Towards Active Linguistic Authentication

Patrick Juola, John Noecker Jr., Ariel Stolerman, Michael Ryan, Patrick Brennan, Rachel Greenstadt

▶ **To cite this version:**

Patrick Juola, John Noecker Jr., Ariel Stolerman, Michael Ryan, Patrick Brennan, et al.. Towards Active Linguistic Authentication. 9th International Conference on Digital Forensics (DF), Jan 2013, Orlando, FL, United States. pp.385-398, 10.1007/978-3-642-41148-9_25 . hal-01460617

HAL Id: hal-01460617

https://inria.hal.science/hal-01460617

Submitted on 7 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Chapter 25

TOWARDS ACTIVE LINGUISTIC AUTHENTICATION

Patrick Juola, John Noecker Jr., Ariel Stolerman, Michael Ryan, Patrick Brennan and Rachel Greenstadt

Abstract Biometric technologies offer a new and effective means for securing computers against unauthorized access. Linguistic technologies and, in particular, authorship attribution technologies can assist in this effort. This paper reports on the results of analyzing a novel corpus that was developed to test the possibility of active linguistic authentication. The study collected the one-week work product of nineteen temporary workers in a simulated office environment. The results demonstrate that techniques culled from the field of authorship attribution can identify workers with more than 90% accuracy.

Keywords: Stylometry, authentication, authorship attribution, profiling

1. Introduction

Standard password-based identification systems are known to have flaws. Passwords can be forgotten, written down, stolen, and guessed. If any of these events occur, an intruder has the keys to the kingdom. Biometric-based identification systems have been proposed to bypass or mitigate some of these problems – it is hard to forget your own thumbprint. But developing and testing these systems can be a challenge precisely because of the need to handle a wide variety of humans, especially when the biometric task is challenging and time-consuming.

One possibility for biometric validation is the individual use of language. Prior work has shown that the authorship of documents containing as little as a few hundred words can be correctly identified. In a typical office environment, a worker types many more words each day and, thus, could continually identify himself or herself to an appropriate security screening program. This paper reports on the development of

a novel corpus to enable this type of analysis and provides the results of a proof-of-concept analysis.

2. Background

This section discusses authentication, stylometry and authorship attribution, and the JGAAP and JStylo systems.

2.1 Authentication

Passwords are commonly used to secure computer systems. The flaws associated with passwords are well known. Passwords can be forgotten, which prevents access. Because passwords are difficult to remember, they are often written down, but this means that they can be stolen and used to enable unauthorized access. Also, passwords can be guessed or recovered using cryptanalytic techniques.

More subtly, traditional passwords are limited in the type of protection they provide. Once a person presents his password, he is authenticated and may use his computer. If this person steps away from his desk for a moment, anyone could step up to the keyboard and use the computer. Chaski [5] defined this as the "who's at the keyboard" dilemma and cites several examples, including the case of a dead body found in a room with a disputed suicide note typed on a computer, and the case of an email sent while an employee was away from her desk at lunch. Coulthard [6] describes a similar case involving a disputed email. In other words, traditional password protection provides only passive "perimeter" security that does not prevent "insiders" from abusing their status.

An improved security model would involve continuous, active authentication where the behavior of the person at the keyboard is monitored. Security measures can be triggered immediately when the behavior of the person changes. The security measures can be as simple as reauthentication or as complicated as triggering a call to security personnel and locking the computer down.

2.2 Stylometry and Authorship Attribution

Authorship attribution, also called stylometry or stylistics, is a well-established field of study [10, 12, 18, 24], although it has typically not been used for authentication. The theory behind authorship attribution is that each person has his own unique "stylome" [28], a unique set of idiolectal choices that describe his speaking and writing style. At a group level, this is the kind of choice that causes the British to walk on "pavements" instead of "sidewalks." At an individual level, it is

the kind of choice that causes a person to place a fork "to" the left of the plate instead of "on" the left or "at" the left. Quantifying these choices, for example, by making a histogram of function words [3, 21] or of character n-grams [25] can enable investigators to develop a computationally tractable summary of stylistic choices and form judgments based on these summaries.

Standard practice in stylometric investigations involves a detailed comparison of stylistic features culled from a training set of documents. The questioned document is then compared against the training set, typically using a classification or machine learning algorithm, and an appropriate decision is taken. A related problem is authorship verification, where a novel document is compared with a summary by a single author. If the novel document is close (stylistically) to the summary, it is inferred to been written by the author; and if it is distant, then not.

