Overview of the 3rd International Competition on Plagiarism Detection

Martin Potthast[1], Andreas Eiselt[1], Alberto Barrón-Cedeño[2],
Benno Stein[1], and Paolo Rosso[2]

[1]Web Technology & Information Systems
Bauhaus-Universiät Weimar, Germany

[2]Natural Language Engineering Lab, ELiRF
Universidad Politécnica de Valencia, Spain

pan@webis.de http://pan.webis.de

Abstract This paper overviews eleven plagiarism detectors that have been developed and evaluated within PAN'11. We survey the detection approaches developed for the two sub-tasks "external plagiarism detection" and "intrinsic plagiarism detection," and we report on their detailed evaluation based on the third revised edition of the PAN plagiarism corpus PAN-PC-11.

1 Introduction

Copying another author's text and claiming its authorship is called plagiarism. While research on automatic plagiarism detection has been conducted for decades, the standardized evaluation of plagiarism detection algorithms has a short history [13]. In this regard we have organized three competitions on plagiarism detection, the latest one held in conjunction with the 2011 CLEF conference. This paper overviews the submitted detectors and evaluates their performances.

1.1 Plagiarism Detection

Let $s = \langle s_{\text{plg}}, d_{\text{plg}}, s_{\text{src}}, d_{\text{src}} \rangle$ denote a plagiarism case where s_{plg} is a passage of document d_{plg} and a plagiarized version of some source passage s_{src} in d_{src}. Given d_{plg}, the task of a plagiarism detector is to detect s by reporting a corresponding plagiarism detection $r = \langle r_{\text{plg}}, d_{\text{plg}}, r_{\text{src}}, d'_{\text{src}} \rangle$. We say that r detects s iff $s_{\text{plg}} \cap r_{\text{plg}} \neq \emptyset$, $s_{\text{src}} \cap r_{\text{src}} \neq \emptyset$, and $d_{\text{src}} = d'_{\text{src}}$. This task can be tackled with external plagiarism detection as well as with intrinsic plagiarism detection.

Algorithms for external plagiarism detection attempt to detect s by retrieving d_{src} from a document collection D (e.g., the web) and by extracting s_{src} and s_{plg} from d_{src} and d_{plg} based on a detailed comparison of the two documents. Algorithms for intrinsic plagiarism detection attempt to detect s by analyzing the writing style of d_{plg}, whereas significant style variations from one passage to another may indicate that s_{plg} has been written by a different author than the rest of d_{plg}.

Table 1. Corpus statistics for 26 939 documents and 61 064 plagiarism cases in the PAN-PC-11.

Document Statistics					
Document Purpose		Plagiarism per Document		Document Length	
source documents	50%	hardly (5%-20%)	57%	short (1-10 pp.)	50%
suspicious documents		medium (20%-50%)	15%	medium (10-100 pp.)	35%
– with plagiarism	25%	much (50%-80%)	18%	long (100-1000 pp.)	15%
– without plagiarism	25%	entirely (>80%)	10%		

Plagiarism Case Statistics			
Obfuscation		Case Length	
none	18%	short (<150 words)	35%
paraphrasing		medium (150-1150 words)	38%
– automatic (low)	32%	long (>1150 words)	27%
– automatic (high)	31%		
– manual	8%		
translation ({de, es} to en)			
– automatic	10%		
– automatic + manual correction	1%		

1.2 Evaluating Plagiarism Detectors

To evaluate plagiarism detectors we have developed an evaluation framework consisting of the PAN plagiarism corpus 2011 (PAN-PC-11) and detection performance measures [13]. The framework was already employed in the 1st and 2nd competition on plagiarism detection, and the corpus has been revised for this year. Table 1 gives an overview of important corpus parameters. Notable changes compared to previous versions of the corpus include the significantly larger portion of plagiarism that is obfuscated by paraphrasing or translation, and the addition of manually translated plagiarism. These changes are based on insights gained from last year's competition, namely, that verbatim plagiarism is detected without problems and that automatically translated plagiarism is detected too easily.

