N. Ferro, J.-Y. Nie, L. Soulier (Eds.), *CLEF 2018 Evaluation Labs and Workshop – Working Notes Papers*, CEUR-WS.org, 2018. URL: <http://ceur-ws.org/Vol-2125/>.
- [41] Z. Zhang, X. Miao, Z. Peng, J. Zeng, H. Cao, J. Zhang, Z. Xiao, X. Peng, Z. Chen, Using Single BERT For Three Tasks Of Style Change Detection, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), *CLEF 2021 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2021.