A third application of stylometric technology is in stylistic profiling [2, 8, 16, 19, 27], where the objective is not to identify a specific person, but to identify characteristics of the writer such as age, gender, social class and native language. Again, a typical study involves collecting samples of (for example) male writing and female writing, and then comparing a questioned document against the stylistic summaries to classify the document as being written by a male or female.

The application of this technology to authentication is straightforward. Instead of using a training set of documents, a pseudo-document containing the user's long-term behavior is used, and the recent behavior at the keyboard is verified as being consistent with the long-term behavior. A significant inconsistency would trigger a security response. We are, therefore, proposing the use of linguistic stylistics as a biometric, similar to the use of typing speed or mouse movements [30].

As a purely text classification technology (i.e., not real time and not using keystrokes, but instead using finished documents and often involving forced-choice comparisons between a fixed group of authors), authorship attribution is in some regards a mature technology. For example, at the PAN-2012 Conference (see `pan.webis.de`), the top three methods all classified more than 80% of 241 documents correctly with (in some cases) more than a dozen distractor authors. We have reason to believe that authorship authentication may be even more accurate because issues such as automatic spelling correction [6] will not normalize individual writing patterns – someone who is a poor typist or who continually misspells "*toutch" [29] will not have this idiosyncratic quirk airbrushed away. We expect that an appropriately chosen analytic method will eventually achieve similar or better results in this novel context.

2.3 JGAAP

In light of the differences among possible analyses, an obvious question is: Which method works best? In order to address this question, the Evaluating Variations in Language Laboratory at Duquesne University has developed a modular system for the development and comparative testing of authorship attribution methods [12, 15]. This system, called JGAAP (Java Graphical Authorship Attribution Program) provides a large number of interchangeable analysis modules to handle different aspects of the analysis pipeline such as document preprocessing, feature selection and analysis/visualization.

The JGAAP project has been very successful, creating one of the most widely-used systems for authorship analysis, leading the way in the search for best practices, and developing a group of protocols accurate enough to have been used in court [14]. Most importantly, it has created a standard, tested set of operational primitives (such as approximately two dozen ways to assess linguistic differences) [26] based on various underlying cognitive models and computational approaches [11]. This toolset will be leveraged in a wide-ranging and systematic exploration of several different types of analysis and relationships. Taking combinatorics into account, the number of different ways to analyze a set of documents numbers in the millions and this can be expanded by an inventive user. JGAAP is freely available (from `www.jgaap.com`), making it a useful testbed for other researchers.

For example, Grant [7] describes a criminal case involving vocabulary comparisons among text messages sent by a number of people. The technical question involved in this case hinged on the existence and number of specific words (or spelling variants such as "wen" for "when" or "4get" for "forget") that were used by the one person who was involved. This can be captured by measuring document similarity using the Jaccard or intersection distance, essentially a measure of vocabulary overlap without regard to specific frequencies. In contrast, the classic Mosteller-Wallace [21] study of historical documents examined frequency differences among common (and therefore shared) vocabulary. The important question was not whether or not people used words like "upon" (because we all do), but whether the disputed document used that word more like person A or person B. This type of analysis can be done by measuring document similarity using frequency measures such as normalized cosine (dot-product) distance [22] or Manhattan distance. JGAAP has been expanded to include all three of these measures as well as many others.

2.4 JStylo

JStylo (see `psal.cs.drexel.edu`) is an open-source authorship attribution platform developed at the Privacy, Security and Automation Laboratory at Drexel University on top of the JGAAP project. It was primarily created to allow cross-feature analysis, where multiple features can be extracted and included in one analysis, an option that was not available in JGAAP at the time. JStylo is complemented by a dual analysis tool, along with Anonymouth [20], a writing-style anonymization platform, whose underlying authorship attribution engine is JStylo.

In JStylo, every feature can be one of two types: a class of feature frequencies (e.g., features "a"..."z" for the "Letters" feature class), or a numeric evaluation of the input documents (e.g., Yule's characteristic K). An additional advantage of JStylo is its fine-resolution feature definition capabilities. Each feature is uniquely defined by a set of its own document-preprocessing tools, one unique feature extractor (core of the feature), feature-postprocessing tools and normalization/factoring options. All of JGAAP core features are available in JStylo, in addition to some newly developed features such as regular-expression-based extraction.