Let S denote the set of plagiarism cases in the corpus, and let R denote the set of detections reported by a plagiarism detector for the suspicious documents. To simplify the notation, a plagiarism case $s = \langle s_{\text{plg}}, d_{\text{plg}}, s_{\text{src}}, d_{\text{src}} \rangle$, $s \in S$, is represented as a set $\mathbf{s}$ of references to the characters of d_{plg} and d_{src}, forming the passages s_{plg} and s_{src}. Likewise, a plagiarism detection $r \in R$ is represented as $\mathbf{r}$. Based on this notation, precision and recall of R under S can be measured as follows [13]:

$$prec(S, R) = \frac{1}{|R|} \sum_{r \in R} \frac{\left| \bigcup_{s \in S} (\mathbf{s} \sqcap \mathbf{r}) \right|}{|\mathbf{r}|}, \qquad rec(S, R) = \frac{1}{|S|} \sum_{s \in S} \frac{\left| \bigcup_{r \in R} (\mathbf{s} \sqcap \mathbf{r}) \right|}{|\mathbf{s}|},$$

$$\text{where} \quad \mathbf{s} \sqcap \mathbf{r} = \begin{cases} \mathbf{s} \cap \mathbf{r} & \text{if } r \text{ detects } s, \\ \emptyset & \text{otherwise.} \end{cases}$$

Observe that neither precision nor recall account for the fact that plagiarism detectors sometimes report overlapping or multiple detections for a single plagiarism case. This is undesirable, and to address this deficit also a detector's granularity is quantified

as follows:

$$gran(S, R) = \frac{1}{|S_R|} \sum_{s \in S_R} |R_s|,$$

where $S_R \subseteq S$ are cases detected by detections in R, and $R_s \subseteq R$ are detections of s. I.e., $S_R = \{s \mid s \in S \wedge \exists r \in R : r \text{ detects } s\}$ and $R_s = \{r \mid r \in R \wedge r \text{ detects } s\}$. The above measures have been computed for every plagiarism detector that took part in PAN'11; however, they do not allow for an absolute ranking among them. Therefore, the three measures are combined into a single overall score as follows:

$$plagdet(S, R) = \frac{F_1}{\log_2(1 + gran(S, R))},$$

where F_1 is the equally weighted harmonic mean of precision and recall.

2 Intrinsic Plagiarism Detection

This section surveys the intrinsic plagiarism detectors and evaluates their performances.

2.1 Survey of Detection Approaches

Intrinsic plagiarism detection has attracted renewed interest in PAN'11. An analysis of the submitted notebooks reveals a generic set of building blocks all of which employ a chunking strategy, a writing style retrieval model, and an outlier detection algorithm; however, the specifics differ significantly. In all cases, the mentioned building blocks are arranged within a retrieval process similar to that described by the authors of [7, 16]: For a given suspicious document, (1) the document is chunked, (2) the chunks are represented under the style retrieval model, and (3) style differences are identified by means of outlier detection among the chunk representations. (4) After post-processing, the identified chunks are returned as potentially plagiarized passages.

Chunking All of the submitted detectors employ sliding window chunking with chunk sizes ranging from 200 to 1000 words. The slide stepping of the window ranges from 40 to 500 words. The best performing detectors use chunk sizes of 400 words [9] and 1000 words [6].

Retrieval Model Retrieval models for intrinsic plagiarism detection are comprised of a model function that maps texts onto feature representations along with a similarity measure to compare representations. The submitted detectors use either word-based features or character-based features: Oberreuter et al. [9] use a word vector including stop words with tf-weighting, Akiva [1] use a binary word vector including only the 100 rarest words that appear in at least 5% of all chunks, while Kestemont et al. [6] use the 2500 most frequent char-3-grams, and Rao et al. [14] use char-3-grams as well as other well-known features that quantify writing style. Notice that the choice of features determines the least sensible chunk length, since some features require a minimum amount of text in order to provide robust results. Regarding similarity measures, all except one detector employ measures similar to Stamatatos' nd_1 [16]. The detector of Akiva employs cosine similarity.

Table 2. Performances of 5 intrinsic plagiarism detectors on the PAN-PC-11. The detectors are ordered by their $plagdet$ performance. As a baseline, the gray columns show the performances of the best performing detector of PAN'09 [16].

Corpus	$plagdet$					$prec$					rec					$gran$				
Subset	[9]	[16]	[6]	[1]	[14]	[9]	[16]	[6]	[1]	[14]	[9]	[16]	[6]	[1]	[14]	[9]	[16]	[6]	[1]	[14]
entire	**.33**	.19	.17	.08	.07	**.31**	.14	.11	.07	.08	.34	.41	**.43**	.13	.11	**1.00**	1.21	1.03	1.05	1.48
Case length																				
short	**.21**	.01	.14	.07	.05	**.26**	.01	.11	.08	.11	.17	.09	**.22**	.08	.06	**1.00**	1.00	1.12	1.15	2.16
medium	**.20**	.12	.08	.03	.04	**.19**	.08	.05	.02	.03	.21	**.27**	.24	.06	.06	**1.00**	1.01	1.00	1.01	1.05
long	**.03**	.10	.01	.01	.00	**.02**	.12	.00	.01	.00	.07	**.19**	.12	.04	.02	**1.00**	1.77	1.00	1.00	1.01
Translation																				
automatic	**.31**	.20	.23	.14	.07	**.36**	.21	.17	.13	.09	.28	.29	**.43**	.16	.09	**1.00**	1.35	1.06	1.07	1.46
manual	**.11**	.07	.04	.02	.02	**.10**	.05	.02	.01	.02	.13	.12	**.15**	.03	.05	**1.00**	1.00	1.00	1.00	1.06
Plagiarism per document																				
hardly	**.37**	.19	.29	.16	.08	**.45**	.14	.23	.16	.16	.32	**.45**	.44	.18	.09	**1.00**	1.15	1.08	1.06	1.69
medium	**.35**	.25	.17	.08	.07	**.33**	.32	.11	.07	.08	.36	.35	**.43**	.12	.11	**1.00**	1.49	1.02	1.06	1.52
Document length																				
short	**.38**	.21	.20	.10	.06	**.37**	.34	.13	.07	.08	.38	.16	**.55**	.18	.11	**1.00**	1.00	1.06	1.05	1.81
medium	**.40**	.28	.28	.13	.10	**.44**	.23	.21	.07	.17	.37	**.48**	.47	.16	.11	**1.00**	1.17	1.03	1.06	1.43
long	**.28**	.18	.17	.04	.12	**.32**	.13	.13	.11	.16	.25	**.53**	.24	.03	.10	**1.00**	1.33	1.00	1.00	1.07

Outlier Detection Based on the style retrieval model, outlier detection attempts to identify chunks of the suspicious document that are noticeably different from the rest. The following two strategies have been applied this year: (1) measuring the deviation from the average document style, and (2) chunk clustering. The former strategy follows the original proposal of [7] by comparing each chunk representation with that of the whole suspicious document [9, 14]. Rationale of this approach is to measure the extent the style of a chunk matches the average style of the whole suspicious document. A significant deviation is interpreted as an indication of different authorship. Chunk clustering, on the other hand, compares the chunk representations and attempts to cluster them into groups of similar styles, whereas the chunks of each group may have been written by a different author [1, 6]. While a lot of finesse has to be applied in order to achieve reasonable performance in outlier detection, it is important to keep in mind that these algorithms also depend crucially on the choice of retrieval model.

Post-processing With regard to post-processing most detectors merge overlapping and consecutive chunks that have been identified as outliers in order to decrease detection granularity.

2.2 Evaluation

Table 2 shows the detection performances of the aforementioned detectors according to the detection performance measures plagdet, precision, recall, and granularity. The overall best performing detector stems from Oberreuter et al. [9]; it outperforms all other detectors on all except one performance measure. The detector of Kestemont et al. [6], however, performs best with regard to recall. Interestingly, both detec-

tors achieve their performances based on different outlier detection strategies, namely chunk-document comparison and chunk-chunk comparison. Under the latter it appears to be more difficult to achieve a good tradeoff between precision and recall. The detector of Stamatatos [16] serves as a baseline for comparison. It has been the best performing intrinsic plagiarism detector of PAN'09, and it still outperforms all except one of the submitted detectors: the detector of Oberreuter et al. performs more than 40% better.

While the baseline detector performs better on medium and long plagiarism cases, the submitted detectors perform better on short and medium length cases. Automatically translated plagiarism is detected better than manually corrected translation plagiarism. Regarding the ratio of plagiarism per document, the picture is not clear, with the first detector performing similar on documents with a hardly and medium ratio of plagiarized text, the second performing better on the latter, and the third performing better on the former. Finally, with regard to document length, a medium length seems to be best for intrinsic plagiarism detection.