With regard to analysis capabilities, the main classification tools available in JStylo are drawn from Weka [9], the popular data mining and machine learning platform. These include classifiers commonly used for authorship attribution such as support vector machines, neural networks, naïve Bayes classifiers and decision trees. In addition, JStylo provides an implementation of the Writeprints authorship attribution technique [1], which is known for its high accuracy in scenarios with large numbers of authors.

Although JStylo lacks the maturity of JGAAP, it compensates with a vast range of features and, more importantly, the ability to combine them. This capability was leveraged to conduct the preliminary analysis described in this paper, where the extensive feature set used with the Writeprints method was applied to the collected data.

3. Work Product Corpus

We created a simulated work environment to generate a suitable corpus for validation. A rented space in downtown Pittsburgh was set up as an office staffed by temporary employees (subjects). The subjects were supervised by Juola & Associates staff and asked, over the course of a week, to research and write blog articles "related to Pittsburgh in some way." This provided them with an incentive to use standard computer tools such as browsers and search engines to conduct research and word

processors to do the actual writing. The task was expected to take approximately six hours per day, except for a shorter first day as described below. The subjects were provided with a reasonable degree of topical similarity, but they had enough freedom to be individually distinctive so that they could not be trivially distinguished on the basis of the type of task they were doing. The subjects were not restricted from accessing personal websites or playing standard games, and they were allowed to copy and paste material as long as the final articles were their own work. As expected, the most commonly-used applications by the subjects were Internet Explorer and Microsoft Word.

The subjects were also advised that their computers were equipped with tracking software, in particular, a macro recorder for measuring keyboard use, including individual keystrokes and dynamic information such as timing, length of keypress and overlap between keys. The macro recorder also measured mouse events such as clicks and movements. Keylogging software was used to record text as it was entered, including mapping text to specific applications, clipboard use and browsing history. The subjects were notified that, although good faith efforts would be made to scrub the data prior to analysis, all input would be captured (including personal information such as Facebook account names and passwords) and that it could not be guaranteed that all this information would be wiped after the experiments were completed. The subjects were given an opportunity to request that specific strings (e.g., user names and passwords) be automatically redacted, but it would have been easier for subjects just to change their Facebook passwords or not log into Facebook from work.

In addition to the main tasks, the subjects had two types of secondary tasks. The morning of the first day was spent in an orientation process that included the administration of a number of psychometric tests that measured traits such as personality, self-esteem and learning styles. (We do not report on this aspect of the study.) During the final two hours of the day, the subjects were asked to perform a set of small, explicitly-defined tasks (microtasks) such as describing a specific local landmark or event or summarizing an article. This provided a set of very detailed, task-specific data with much tighter control, possibly creating a different environment for task-focused inter-individual comparisons.

Data collection is ongoing. By the end of the project, we expect to have the work product of at least 80 subjects.

4. Analysis

Our analysis is based on data gathered over three weeks according to the protocol described in the previous section. The data set comprised five days of work for each of fourteen subjects (one subject failed to show up for work after being hired). One day of work for one participant was temporarily mislaid; this has since been addressed, but the work product was not analyzed. Consequently, we have a total of 69 days of work product. This work product comprised approximately 280 MB of data, including 17.5 million mouse and keyboard events and 23,000 website visits. Our analysis in this paper only focuses on the language used in the keystroke events.

4.1 Daily Data

In the first phase, we analyzed each day of work as a unit, using hold-one-out cross-validation (i.e., each document was analyzed individually against the other 68 documents) in a content-agnostic way using only character n-gram distribution frequencies. We recognize that requiring a full day of work prior to making security decisions is impractical, but this provides a baseline against which smaller samples can be measured.

Our analysis was performed using JGAAP with three canonicizers: (i) Unify Case, which neutralizes all case distinctions; (ii) Normalize Whitespace, which replaces tabs, newlines and multiple spaces with a single space character; and (iii) Keylogger, which cleans up the logs in several ways. The Keylogger canonicizer removed anything that did not represent a keystroke, including time/date stamps from the keylogger, information about the window from which keystrokes originated, and whitespace to make the logs readable. Note that this means that if a subject typed something in a browser and then switched to Word, there would be nothing left in the log to record this activity. Thus, we would have "google.com-ENTER-is a PittsburghHotels in Pittsburgh institution" (i.e., multiple window inputs mashed together).

Next, we converted special keys to single character representations. For example, -ENTER- was converted to a newline. Or, arrow key -UP- was was converted to a placeholder that was unlikely to appear in the actual input.

Finally, the analysis was performed using a simple nearest-neighbor classifier with the Manhattan distance (i.e., L_1 distance) or intersection distance (i.e., Jaccard distance) based on histograms of character n-grams of lengths ranging from one through five.

The results are shown in Table 1. All results are based on the number of definitive classifications, nominally 69 documents (denominator).

Table 1. Daily analysis classification results.

Analysis Method	Results
Manhattan 1-grams	37/69
Manhattan 2-grams	53/69
Manhattan 3-grams	62/69
Manhattan 4-grams	58/69
Manhattan 5-grams	50/69
Intersection 1-gram	6/21
Intersection 3-gram	23/68
Intersection 4-gram	22/69
Intersection 5-gram	22/69

However, all ties are reported as a single author, which yields a lower number in the denominator (e.g., 21 and 68).

Based on these results, it is clear that individual subjects can be distinguished with high accuracy. The result is 88.4% accurate, better, in fact, in purely nominal terms than the PAN-2012 Conference winner. It is also clear that, in this particular framework, the Manhattan distance is a more promising and accurate measure than the intersection distance, suggesting that it is more useful to measure frequency differences than mere presence/absence distinctions. Nevertheless, the fact that decisions were possible at all in more than 21 cases using individual characters and intersection distance hints at the power of using keyboard interactions in a forensic or security tool. In the 21 cases, there were certain keys that some individuals did not hit at all over the course of an entire day. Clearly, this is a much richer set of features and events than just alphanumeric characters and punctuation.

4.2 Fixed-Size Sliding Window

In the second phase of the analysis, we concatenated the keystroke data of all the users and re-divided the result into consecutive non-overlapping documents (windows) with predefined fixed sizes of 100, 500 and 1,000 words. This type of analysis is closer to the active authentication problem we aim to solve, because any real-time monitoring system would eventually be based on evaluating sliding windows of user input on-the-fly in an attempt to catch unauthorized users. One of the challenges is to decrease the window size as much as possible (leading to a quicker response time) while maintaining high accuracy and low false positives and false negatives (i.e., undetected unauthorized users and false alarms for authorized users, respectively).

Table 2. Writeprints-inspired feature set.

Group	**Features**
Lexical	Character count
	Average word length
	Letters
	50 most common letter bigrams
	50 most common letter trigrams
	Percentage of letters
	Percentage of uppercase letters
	Percentage of digits
	Digits
	2-digit numbers
	3-digit numbers
	Word length distribution
	Special characters
Syntactic	50 most common function words
	Punctuation
	Part-of-speech (POS) tags
	50 most common POS bigrams
	50 most common POS trigrams
Content	50 most common words in the corpus
	50 most common word bigrams in the corpus
	50 most common word trigrams in the corpus
Idiosyncrasies	Common misspellings

As in the first phase, the data was stripped of keylogger metadata, special keys were converted to unique single-character placeholders and whitespace was normalized. However, the raw data was not case unified and all special keys (e.g., -ENTER- and -TAB-) were replaced with placeholders (rather than being converted to newline and tab, respectively) in order to preserve user typing characteristics to the extent possible. Since the data includes special characters, it is more accurate to measure document length in terms of tokens than words (e.g., ch$\beta\beta$Cch$\beta\beta$hicago where β represents backspace).

The second phase used a different feature set from the first phase. Specifically, a close variation of the Writeprints [1] feature set was used; this set includes a vast range of linguistic features across different levels of text (summarized in Table 2). The rich linguistic feature set better captures user writing styles. With the help of the special-character placeholders, some features capture aspects of user style that are usually not found in standard authorship problem settings. For example,

Table 3. Sliding-window analysis classification results.

Window Size	Accuracy	Weighted Avg. FN	Weighted Avg. FP
SMO			
100	81.07%	18.93%	2.0%
500	93.06%	6.94%	0.9%
1,000	93.33%	6.77%	0.9%
KNN			
100	71.79%	28.21%	2.4%
500	83.04%	16.96%	1.5%
1,000	83.13%	16.87%	1.1%

frequencies of backspaces and deletes provide some evaluation of a user's typo rate or lack of decisiveness.