2.3 Discussion

The outstanding performance of the detector of Oberreuter et al. [9] looks very encouraging, but it should be taken with a grain of salt. The detector's retrieval model quantifies the uniqueness of a word with regard to the whole suspicious document. However, during construction of the PAN-PC-11, plagiarism cases have been inserted into the suspicious documents from randomly chosen source documents, so that no topic overlap between a suspicious document and its sources can be expected. I.e., with a high probability, words have been inserted into the suspicious documents that did not occur beforehand. A retrieval model which builds on computing word uniqueness hence benefits from this construction principle. Moreover, it is surprising that a retrieval model which builds on words instead of writing style features which have been shown in the past to outperform word-based style quantification should perform that well on intrinsic plagiarism detection. Presumably, the PAN'11 performance may not be achieved in different settings.

These results are nonetheless important as they pinpoint a problem with constructing a corpus for intrinsic plagiarism detection. Randomly inserting text into a document may preserve writing style, but it obviously doesn't represent plagiarist behavior, and it hence opens the door to detection approaches which may not be applicable in practice. Though, at the time of writing, no better way of constructing a corpus for intrinsic plagiarism detection evaluation is at hand, this will be an important subject for future research. Also, the second best performing detector of Kestemont et al. [6] points into new directions to improve detection performance in terms of recall.

3 External Plagiarism Detection

This section surveys the external plagiarism detectors and evaluates their performances.

3.1 Survey of Detection Approaches

External plagiarism detection continues to be an important part of the PAN plagiarism detection competition. Several of the plagiarism detectors that have been evaluated this year are enhanced versions of detectors that have been evaluated in previous years. An analysis of the submitted notebooks reveals that the generic retrieval process for external plagiarism detection described at length in the overview papers of PAN'09 and PAN'10 did not change much [11, 12]: For a given suspicious document and a collection of potential source documents, (1) all documents are pre-processed using an indexing pipeline that normalizes the word tokens by removing stop words, stems the remainder, and replaces words with one of their synonyms. Moreover, non-English documents are typically translated to English using Google Translate or other translation services. (2) A set of candidate source documents is retrieved from the collection of source documents, (3) each candidate document is compared in detail to the suspicious document in order to extract similar passages of text, and finally, (4) the extracted passages are post-processed to filter false positive detections, while the remainder is returned to the user as potential plagiarism detections.

While the specifics of the algorithms applied in each of the aforementioned steps have not changed much, it can be observed that some participants focus on certain aspects such as runtime performance, cross-language detection, and obfuscation. For instance, Cooke et al. [2] report to process the entire PAN-PC-11 in about 12 minutes excluding translation, whereas Rodríguez Torrejón and Martín Ramos [15] report to require 30 minutes including translation. Many other detectors need hours. The top three plagiarism detectors of Grman and Ravas [4], Grozea and Popescu [5], and Oberreuter et al. [9], however, follow the best practices that emerged in the previous editions of PAN.

3.2 Evaluation

Table 3 shows the detection performances of the detectors that took part in PAN'10; the table reports the detection performance measures $plagdet$, precision, recall, and granularity. The best-performing detector of Grman and Ravas [4] dominates all other detectors on all measures and on almost all variations of the corpus parameters. The second and third best-performing detector of Grozea and Popescu [5] and Oberreuter et al. [9] achieve 33% and 60% less $plagdet$ performance respectively.

Table 3. Performances of 9 external plagiarism detector on the PAN-PC-11. The detectors are ordered by their $plagdet$ performance.