Our analysis was performed in JStylo using 10-fold cross-validation for evaluation. We used two Weka classifiers: KNN classifier with $K = 1$ and Manhattan distance (similar to the previous phase) and SMO SVM [23] with a soft margin constant $C = 1$ and polynomial kernel of degree one. SMO solves multi-class problems using pairwise binary classification. The features are normalized by default for both classifiers.

The results for the second phase are shown in Table 3. If it was clear from the first phase that the subjects can be distinguished with high accuracy based on one day of work product, the results in the second phase demonstrate that high distinguishability can be achieved by examining token sequences of up to 1,000 in length. Moreover, the statistically insignificant ($p < 0.01$) difference between the results for 500-token and 1,000-token windows implies that verification could be achieved after merely 500 tokens of user input. However, the statistically significant ($p < 0.01$) difference in accuracy when dropping down to 100-token windows suggests that there is a minimal threshold to consider for these settings. Finally, support vector machines, which are used extensively in authorship attribution due to their high performance and accuracy prevailed in these settings as well. In particular, the support vector machines outperformed KNN ($p < 0.01$), with the best results being 93.33% accuracy for 1,000-token windows and 93.06% accuracy for 500-token windows.

5. Conclusions

From a practical standpoint, waiting for almost an entire day to determine whether or not an unauthorized individual is using a work com-

puter leaves much to be desired. Nevertheless, the results demonstrate the success of the proof-of-concept system and that it is easy enough to vary the window size and other parameters to obtain good results. For example, it could be useful to compare and analyze the accuracy achieved with temporal windows that consider hour-long, five-minute long or minute-long slices of work instead of length-based windows.

Similarly, we have chosen only a few types of features/events to analyze out of the dozens provided by JGAAP and JStylo, just a few types of classifiers and ignored the possibility of using other classification techniques such as binary-class support vector machines, neural networks, linear discriminant analysis and latent Dirichelet allocation. In the longer run, it has been shown that ensemble methods like mixture-of-experts [13] tend to outperform individual analyses, and it is necessary to investigate that, even if small/short samples do not work well under single analysis, it is possible to achieve good authentication with multiple independent analyses. Also, it may be promising to combine linguistic biometrics with other data sources (e.g., mouse movements that have yielded good results [30]) as a base for authentication.

From a security standpoint, a key question is accuracy in the face of active deception that fools the monitor. While our experiments do not consider this issue, recent research in stylistic deception [4, 17, 20] provides avenues for future work in this area.

The U.S. Department of Defense has suggested (in DARPA BAA 12-06) that computer-captured biometrics can be "used to uniquely recognize humans" with high accuracy and minimal intrusiveness. We believe that our work confirms this suggestion, providing approximately 90% accuracy across nearly two dozen subjects. While further work is required to improve accuracy and to address verification in unknown settings – and possibly to integrate with other sources of biometric information and to develop a commercial-scale security system – the results presented in this paper strongly confirm the promise of our approach.

Acknowledgement

This research was supported by the National Science Foundation under Grant No. OCI-1032683 and by DARPA under BAA-12-06.

References

[1] A. Abbasi and H. Chen, Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace, *ACM Transactions on Information Systems*, vol. 26(2), pp. 7:1–7:29, 2008.

[2] S. Argamon, M. Koppel, J. Pennebaker and J. Schler, Automatically profiling the author of an anonymous text, *Communications of the ACM*, vol. 52(2), pp. 119–123, 2009.

[3] J. Binongo, Who wrote the 15th Book of Oz? An application of multivariate analysis of authorship attribution, *Chance*, vol. 16(2), pp. 9–17, 2003.

[4] M. Brennan and R. Greenstadt, Practical attacks against authorship recognition techniques, *Proceedings of the Twenty-First Conference on Innovative Applications of Artificial Intelligence*, pp. 60–65, 2009.

[5] C. Chaski, Who's at the keyboard: Authorship attribution in digital evidence investigations, *International Journal of Digital Evidence*, vol. 4(1), 2005.

[6] M. Coulthard, On the admissibility of linguistic evidence, *Brooklyn Law School Journal of Law and Policy*, vol. 21(2), pp. 441–466, 2013.

[7] T. Grant, TXT 4N6: Method, consistency and distinctiveness in the forensic authorship analysis of SMS text messaging, *Brooklyn Law School Journal of Law and Policy*, vol. 21(2), pp. 467–494, 2013.

[8] C. Gray and P. Juola, Personality identification through on-line text analysis, presented at the *Chicago Colloquium on Digital Humanities and Computer Science*, 2012.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. Witten, The Weka Data Mining Software: An update, *SIGKDD Explorations Newsletter*, vol. 11(1), pp. 10–18, 2009.

[10] M. Jockers and D. Witten, A comparative study of machine learning methods for authorship attribution, *Literary and Linguistic Computing*, vol. 25(2), pp. 215–223, 2010.

[11] P. Juola, Operationalizing lexical choice in language change, presented at the *First Conference on Quantitative Investigation in Theoretical Linguistics*, 2002.

[12] P. Juola, Authorship attribution, *Foundations and Trends in Information Retrieval*, vol. 1(3), pp. 233–334, 2008.

[13] P. Juola, Authorship attribution: What mixture-of-experts says we don't yet know, presented at the *American Association for Corpus Linguistics Conference*, 2008.

[14] P. Juola, Authorship and immigration: A case study, *Brooklyn Law School Journal of Law and Policy*, vol. 21(2), pp. 287–298, 2013.

[15] P. Juola, J. Noecker, M. Ryan and S. Speer, JGAAP 4.0 – A revised authorship attribution tool, presented at the *Digital Humanities Conference*, 2009.

[16] P. Juola, M. Ryan and M. Mehok, Geographically localizing tweets using stylometric analysis, presented at the *American Association for Corpus Linguistics Conference*, 2011.

[17] P. Juola and D. Vescovi, Empirical evaluation of authorship obfuscation using JGAAP, *Proceedings of the Third ACM Workshop on Artificial Intelligence and Security*, pp. 14–18, 2010.

[18] M. Koppel, J. Schler and S. Argamon, Computational methods in authorship attribution, *Journal of the American Society for Information Science and Technology*, vol. 60(1), pp. 9–26, 2009.

[19] M. Koppel, J. Schler and K. Zigdon, Determining an author's native language by mining a text for errors, *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 624–628, 2005.

[20] A. McDonald, S. Afroz, A. Caliskan, A. Stolerman and R. Greenstadt, Use fewer instances of the letter "i": Toward writing style anonymization, in *Privacy Enhancing Technologies*, S. Fischer-Hubner and M. Wright (Eds.), Springer-Verlag, Berlin, Germany, pp. 299–318, 2012.

[21] F. Mosteller and D. Wallace, *Inference and Disputed Authorship: The Federalist*, Addison-Wesley, Reading, Massachusetts, 1964.

[22] J. Noecker and P. Juola, Cosine distance nearest-neighbor classification for authorship attribution, presented at the *Digital Humanities Conference*, 2009.

[23] J. Platt, Fast training of support vector machines using sequential minimal optimization, in *Advances in Kernel Methods: Support Vector Learning*, B. Scholkopf, C. Burges and A. Smola (Eds.), MIT Press, Cambridge, Massachusetts, pp. 185–208, 1999.

[24] E. Stamatatos, A survey of modern authorship attribution methods, *Journal of the American Society for Information Science and Technology*, vol. 60(3), pp. 538–556, 2009.

[25] E. Stamatatos, On the robustness of authorship attribution based on character n-gram features, *Brooklyn Law School Journal of Law and Policy*, vol. 21(2), pp. 421–439, 2013.

[26] S. Stein and S. Argamon, A mathematical explanation of Burrows's delta, presented at the *Digital Humanities Conference*, 2006.

[27] H. van Halteren, Author verification by linguistic profiling: An exploration of the parameter space, *ACM Transactions on Speech and Language Processing*, vol. 4(1), pp. 1:1–1:17, 2007.

[28] H. van Halteren, R. Baayen, F. Tweedie, M. Haverkort and A. Neijt, New machine learning methods demonstrate the existence of a human stylome, *Journal of Quantitative Linguistics*, vol. 12(1), pp. 65–77, 2005.

[29] F. Wellman, *The Art of Cross-Examination*, Macmillan, New York, 1936.

[30] N. Zheng, A. Paloski and H. Wang, An efficient user verification system via mouse movements, *Proceedings of the Eighteenth ACM Conference on Computer and Communications Security*, pp. 139–150, 2011.