Corpus	$plagdet$									$prec$									rec									$gran$								
Subset	[4]	[5]	[9]	[2]	[15]	[14]	[10]	[8]	[3]	[4]	[5]	[9]	[2]	[15]	[14]	[10]	[8]	[3]	[4]	[5]	[9]	[2]	[15]	[14]	[10]	[8]	[3]	[4]	[5]	[9]	[2]	[15]	[14]	[10]	[8]	[3]
entire	**.56**	.42	.35	.25	.23	.20	.19	.08	.00	**.94**	.81	.91	.71	.85	.45	.44	.28	.01	**.40**	.34	.23	.15	.16	.16	.14	.09	.00	**1.00**	1.22	1.06	1.01	1.23	1.29	1.17	2.18	2.00
Paraphrasing																																				
none	**.97**	.85	.91	.81	.81	.60	.66	.40	.01	**.97**	.84	.94	.75	.82	.53	.58	.32	.08	**.97**	.90	.88	.87	.79	.70	.81	.72	.01	**1.00**	1.02	1.01	1.00	1.00	1.00	1.02	1.14	4.64
manual	.49	**.50**	.47	.17	.32	.05	.29	.02	.00	**.99**	.96	.98	.86	.93	.38	.80	.43	.01	.33	**.36**	.31	.09	.20	.03	.18	.02	.00	1.01	1.06	1.01	**1.00**	1.03	1.05	1.01	2.71	1.31
auto-low	**.71**	.60	.55	.25	.32	.35	.29	.15	.00	**.95**	.90	.93	.74	.92	.62	.57	.45	.01	.56	**.58**	.42	.15	.25	.32	.23	.18	.00	**1.00**	1.27	1.08	1.01	1.34	1.33	1.22	2.29	1.32
auto-high	**.15**	.13	.05	.04	.01	.03	.00	.00	.00	**.77**	.64	.67	.48	.39	.18	.04	.01	.00	**.08**	.08	.03	.02	.01	.02	.00	.00	.00	**1.00**	1.19	1.02	1.01	1.16	1.12	1.01	1.21	1.31
Translation																																				
manual	**.37**	.12	.00	.15	.03	.03	.04	.00	.00	**.69**	.25	.00	.23	.12	.04	.03	.00	.00	**.26**	.08	.00	.11	.01	.03	.06	.00	.00	**1.00**	1.01	1.00	1.00	1.00	1.03	1.04	1.00	1.00
automatic	**.92**	.28	.00	.61	.36	.08	.13	.00	.00	**.97**	.40	.22	.71	.69	.14	.16	.00	.00	**.88**	.23	.00	.55	.24	.09	.13	.00	.00	**1.00**	1.06	1.10	1.01	1.00	1.42	1.16	1.00	1.00
Case length																																				
short	**.23**	.18	.11	.03	.02	.07	.08	.04	.00	**.70**	.47	.54	.09	.05	.12	.14	.04	.00	**.14**	.11	.06	.02	.01	.05	.05	.04	.00	**1.00**	1.01	1.00	1.00	1.00	1.00	1.00	1.04	1.06
medium	**.33**	.28	.26	.14	.19	.14	.11	.05	.00	**.82**	.63	.81	.45	.68	.27	.30	.22	.00	**.21**	.19	.16	.08	.12	.10	.08	.05	.00	**1.00**	1.06	1.01	1.00	1.04	1.05	1.18	2.42	1.33
long	**.16**	.08	.03	.12	.07	.03	.02	.01	.00	**.71**	.60	.65	.35	.55	.35	.13	.13	.01	**.09**	.07	.02	.07	.06	.03	.02	.01	.00	**1.00**	1.89	1.32	1.02	1.62	2.00	1.51	3.11	3.46
Document length																																				
short	**.63**	.56	.40	.22	.21	.33	.27	.15	.00	**.99**	.86	.94	.71	.91	.65	.76	.71	.03	**.46**	.44	.26	.13	.13	.23	.18	.13	.00	**1.00**	1.05	1.02	1.00	1.06	1.06	1.09	1.82	1.37
medium	**.55**	.40	.34	.26	.24	.21	.19	.09	.00	**.97**	.86	.94	.78	.91	.56	.61	.45	.01	**.39**	.33	.22	.16	.17	.16	.13	.09	.00	**1.00**	1.27	1.06	1.01	1.22	1.32	1.19	2.25	2.17
long	**.53**	.35	.32	.25	.24	.13	.16	.05	.00	**.92**	.81	.93	.70	.85	.43	.38	.21	.00	**.37**	.28	.21	.15	.18	.11	.12	.06	.00	**1.00**	1.28	1.10	1.01	1.32	1.49	1.22	2.46	2.41
Plagiarism per document																																				
hardly	**.57**	.47	.40	.19	.24	.19	.19	.08	.00	**.89**	.75	.89	.73	.78	.31	.30	.14	.00	**.42**	.40	.27	.11	.16	.16	.16	.10	.00	**1.00**	1.14	1.03	1.00	1.13	1.16	1.10	1.90	1.95
medium	**.63**	.43	.35	.29	.28	.19	.21	.07	.00	**.94**	.75	.89	.73	.84	.39	.42	.23	.00	**.48**	.36	.23	.18	.19	.15	.15	.08	.00	**1.00**	1.17	1.06	1.01	1.16	1.27	1.14	2.14	1.78
much	**.52**	.38	.33	.25	.23	.21	.18	.09	.00	**.95**	.85	.94	.71	.88	.60	.55	.44	.01	**.36**	.31	.21	.15	.16	.16	.13	.09	.00	**1.00**	1.26	1.07	1.01	1.29	1.35	1.20	2.26	2.39
entirely	**.54**	.38	.32	.28	.22	.22	.17	.08	.00	**.96**	.88	.94	.71	.90	.59	.65	.51	.01	**.38**	.31	.20	.18	.16	.17	.12	.09	.00	**1.00**	1.29	1.08	1.01	1.29	1.35	1.24	2.35	1.72

The precision performance of the top five detectors is very high, while the recall performances varies from poor to medium, depending on the corpus parameter. In fact, only verbatim plagiarism is detected with high recall. The granularity of the top five detectors is close to 1.0 in most cases, while the remaining detectors show comparably unstable performance characteristics. With regard to the corpus parameters it can be seen that manual obfuscation in terms of paraphrasing and translation is much more difficult to be detected than automatic obfuscation. While the length of a case has a certain influence on detection performance (short cases are less well detected), the document length as well as the ratio of plagiarism per document have no effect on detection performance.

Compared to the performances reported in PAN'09 and PAN'10, a drop in the $plagdet$ performance can be observed in PAN'11. This fact does not indicate a worse detection performance compared to previous years, but should be attributed to an increased detection difficulty: during the construction of the PAN-PC-11 corpus we have lowered the ratio of plagiarism cases that are not obfuscated, while we have increased the number of cases that are manually or automatically obfuscated. Within the previous as well as this competition, it became clear that verbatim plagiarism poses no challenge to detection anymore, so that the high absolute performance values reported may lead to the false conclusion that plagiarism detection in general is close to being solved. By changing the respective corpus parameters, we address this issue; however it forecloses a direct comparison between performances of this year and those of earlier years.

3.3 Discussion

The external plagiarism detection sub-task of the PAN'11 competition on plagiarism detection has matured in the past three years: in the first year the size of the test corpus posed the biggest difficulty, in the second year the revised corpus introduced new challenges such as manual obfuscation, whereas in the third year the difficulty of the task was increased. Furthermore, many of the plagiarism detectors have been evaluated more than once. Unfortunately, however, some participants chose not to share their approaches, so that some of the achieved performances cannot be replicated or verified.

4 Conclusion

The results of the 3rd international competition on plagiarism detection, PAN'11, can be summarized as follows: 11 plagiarism detectors have been developed from which 7 detectors tackle external plagiarism detection, 2 detectors tackle intrinsic plagiarism detection, and 2 detectors handle both. Five of the detectors have been evaluated the second time, and one for the third time. One of the new detectors dominates all other detectors in terms of detection performance on the third revised version of the PAN plagiarism corpus PAN-PC-11. The corpus features plagiarism cases with an increased level of detection difficulty compared to previous corpus versions.

The lessons learned from the competition include that the portion of our corpus which is dedicated to intrinsic plagiarism detection may be biased: the evaluation may not favor realistic detection approaches over less realistic ones. Moreover, the increased

detection difficulty draws a clearer picture of the detection performances of today's plagiarism detectors.

Acknowledgements

We would like to thank Yahoo Research for sponsoring the PAN competition for the third time now. We further thank the participants of PAN for their dedicated work without which this event would not be possible. This work was partly funded by the European Commission as part of the WIQEI IRSES project (grant no. 269180) within the FP7 Marie Curie People Framework, by MICINN as part of the Text-Enterprise 2.0 project (TIN2009-13391-C04-03) within the Plan I+D+i, and as part of the VLC/CAMPUS Microcluster on Multimodal Interaction in Intelligent Systems.

Bibliography

[1] Navot Akiva. Using Clustering to Identify Outlier Chunks of Text: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[2] Neil Cooke, Lee Gillam, Henry Cooke Peter Wrobel, and Fahad Al-Obaidli. A High-performance Plagiarism Detection System: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[3] Aniruddha Ghosh, Pinaki Bhaskar, Santanu Pal, and Sivaji Bandyopadhyay. Rule Based Plagiarism Detection using Information Retrieval: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[4] Ján Grman and Rudolf Ravas. Improved Implementation for Finding Text Similarities in Large Collections of Data: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[5] Cristian Grozea and Marius Popescu. The Encoplot Similarity Measure for Automatic Detection of Plagiarism: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[6] Mike Kestemont, Kim Luyckx, and Walter Daelemans. Intrinsic Plagiarism Detection Using Character Trigram Distance Scores: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[7] Sven Meyer zu Eißen and Benno Stein. Intrinsic Plagiarism Detection. In Mounia Lalmas, Andy MacFarlane, Stefan Rüger, Anastasios Tombros, Theodora Tsikrika, and Alexei Yavlinsky, editors, *Advances in Information Retrieval. 28th European Conference on IR Research (ECIR 06)*, volume 3936 LNCS of *Lecture Notes in Computer Science*, pages 565–569, Berlin Heidelberg New York, 2006. Springer. ISBN 3-540-33347-9. doi: http://dx.doi.org/10.1007/11735106_66.

[8] Rao Muhammad Adeel Nawab, Mark Stevenson, and Paul Clough. External Plagiarism Detection using Information Retrieval and Sequence Alignment: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[9] Gabriel Oberreuter, Gaston L'Huillier, Sebastián A. Ríos, and Juan D. Velásquez. Approaches for Intrinsic and External Plagiarism Detection: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[10] Yurii Palkovskii, Alexei Belov, and Iryna Muzyka. Using WordNet-based Semantic Similarity Measurement in External Plagiarism Detection: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[11] Martin Potthast, Benno Stein, Andreas Eiselt, Alberto Barrón-Cedeño, and Paolo Rosso. Overview of the 1st International Competition on Plagiarism Detection. In Benno Stein, Paolo Rosso, Efstathios Stamatatos, Moshe Koppel, and Eneko Agirre, editors, *SEPLN 09 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09)*, pages 1–9. CEUR-WS.org, September 2009. URL http://ceur-ws.org/Vol-502.

[12] Martin Potthast, Alberto Barrón-Cedeño, Andreas Eiselt, Benno Stein, and Paolo Rosso. Overview of the 2nd International Competition on Plagiarism Detection. In Martin Braschler and Donna Harman, editors, *Notebook Papers of CLEF 10 Labs and Workshops*, September 2010. ISBN 978-88-904810-0-0.

[13] Martin Potthast, Benno Stein, Alberto Barrón-Cedeño, and Paolo Rosso. An Evaluation Framework for Plagiarism Detection. In Chu-Ren Huang and Dan Jurafsky, editors, *23rd International Conference on Computational Linguistics (COLING 10)*, pages 997–1005, Stroudsburg, PA, USA, August 2010. Association for Computational Linguistics.

[14] Sameer Rao, Parth Gupta, Khushboo Singhal, and Prasenjit Majumder. External & Intrinsic Plagiarism Detection: VSM & Discourse Markers based Approach: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[15] Diego Antonio Rodríguez Torrejón and José Manuel Martín Ramos. Crosslingual CoReMo System: Notebook for PAN at CLEF 2011. In *Notebook Papers of CLEF 2011 LABs and Workshops, 19-22 September, Amsterdam, The Netherlands*, September 2011.

[16] Efstathios Stamatatos. Intrinsic Plagiarism Detection Using Character n-gram Profiles. In Benno Stein, Paolo Rosso, Efstathios Stamatatos, Moshe Koppel, and Eneko Agirre, editors, *SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09)*, pages 38–46. Universidad Politécnica de Valencia and CEUR-WS.org, September 2009. URL http://ceur-ws.org/Vol-502.